

# Point Specification in Collaborative Visualization for 3D Scalar Fields Using Augmented Reality

Mickael Sereno · Llonni Besançon · Tobias Isenberg

Received: September 25, 2020 / Accepted: December 3, 2021 / Author version

**Abstract** We compared three techniques to specify 3D positions for collaborative augmented reality (AR) visualization. AR Head-Mounted Displays allow multiple users to share the same physical space, while keeping seamless social interactions. Interactions being key parts of exploratory visualization tasks, we adapted from the Virtual Reality literature three distinct techniques to specify points in 3D space, such as for placing annotations for which they cannot rely on existing data objects. We evaluated these techniques on their accuracy and speed, the user's subjective workload and preferences, as well as their co-presence, mutual understanding, and behavior in collaborative tasks. Our results suggest that all the three techniques provide good mutual understanding and co-presence among collaborators. They differ, however, in the way users behave, their accuracy, and their speed.

**Keywords** AR · CSCW · 3D visualization · 3D point specification

## 1 Introduction

Well-designed interaction is essential for exploratory 3D visualization tasks (Johnson, 2004; Keefe and Isenberg, 2013; Rheingans, 2002; Tory and Möller, 2004) within immersive analytics (Büschel et al., 2018; Marriott et al., 2018), including the selection of Region-Of-Interests (ROIs) (Besançon et al., 2019; Top et al., 2011; Yu et al., 2012, 2016), specific

---

Mickael Sereno  
Université Paris-Saclay, CNRS, Inria, LISN, France  
E-mail: serenomickael@gmail.com

Llonni Besançon  
Linköping University, Sweden  
E-mail: lonni.besancon@gmail.com

Tobias Isenberg  
Université Paris-Saclay, CNRS, Inria, LISN, France  
E-mail: tobias.isenberg@inria.fr

features (e. g., vector orientation) (Jackson et al., 2013; Sherbony et al., 2005), and 3D positions (Besançon et al., 2017b; Fu et al., 2010; Hinckley et al., 1994; Isenberg, 2016; Klein et al., 2012; Yu et al., 2010). To select objects, systems generally propose pointing actions, e. g., via raycasting (Arge-laguet and Andujar, 2013; Forsberg et al., 1996; Grossman and Balakrishnan, 2006; Loosler et al., 2007), or via direct manipulation (Stoakley et al., 1995). In 3D data such as volumes or point/line samples, however, raycasting techniques are of limited use because there are no physical objects that can serve as proxies to be intersected. Research on selection in such dataset exist (Besançon et al., 2019; Hurter et al., 2019; Yu et al., 2016) but focuses mostly on selecting ROIs and does not use AR Head-Mounted Displays (AR-HMD).

Compared to ROIs, pointing at a 3D position allows users to show this point to others, to probe its data (e. g., velocity, temperature), and to place objects at this location. These objects can be annotations, which are important in visualization systems (Czauderna et al., 2018; Harmon et al., 1996; Springmeyer et al., 1992) and important in collaborative systems (Irlitti et al., 2016; Klapperstuck et al., 2016; Szalavári et al., 1998). To study the efficiency of 3D point specification techniques,<sup>1</sup> we focus on whether users are (1) efficient with them in solo tasks, and (2) if they understand what another user is doing, as both are important in collaborative environments (Billinghurst et al., 2018). For example, students in medicine should be able to specify the location of a tumor, for which their teachers should be able to guide them.

AR-HMDs seem particularly suitable for such collaborative exploratory tasks. First, they give each collaborator a custom view adapted to their needs (e. g., (Nilsson et al., 2009)). They also allow users to both see the workspace and their collaborators without requiring active transitions be-

---

<sup>1</sup> We use the term *specification* and not *selection* as we are interested in arbitrary points inside the dataset that are not defined *a priori* as data elements, in contrast to the *selection* of such well-defined points.

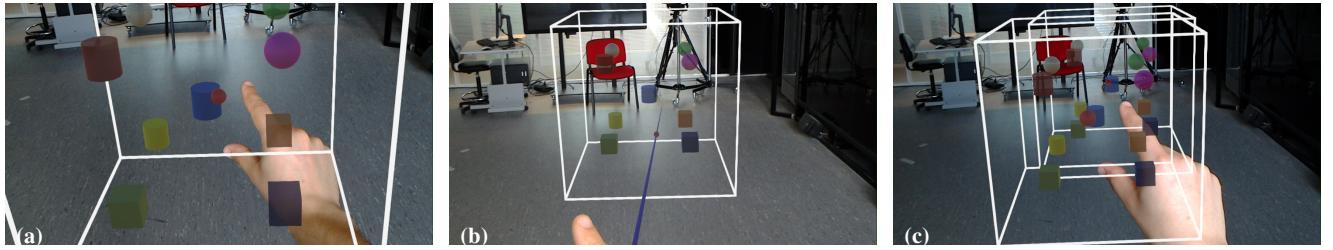


Fig. 1: The three techniques we studied: (a) *Manual*, (b) *AR-Go-Go*, (c) *World-In-Minature (WIM)*. In all cases we display a red spherical 3D cursor ( $r = 2 \text{ cm}$ ) for participants to point to. As we study point specification in general, we use abstract shapes as stand-ins for 3D data. Our participants can thus refer to these colored geometries and do not have to be familiar with a particular type of data. The outline of the dataset turns white when the 3D cursor enters the dataset’s bounding box.

tween communication and analysis. Third, the perspective stereoscopic view matches the human visual system which gives a better feeling of depth perception and object size compared to non-stereoscopic 2D screens (McIntire et al., 2014). Even with these benefits, visualization in AR has not been well studied, with only 36 papers published at IEEE ISMAR (the main AR conference) from 1998 to 2018 discussing it, i.e.,  $\approx 6.7\%$  of the total number of papers published in that venue during that period (Kim et al., 2018; Zhou et al., 2008).

Our contributions are threefold. First, we adapted and implemented three Virtual Reality (VR) 3D selection techniques to AR-HMD environments using Poupyrev et al.’s taxonomy (1998) (Fig. 1). In the first users physically reach into the data to specify 3D points. The second extends a ray from the user’s chest to their dominant hand, with their separating distance controlling the position along the ray. The third creates a proxy of the dataset, which acts for the original. Second, we studied these techniques in a collaborative setting using guidance tasks, in which participants specify 3D points to measure their subjective co-presence and respective understanding. Third, in another experiment, we evaluated the techniques’ performance in non-collaborative AR work. We report on participants’ behavior, performance, fatigue, and the efficiency of the awareness cues they perceived.

## 2 Related Work

Our work relates to 3D point specification for 3D visualization in AR collaborative systems.

### 2.1 Collaborative Systems

Workspace awareness (WA) is a major issue in collaborative environments (Gutwin and Greenberg, 2002). Gutwin and Greenberg (2002) defined WA as “the up-to-the-moment understanding of another person’s interaction with the shared workspace.” The user’s embodiment in a general way, workspace artifacts, or conversational cues (e.g., speech, body language) provide awareness cues. According to Milgram and Kishino’s taxonomy (1994) and Benford et al.’s classification

(1998), AR is closer to the real than to the virtual world. In AR, users thus have access to their traditional tools where communication mostly relies on social protocols, with no particular need for embodiment and conversational cues—the two most important cues according to Gutwin and Greenberg (2002). Co-located AR setups tend to provide better awareness among collaborators, yet there may not be enough of it. Spaces relying on both virtual and real objects must be coherent regarding these two object types. Since more literature addresses interactive data visualization in VR than in AR, we adapted existing VR interaction techniques to AR spaces if they do not directly apply to the changed environment. For example, we adjusted the *Go-Go* technique (Poupyrev et al., 1996) by changing the arm extension (which would not be aligned with reality) to a displayed 3D cursor.

Using AR-HMD, awareness significantly improves user performance (Adcock et al., 2013; Gupta et al., 2016; Pi-umsomboon et al., 2019a; Teo et al., 2018). Because of the strong embodiment and conversational cues provided in synchronous co-located setups, however, awareness has mostly been studied in remote setups (Sereno et al., 2021). The only work we found focusing on co-located 3D selection using AR-HMDs as support is Oda and Feiner’s *GARDEN* system (2012), which facilitates co-located 3D referencing in AR based on the scene’s depth data. They studied both the roles of *indicator* (the one pointing at objects) and *recipient* (the one understanding the intended location).

### 2.2 3D Visualization

3D visualization is considered to require multi-disciplinary work (Keefe, 2010). Its main purpose is to give users better insight into their datasets, e.g., to compare their measured data with theories or to better understand both in a coarse and in a fine way their data. For better understanding, users need to interact with these datasets, e.g., to change the visualization parameters, select region-of-interests (ROIs) which allow fine-grain analysis, or to anchor annotations (Keefe and Isenberg, 2013) to record insights (Lenne et al., 2009) such as annotating blood vessels aneurysms (Saalfeld et al.,

2017). Such visualizations can happen in non-stereoscopic virtual workspaces (e. g., on a 2D screen), in AR, or in VR.

Researchers investigated AR-HMD-based 3D visualization using hybrid PC+AR-HMD systems (Bornik et al., 2006; Wang et al., 2019), where most interactions are initiated and/or controlled by traditional 2D interfaces. Wang et al. (2019) showed that scientists prefer this kind of interface which are close to what scientists currently use. Those hybrid interfaces, however, do not encourage users to move around their workspace, and do not merge the social and visualization spaces. Yet this merger would be one of the main benefits of using AR-HMDs for collaborative work. We argue that replacing workstations by multi-touch tablets can solve these issues, while still providing the familiar 2D interfaces.

Keefe and Isenberg (2013) showed why tangible and spatial input, on which AR and VR systems usually rely, may improve the visualization process, and discussed research challenges for this input in 3D visualization. Navigation and selection, e. g., are crucial in visualization software and scientific workflows. Below we review such interaction techniques related to our own work, in particular those used in VR.

### 2.2.1 VR-Based Selection Techniques

Because of the 3D and direct interaction nature of VR (Bruckner et al., 2019), researchers intensively studied point and region selection in this environment (Argelaguet and Andujar, 2013; Besançon et al., 2019). Most of these techniques rely on geometry or ray casting metaphors to allow users to reach far targets (Argelaguet and Andujar, 2013; Besançon et al., 2019). Techniques relying on geometries such as cubes and spheres cannot specify points due to their volumetric nature. Pure ray casting metaphors are also not applicable in our context as they rely on intersections with physical geometries which do not exist in many 3D datasets such as volumetric data (Besançon et al., 2019; Yu et al., 2012).

Work addressing occlusion in dense environments may also apply to 3D point specification, in particular methods that enhance ray casting with an additional degree of freedom (DoF) such as a movable cursor (Baloup et al., 2019; Grossman and Balakrishnan, 2006; Ro et al., 2017; Rosa and Nagel, 2010). This additional DoF allows users to select an object in space and, while being categorized as ray casting, they are closer to the remote virtual hand metaphor such as Poupyrev et al.'s *Go-Go* technique (1996). We can, for instance, define a ray from the user's chest to their hand and beyond, enhanced with a 3D cursor that is controlled by the distance between the user's chest and hand, to target 3D points. Compared to other ray+cursor metaphors, this adaptation of the *Go-Go* technique does not need external interactive devices to determine the position of the cursor along the ray. It then seems suitable for AR-HMD systems.

Ray+cursor techniques, however, are limited to big and

close areas as the orientational jitter amplified by the ray cannot benefit from conflict-free algorithms (e. g., the *Bubble Cursor* (Grossman and Balakrishnan, 2005)) or by adapting the radius of the cursor (Rosa and Nagel, 2010) depending on its location, as there is no well defined target in our case.

### 2.2.2 Multi-touch device-based Selection Techniques

AR-HMDs are limited to mid-air gestures (which causes fatigue (Hincapié-Ramos et al., 2014)), speech, and eye-based interactions. They can, however, be coupled with hand-held devices as users with AR-HMDs still perceive the real-world, in contrast to VR. Such devices provide various input modalities (see also Surale et al.'s design space (2019)). For instance, the 2D interface facilitates the opening of datasets and 2D annotation sketching. They also facilitate touch input for 3D ROIs selections such as *SpaceCast* (Yu et al., 2016), which combines 2D input and the density of a volume to infer the ROI intended for selection, or *TangibleBrush* (Besançon et al., 2019) which relies on a position-aware tablet to extrude a lasso drawn on the tablet into 3D space. Their touch display also allow users to specify 3D points using, e. g., the *Balloon* technique (Benko and Feiner, 2007; Daiber et al., 2012).

For collaboration, however, non-tangible tablets may not allow other co-workers to be aware of actions because they cannot see this personal screen. Moreover, interaction based on motions may be physically too demanding because users are forced to hold the device in specific positions. We thus decided to focus on the AR capabilities to provide direct manipulations in the real-world using hands instead of tracked devices or non-versatile techniques.

## 3 Motivation

We begin by explaining our design decisions, which we later investigated in our experiments.

### 3.1 Three Interaction Techniques for Point Specification

We wanted to study 3D point specification techniques in AR-HMDs collaborative contexts. Since most direct manipulation techniques in VR rely on hand manipulations (with special controllers for the Oculus Rift or the HTC Vive VR-HMDs) and because the VR literature is more mature than the AR literature regarding interaction techniques, we chose from the VR literature three state-of-the-art 3D interaction techniques using hands as primary input to cover all of Poupyrev et al.'s taxonomy (Fig. 2). We aimed, with this approach, to understand the effects of most techniques in collaborative contexts by studying three techniques which represent the whole taxonomy. Since pure ray casting techniques do not support 3D point specification in our setup, we ignored that

part of the taxonomy and selected from the hand metaphors part (e.g., Poupyrev et al., 1996) and exocentric techniques (e.g., Pierce et al., 1999; Stoakley et al., 1995). Poupyrev et al.’s taxonomy (1998), however, does not consider the techniques’ reach. Users can indeed virtually reach any object in VR setups with their hands (e.g., Poupyrev et al., 1996), no matter if they are within arm’s reach or not. Because in AR we cannot occlude the user’s real arm, we extended Poupyrev et al.’s taxonomy to consider the proximity of objects of interest (remote or within arm’s reach) and decided to implement two virtual hand techniques (one with remote and one with local targets) as well as an exocentric interaction technique.

While we focus on co-located contexts, we hypothesize that direct manipulations (virtual hand metaphor—non-remote) may not always be suitable. Indeed, by manipulating the public view, one user may physically occlude the workspace to others, or several users may want to reach the same point in space, introducing conflicts. Remote interactions may resolve these issues. We summarize our choices in Fig. 2 and show our techniques schematically in Fig. 3.

### 3.2 Scenarios

We want to determine the techniques that work best in collaborative scenarios. We identified particularly the three following ones: (1) a solo task where the user works alone, (2) a collaborative task where collaborators work on the same object, and (3) a collaborative task where collaborators are working in parallel with two distinct regions of the environment. We foresee, moreover, that users should be able to do more than only specifying points in a dataset, e.g., they also may need to manipulate it in the 3D space, select ROIs, and add annotations. Finally, in such collaborative contexts, users should be able to move freely. We believe that adding a general-purpose device like a multi-touch tablet may solve these requirements, which we included in our setup to be as closed as possible to a real use-case. Studying these three scenarios in one study would, however, be cumbersome due to the limited time participants can focus their attention ( $\approx 1$  hour). We thus focused on only the two first scenarios and give hypotheses regarding parallel tasks based on the results gathered from these two scenarios. To study these scenarios, we created a collaborative system which we present next.

## 4 Implementation

Our collaborative system uses multiple Microsoft HoloLens (1<sup>st</sup> gen.). Their software relies on Unity 2019/C# .Net, and multiple homemade C++ plugins, including one relying on OpenCV. In addition, each user uses a tablet (Samsung’s Galaxy Tab S4 with Android 9) to ultimately be able to provide rich input in a 3D data analysis system. The tablet’s

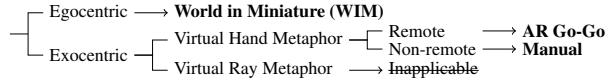


Fig. 2: Summary of our techniques’ choice to cover an extended version of Poupyrev et al.’s taxonomy (1998).

software relies on Android NDK/SDK. In the more complete software we are developing, we use the tablets to handle the views and the datasets, e.g., opening/closing, 3D transformations, filtering, entering texts. In our experiments, we used them to control the tasks. To be able to support two or more collaborators, we implemented a client-server system that connects the HMDs and tablets via WiFi. The server runs on a Linux distribution (ArchLinux, kernel v5.0.11, i5-7300HQ, 6 GB RAM) which handles the communication between all the devices and logs all received and sent network messages. We wrote the server software from scratch using C++ and the associated UNIX API. We based the communication on TCP/IP with a custom communication format that is energy-efficient w.r.t. the computation each device has complete to parse the messages. We provide the server software and links to other parts of the system at [https://github.com/MickaelSERENO/SciVis\\_Server/tree/CHI2020](https://github.com/MickaelSERENO/SciVis_Server/tree/CHI2020).

Using the HoloLens’ API, the first connected HMD sends the room’s data to the server, which relays it to the other HMDs to create a shared coordinate system. Meanwhile, each HMD displays an IP address that users must enter on their respective tablet to pair the devices. During pairing, each user selects their handedness. After pairing, the tablet loads the system’s status (e.g., opened datasets, created annotations). Each HMD sends, at each frame update ( $\approx 60$  Hz), its status (position, orientation, interaction technique data) to the server. The server dispatches this data to all clients at  $\approx 20$  Hz, depending on the network connection quality and the time each client has to parse the TCP/IP packets. We choose this frequency to provide a fluid experience, while avoiding too much computational overhead for all clients.

Next, we describe the techniques we chose and implemented (see also Fig. 1 and 3). Based on our discussion of related work, we implemented an exocentric and two virtual hand metaphor techniques (one scale 1:1 and one remote interaction technique). These techniques rely on the 3D position of the user’s dominant hand. All techniques display a 3D cursor (radius=2 cm) representing the currently highlighted position. Since this feedback is impossible for every hand shape with the hand tracker of the HoloLens’ API (version: MRTK RC1), we implemented a custom computer vision tracking algorithm on top of OpenCV. We analyze one of the HoloLens’ depth camera streams that we access via the Microsoft’s Media Foundation API in “research mode.” To validate a position, the user performs a “Tap” gesture, which the HoloLens’ API natively recognizes. This gesture is com-

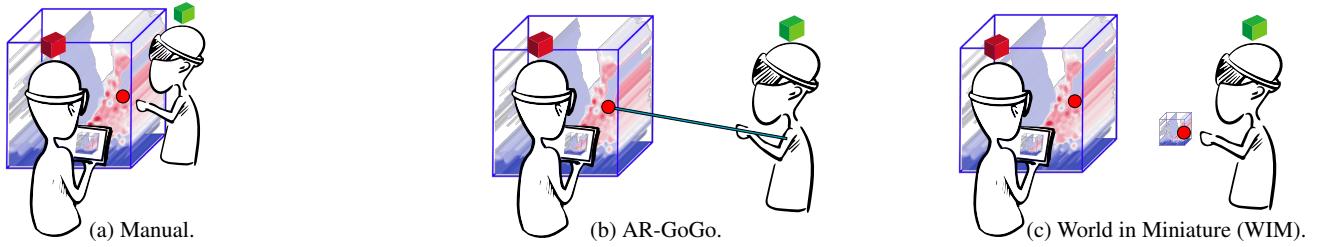


Fig. 3: Sketches of the representative techniques we selected. See Fig. 1 for screenshots of their respective implementations.

patible with our hand detection but generates a small shift (called the Heisenberg effect (Bowman et al., 2001)) in the detected hand position due to the change of the hand shape and the shaking created by the muscles. After further analysis, we found that the shift is higher in the native API than in our custom algorithm, but without removing it. To reduce this effect, we applied an exponential low-pass filter ( $\alpha = 0.3$ ) to the position input. Participants in our study did not report any issue (such as latency) regarding the filtering.

We did not use another tracking solution such as data gloves or external tracking system because we expect that such AR collaborative systems should be portable with low maintenance to be adopted by scientists, as also argued by Wang et al. (2019) and Issartel et al. (2016). Current standalone HMDs, including VR (e.g., Oculus Quest) and AR (e.g., Microsoft’s HoloLens 2<sup>nd</sup> gen.) HMDs, show a tendency of using visual tracking based on RGB-D cameras. The technology is mature enough to be used, thus we do not foresee that other types of sensors would be used for public devices in the near future: markerless trackers are faster to deploy by not needing active calibration and specific cumbersome external devices.

#### 4.1 Manual technique

In the *Manual* technique (Fig. 1(a), 3(a)), a user must physically reach the position of interest—a local virtual hand metaphor. We display a small red sphere near the user’s thumb that acts as a 3D cursor, visible by all co-workers. We use an approximate thumb position because we think it is more comfortable than using the tracked palm position. To test our intuition, we ran a small pilot study with 8 volunteers from our lab (4 females, 4 males; 3 using the HoloLens daily) and asked them to place ten 3D objects in an empty space with both modes. We found that four people strongly agreed and three agreed that they preferred using the estimated thumb position, compared to the accurate palm position. Only one person preferred the palm mode because of its accuracy, the ball not representing an extension of his hand in thumb mode. When looking at Bruder et al.’s study (2013), users tend to be more accurate when a small offset is applied, at the cost of speed. We prefer, however, to em-

phasize accuracy over speed due the tracking system relying on image processing and, hence, being subject to noise and because 3D data visualization usually requires high accuracy.

We heuristically compute a thumb position by taking the palm position and translating it by 6 cm to the left (or to the right for left-handed users), 3 cm to the top, and 11 cm forward. The left and forward axes are first defined in the user-centric coordinate system (i.e., head orientation) and then projected onto the horizontal plane. We derived this translation in another pilot study and it corresponds approximately to the thumb position when a user is in the HoloLens API’s “ready” state. However, this design suffers from the noise in the user’s head orientation since we use this orientation to determine the forward and left axis. This noise is comparable to the noise of the head-gaze interaction technique (as both are based on the head orientation) and it has a low impact as we only re-orient a small vector (i.e., magnitude < 13 cm).

We did not use a Leap Motion for two reasons. First, users need to move, so adding a device that requires to be wired to a computer is highly inconvenient. Other solutions such as a separate Raspberry Pi would decrease the portability and thus would create new limitations (Issartel et al., 2016). Second, the Leap Motion also does not provide accurate finger detection due to occlusion and low pixels count for the fingertips. Consequently, hand-tracking without finger-tracking provides us with more stability as there are more pixels in the depth image. Moreover, the small mismatch between the true and the heuristic finger position did not bother participants during our pilots or during the experiments.

#### 4.2 AR Go-Go interaction technique

The *AR Go-Go* technique (Fig. 1(b), 3(b)) allows users to remotely specify points using a virtual hand metaphor. We cast a ray from the user’s chest that follows their dominant hand and place a 3D cursor along the ray, both being shown to all users. We compute the 3D cursor distance as

$$d = 13.0 \times |(\text{hand} - (\text{head} + (0, -0.15, 0)))| - 0.20 \quad (\text{in m}).$$

With this formula we determine the chest position as being 15 cm below the head position for comfortable ray steering. We then determine the absolute distance between the chest

and the hand position and set the ray origin (i.e., where the cursor overlaps the user’s hand) to be  $\frac{0.20 \times 13.0}{12.0} \approx 0.21$  m forward, which roughly corresponds to the distance where users can clearly see their hands and we can reliably track them using the depth camera. We use an application- or environment-dependent gain factor of 13.0 to determine the final 3D cursor position. This gain factor trades-off the noise amplification with the reachable distance which is about 8 m for a chest–hand distance of 80 cm with our applied offset and gain factor. This 8 m reachable distance justifies its remote nature. Compared to the original *Go-Go* technique (Poupyrev et al., 1996), we did not want to use a progressive gain-factor because we want users to be accurate regardless of the distance between the user and the position of interest (see Bowman et al.’s results Bowman et al. (1999)), because a 8 m distance is typically sufficient for in-door AR applications, and because we want to compare it to local hand metaphors. Moreover, due to its remote nature, it would not have made sense to apply the “finger” offset as we did in *Manual*. Thus, we did not apply an offset in *AR Go-Go* on the hand 3D position. Also, compared to the original *Go-Go* technique, our virtual hand metaphor does not suffer from the users seeing their real hands, because we use a ray+cursor representation instead of an arm extension representation.

As Looser et al. (2007) showed with a Wiimote for ray casting, the noise is more perceivable in this technique than in *Manual*, as we convert a positional jitter to an orientational jitter. It is well known, however, that ray casting is not well suited for selecting small objects/regions at a great distance (Forsberg et al., 1996; Poupyrev et al., 1998). The perceived noise is also affected by the head position noise used to determine the chest position. To limit all these effects, we applied an exponential filter ( $\alpha = 0.3$ ) on the target location in addition to the already filtered hand position.

#### 4.3 World In Miniature (WIM)

The *WIM* technique (Fig. 1(c), 3(c)) creates a copy of the dataset placed in front of the user as an exocentric metaphor. For this purpose, we place the copy 45 cm away and 15 cm below of the HMD’s position and resize it to be at most 25 cm in its longest axis. We used these values to maximize the copy size under the constraint of selecting any point in the WIM, without walking, while seeing the entire WIM through the HMD. This copy cannot be modified (e.g., no translation, rotation, scaling, or visual changes); it only serves as an interaction proxy. We use the same orientation for the copy as in the original, allowing co-workers to communicate using normal language (e.g., left, forward). We then display the 3D cursor both near the user’s thumb (as in *Manual*) and inside the original dataset at the corresponding local position. Both cursors have the same radius (2 cm). We show both the data copy and the two cursors to all collaborators, providing them

cues about all the collaborator’s actions and the intended 3D position in the local space of both the WIM and the original.

## 5 User Study

We designed a within-subject study to measure the performance of the three chosen techniques, which we consider to be different enough to discuss the usability of the metaphors. Because replicating real scenarios is complicated due to knowledge barriers, in a first experiment we created a pair guidance task done using a simple scene to measure the collaborative aspects of the techniques. These tasks both rely on the “actor” performing the task and the “guider” who has more knowledge. In our study, a guider is someone who knows the target position and guides the actor to it. The advantages of this setup are that (1) our study does not rely on visualization features (we thus reduce the bias) and (2) we do not need participants to understand specific datasets to be able to specify 3D locations, while still forcing them to collaborate based on the visual feedback these techniques provide. Such guidance tasks were used in the past to measure the efficiency of awareness cues (e.g., Lee et al., 2018). While we did not expect usual collaborative scenarios relying on guidance except for teaching, such tasks allow us to see how users can communicate, use, and, hence, understand the awareness cues. Moreover, our motivation is interactive 3D visualization such as voxel datasets rendered volumetrically. One main task in this domain is the definition of transfer functions that modify the transparency of cells such that users see the needed properties (e.g., bones and organs in medicine, hurricane features in meteorology). Then, scenarios where occlusions are an issue should be rare. Our sample dataset thus seems suitable to represent such cases, without participants actually needing to understand these sciences. Finally, we are not aware of prior work which compared these techniques in AR and VR contexts. We thus also discuss performance results for all techniques based on a second experiment.

### 5.1 Hypotheses

Our goal was to measure the performance and the collaborative metrics of each interaction technique mentioned previously. Based on a pilot study we carried out to test the system ourselves, we aimed to test the following six hypotheses:

- H1 *AR Go-Go* is the slowest and least accurate technique because ray metaphors are bad at accurate positioning at great distances, which then require time to adjust,
- H2 *Manual* is preferred in general due to the accuracy and speed of manual interaction,
- H3 *WIM* is the fastest technique as users can interact, similar to *Manual*, without the need to physically move toward (or into) the dataset,

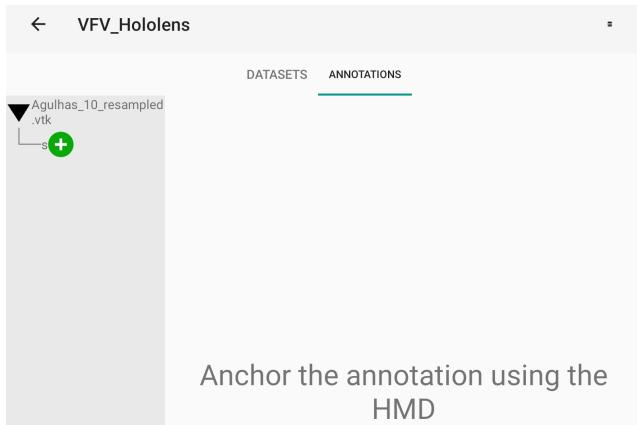


Fig. 4: Interface to start annotation anchoring (via green “+” button). Users can select which technique to use during training using additional menus (top-right menu icon).

- H4 *Manual* provides the best co-presence/mutual understanding as the users face each other,
- H5 *Manual* is the most accurate technique since the original dataset is larger than its *WIM* version and since **H1** hypothesizes that *AR Go-Go* is the least accurate, and
- H6 *AR Go-Go* supports collaboration best because the ray displayed to all users may support better directional cues, e.g., the meaning of expressions such as “forward.”

## 5.2 Common Design

For both experiments, we asked participants to anchor annotations using the mentioned interaction techniques and the provided touch tablet. We asked participants to anchor the annotations in a  $50\text{ cm} \times 50\text{ cm} \times 50\text{ cm}$  dataset that comprised simple colored geometries (cubes, spheres, and cylinders). While the tablet was not mandatory for such experiment, we do not foresee that scientists will explore 3D datasets using only AR-HMDs; see Wang et al.’s study (2019) regarding the use of a hybrid interface for particle physics scientists and Bornik et al.’s work (2006) for 3D liver segmentation in medicine. Our setup then closely matches a realistic setup where the touch tablet is used as a 2D interface for usual interaction, while ensuring mobility compared to workstations.

To start the HMD-based annotation anchoring, participants tapped the “+” button on the tablet, next to the dataset label (Fig. 4). We thus clearly separated the annotation placement from normal tracking. Once the annotation was anchored using the chosen technique, we rendered a small yellow 3D star at that location in AR space for all collaborators and stopped the hand tracking. As a constraint in every condition, we asked participants to start the anchoring process at least 1.5 m away from the center of the dataset (they can then move if needed), forcing them to move in the *Manual*

	Manual	WIM	Go-Go
study 1	(2;1)	(4;2)	(3;2)
study 2	(5;5)	(6;3)	(5;4)

Table 1: Number of (*accuracy; speed*) data points discarded per technique per study. Each condition comprises 192 trials.

condition. We represented this radius with 4 markers on the floor, one for each direction. This constraint models normal sessions in which users will not always stay near the dataset as they occasionally need to get a global overview of the environment. Without this constraint, the *Manual* condition would not require participants to physically move around the room, which may bias the results compared to real use-cases.

Each participant experienced all interaction techniques. We counter-balanced the technique order with a  $3 \times 3$  cyclic Latin square. Each participant used, for every trial, a position from his/her assigned pre-generated pool to reduce learning effects. We reused both pools for all participant pairs. Each pool contained eight 3D positions, determined by a uniform random generator and rejecting points inside the colored shapes. Values ranged from  $-0.5$  to  $+0.5$  along the  $x$ -,  $y$ - and  $z$ -axes, defined in the dataset’s local coordinates. We recomputed the target position order at the start of each condition using a uniform random generator, without replacement.

Participants could take a break at the end of each condition in both experiments. To end the break, we asked both participants to tap on the corresponding button on their tablets.

## 5.3 Data analysis and pre-registration

Study data is often analyzed with Null Hypothesis Significance Testing (NHST) and methods such as ANalysis Of VAriance (ANOVA). Yet we accept the criticism of NHST-based data analysis (Amrhein et al., 2018, 2019; Dragicevic et al., 2014; Gigerenzer, 2018; McShane and Gal, 2017; Valentine et al., 2015) which often results in dichotomous conclusions (Besançon and Dragicevic, 2019; Helske et al., 2021), which in turn can hinder the robustness of findings and push for questionable research practices—both of these increasing the risk of a replication crisis (Amrhein et al., 2019; Cockburn et al., 2020). We follow current APA recommendations (VandenBos, 2009) and report our results using estimation techniques with effect sizes and confidence intervals (CIs) instead of  $p$ -values, which users can still “infer[ence] by eye” (Krzywinski and Altman, 2013). The effect sizes refer to our measured means and not to standardized effect sizes (Coe, 2002), whose reporting can be problematic (Baguley, 2009). We interpret our effect sizes and CIs as providing different degrees of evidence about the population mean (Besançon and Dragicevic, 2019; Cumming, 2014; Dragicevic, 2016) and thus avoid words such as ‘significant,’ we rather adapt our language to our results Besançon et al. (2021).

We pre-registered our experiments at [osf.io/43j9g](https://osf.io/43j9g/), following current best practices (Cockburn et al., 2018, 2020), and we provide the current scripts and our gathered data at [osf.io/7a6yr](https://osf.io/7a6yr). We modified the pre-registered scripts to correct for logical errors. We also removed miss-clicks performed by users, i. e., detection of false-positive gestures which users reported during the sessions. We noted the users' respective trials and removed these from our statistical analysis, resulting in six discarded trials. After further analysis, we replaced the arithmetic mean with the geometric mean for all pair-wise comparison ratios, as it came to our attention that the geometric mean is a more adequate measure of central tendency for ratios (Fleming and Wallace, 1986). Although we pre-registered our analysis without outlier removal, the noisy data obtained from our sensors due to image-processing tracking led us to question if our results would also hold if we remove outliers for the speed and accuracy metrics. We define outliers as values being distant of the mean by more than three times the standard deviation. We summarize the numbers of discarded values in Table 1. Except for speed and ranking analysis, we represent all statistical results using a 95% Bootstrap CI (BCA method). Finally, we also video-recorded every session for exploratory analysis.

Finally, we asked participants to rate their familiarity with immersive HMDs on a 5-point Likert scale (separately for AR and VR, from “never used” (1) to “daily use” (5)) and the subjective accuracy of the hand tracking on a 7-point Likert scale. We also asked them for comments and to rank each technique (best: 1) w.r.t. *co-presence, message understanding, accuracy, speed, parallel tasks, and overall ranking*.

#### 5.4 Participants

A total of 12 pairs of participants (6 females, mean age=28.4, median=25.0, SD=7.37, range=23–56) completed our experiments as we pre-registered. 23 participants were right-handed. 22 participants were students or researchers in computer science. Users reported their VR usage (mean=2.2, median=2, SD=0.9, range=1–4), AR usage (mean=2.5, median=2, SD=1.1, range= 1–5), and the hand tracking subjective accuracy (mean=4.1, median=5, SD=1.4, range=1–6). We denote each pair as  $P_i$ ,  $i \in [0, 11]$ ; for each of these we denote participants as either ID 0 or ID 1; so overall we refer to  $P_{ij}$ ,  $j \in [0, 1]$ . Overall, we repeat the  $3 \times 3$  Latin square we used to counter-balance our techniques  $12/3 = 4$  times. This gives us enough data to perform statistical data analyses.

#### 5.5 First Experiment – Collaborative Tasks

The purpose of the first experiment was to measure the efficiency of awareness cues perceived by users in collaboration. For that, we used a guidance task which we explain next.

##### 5.5.1 Procedure

We first welcomed the participants and asked them to read and sign the consent and media-release forms. The experimenter then explained them the overall context of our work. Next, to ensure a comparable setup, he used experimenter-only functions on the tablets to place and orient the dataset correctly in the physical space for participants to move around it with ease. Each participant then put on a Microsoft's HoloLens 1<sup>st</sup> gen. and took a Samsung Galaxy Tab S4. The experimenter then explained the three interaction techniques and the overall system. For the purpose of the study, we used the previously mentioned  $50 \times 50 \times 50 \text{ cm}^3$  dataset.

**Training.** We began with a training session for participants to test all the techniques and anchor as many annotations as they wanted. The experimenter explained each technique and asked participants to anchor multiple annotations, until they felt comfortable. At the end of the training session, we launched the first part of the study.

**Tasks.** In each condition, we asked each participant to anchor eight annotations, resulting in 16 trials per condition per pair. For each trial, one participant saw where the annotation should be anchored as a transparent star as a proxy position for structures that users would mentally target in a real dataset, e. g., the eye of a hurricane. Because in reality such explicit object may not exist and certainly is not known *a priori*, we did not highlight it when the 3D cursor was near the location. We then asked this participant to guide his/her partner to the marked location, as accurately as possible. Once anchored, we displayed a star annotation at the specified location for 2 seconds to both participants as visual feedback, before launching the next trial and switching the participants' roles. At the end of this first experiment, we proposed a break and asked participants to fill a questionnaire.

We purposefully did not instruct participants on how to interact with each other to see how they would intuitively collaborate in realistic scenarios for each technique. Here we thus focused on people's behavior, while we measured each technique's performance in our second experiment.

**Questionnaire.** We asked participants to rate, on a 7-point Likert scale, their perceived *co-presence* and *message understanding*, which we took from Harms and Biocca's (2004) *Networked Minds Measure of Social Presence* questionnaire. We were inspired by related studies, e. g., by Piumsomboon et al. (2019b) and Lee et al. (2018) who also used a subset of the proposed six metrics. We averaged the results of all questions per factor item to get one metric per factor, considering positive and negative questions.

##### 5.5.2 Pre-registered Results

We now present the planned analysis results of this first experiment. Our graphs include CIs with and without outliers.

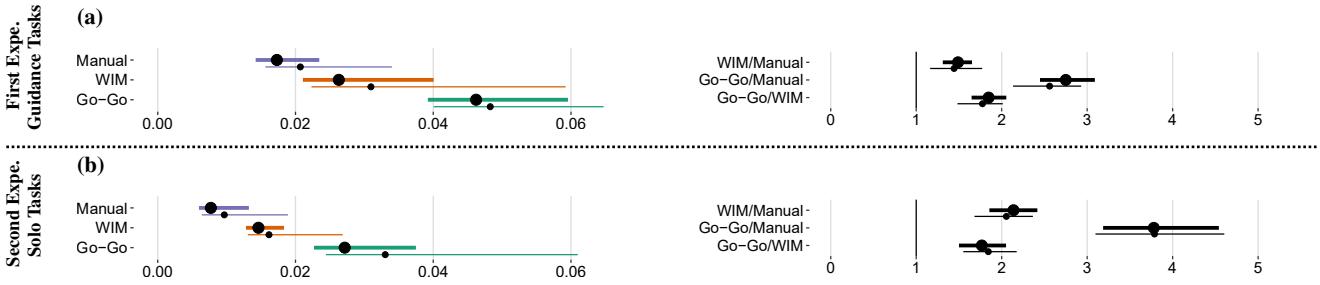


Fig. 5: Accuracy results for the two experiments of the study. For each, the left graph represents using a 95% Bootstrap CI the distance error of each technique in meters. The right graph represents the pair-wise comparisons (ratio) 95% CI for the three techniques. For all metrics, the top CI is computed with outliers removed and the bottom CI represents all data.

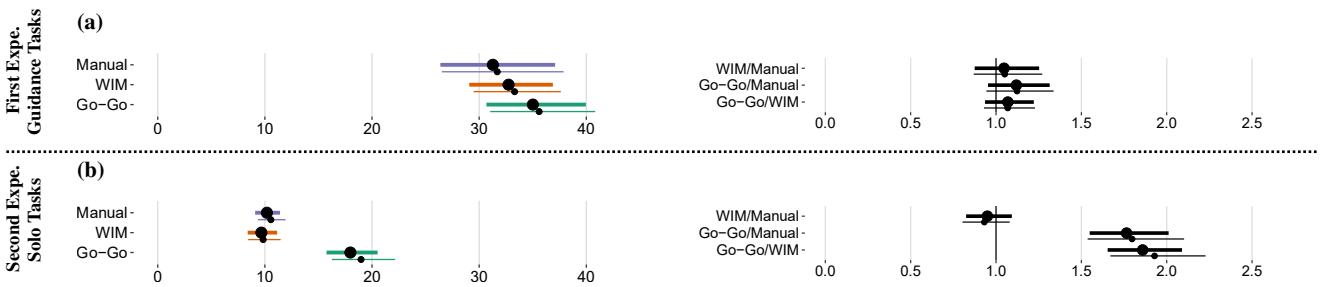


Fig. 6: Time completion task (TCT) for the two experiments of the study. For each, the left graph represents using a 95% CI the TCT of each technique in seconds. The right graph represents the pair-wise comparisons (ratio) 95% CI for the three techniques. For all metrics, the top CI is computed with outliers removed and the bottom CI represents all data.

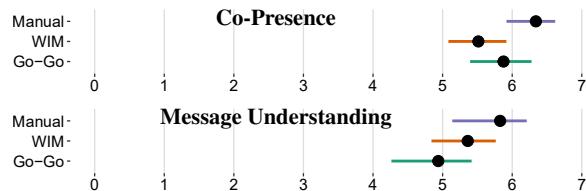


Fig. 7: 95% Bootstrap Confidence Interval (BCA method) of the Co-Presence and Message Understanding metrics.

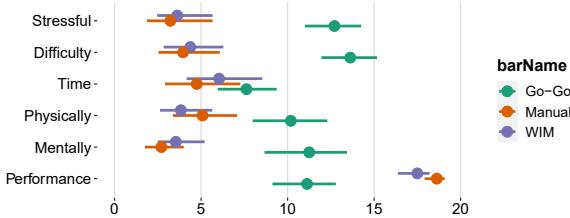


Fig. 8: The results of the NasaTLX questionnaire represented using a 95% Bootstrap Confidence Interval (BCA method).

We found that our pairwise comparison results show similar patterns for both cases. We will use by default the CIs computed without outliers during the discussion.

**Accuracy.** Fig. 5a shows the distance error per technique and pair-wise comparisons for all the three techniques. We computed the values in the dataset's local coordinate system,

which means that a physical 2 cm accuracy in the WIM technique is reported as a 4 cm accuracy because copies were halved. We see strong evidence for *Manual* being better than *Go-Go* but the overlap of CIs between *Manual* and *WIM* as well as *WIM* and *Go-Go* do not reveal a difference between these techniques. When inspecting the pair-wise ratios, however, the CIs reveal additional strong evidence. We see that *Manual* is about 1.5× as accurate as *WIM* and about 2.5× as accurate as *Go-Go*. *WIM* is a bit less than 2× as accurate as *Go-Go*. These results then support **H1** and **H5**.

**Task Completion Time.** We analyzed log-transformed measurements to correct for positive skewness and present anti-logged results, a standard data analysis process (Sauro and Lewis, 2010) commonly used in HCI (e.g., (Besançon et al., 2017a; Jansen et al., 2013; Le Goc et al., 2016; Yu et al., 2016)). We thus computed geometric means (Fig. 6).<sup>2</sup> The participant starts the timer by pressing the “+” button and ends it when he/she anchors the annotation. Re-pressing the “+” button restarts the timer, which allows the experiment to resolve issues just-in-time. Indeed, because a participant needs both hands to restart the timer (one hand to hold the tablet, and one hand to press the button), the whole phase of

<sup>2</sup> While an arithmetic mean is based on the sum of a set of values, a geometric mean uses the product of the values. It dampens the effect of potential extreme completion times, otherwise these could have biased an arithmetic mean. We plotted results and pair-wise ratios in Fig. 6a.

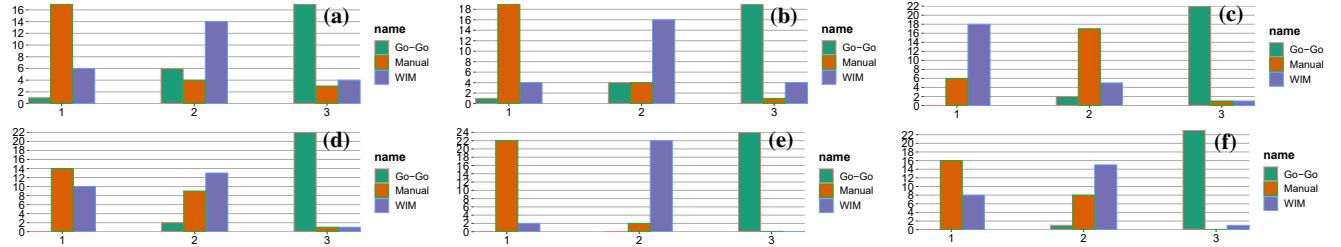


Fig. 9: Ranking counting results: (a) co-presence, (b) message understanding, (c) parallelism, (d) speed, (e) accuracy, and (f) general. 1 is best.

refinement, the longest phase, needs to be redone from the beginning, which preserves the validity of our speed results. Fig. 6 shows no evidence that one technique may be faster than another one. This contradicts **H1** and **H3**.

**Co-Presence and Message Understanding.** Fig. 7 shows the subjective *co-presence* and *message understanding* metrics measured at the end of the first experiment. For all techniques, values ranged from 4 to 7 (values below 5: AR Go-Go:  $n = 3$ , WIM:  $n = 6$  and Manual:  $n = 2$ ) on a 7-point Likert scale. We see weak evidence that users feel more present in *Manual* than in *WIM*. We find no evidence that *WIM* and *Manual* would differ from AR Go-Go at supporting *co-presence*. We see weak evidence that users understand their partner better with *Manual* than with AR Go-Go. We do not see differences, however, between *WIM* and *Manual*, and between *WIM* and AR Go-Go, which could be seen as contradicting **H4**. Nonetheless, the ranking questionnaire shows different results regarding these metrics, with *Manual* providing the most co-presence and best supporting mutual understanding.

### 5.5.3 Exploratory Results

Through the media recordings, we extracted behavior patterns in an exploratory way. We watched all video recordings, and now report and discuss the extracted results.

First, framing the dataset inside a cube (Fig. 1) allowed participants to use the edges and faces to support their instructions (e.g., “near the top”). These references were used a lot during our experiment by all participants, in addition to the available geometries. These observations confirm the usefulness of 3D windows (Schmalstieg et al., 2002) because concepts such as ceiling and wall exist for virtually every visualization. We also observed P0–9<sub>0,1</sub> to often go back to a particular place, after completing their trials as actors in the remote conditions (AR Go-Go and WIM). In most trials, actors started the interaction before being instructed by the guider, which may have affected the computed speed.

**Manual** is the most direct method, i.e., there is no artifact between the object of interest and the interaction. Here, participants generally used the actor’s local coordinate system (left-right, up-down, forward-backward). They sometimes

added their body as a direction, using cues such as “towards me” or using pointing directions such as “this way” (pointing to the right). With this technique, participants tended to use “here” cues by pointing to the correct location with their fingers, fingers that they took back because of tracking interferences. Those interferences are inevitable when using markerless tracking system, on which standalone HMDs such as Microsoft’s HoloLens and the Oculus Quest rely—they cannot determine which hand belongs to whom. Due to these tracking interferences, P6<sub>1</sub> created a “v” sign using two fingers to show the target location. To be as accurate as possible, P0<sub>1</sub> grabbed his partner’s hand for one trial, using it as a remote controller to adjust the position correctly. In contrast, P2 seemed reluctant about using deictic communication and instead relied on voice instructions only. We also saw some participants (P4<sub>1</sub>, P8<sub>1</sub>, P9<sub>0</sub>) staying behind their partner. P4<sub>1</sub> said that he preferred staying behind his partner, as opposed to being in front, to keep the “left” and “right” directions the same for both users, even if it interfered with the tracking system due to the proximity of both users’ hands. Except for occasionally confusing “left” with “right,” we did not observe any ambiguity for any participant.

In our study, we forced participants to start the interaction at 1.5 m away from the dataset, which confused some participants who did not see any visual feedback when pressing the “+” button within that radius. We then reminded them about this minimal distance and to repress the “+” button for resetting the timer. Even with this reset, this design may have affected some results (e.g., frustration metrics).

**AR Go-Go** allows users to communicate about the same object without standing near each other. While it facilitates hand tracking, it also provides better deictic and gestural communication in the guidance task as P7<sub>0</sub> pointed out. Except for P2, all participants pointed at positions with their fingers and used non-verbal gestures as cues (e.g., gestures for moving to the left and right directions). In addition, P4 used mostly gestural communication and nearby geometries, and P11 guided their partner with their palms, instead of using fingers. By manipulating a remote object, however, occlusion problems could appear, which P1 confirmed by asking his partner to physically move to a location with better visibility.

Except for *P4* and *P8*, participants used the ray of the *AR Go-Go* as a support for *forward* and *backward* directions, instead of the actor's local coordinate system. *P4* and *P8* just pointed to the location and gave directions using the provided geometries. *AR Go-Go* allows users to adjust the ray orientation (2 DoFs) and then the depth position along this ray (1 DoF). Because both users use the same object of interest (i.e., the guider can point at the original dataset, while the actor performs the task), the actor can align the ray with the target or the guider's finger. We then saw *P10* (actor) asking multiple times to his partner if his ray is aligned with the target, and *P11* (guider) asked his partner to align the ray with his finger, before adjusting the depth and notifying his partner when the trajectory was correct. *P60* suggested to his partner to first manage the orientation before managing the depth position. All these observations support **H6**.

However, hand movements are amplified, hence covering a larger area compared to the *Manual* technique. Participants thus tended to say a lot "yes" followed by "no" to specify that the position was correct during the whole action. While showing first an accuracy limitation, it allows users to get a coarse impression of the target position, before refining their input. This "amplification", however, creates a noticeable Heisenberg effect when clicking, which frustrated the participants a lot. It is also known that mid-air hand gestures lack stability, in addition to this perceived shift during the movement (Argelaguet and Andujar, 2013).

In the **WIM** condition, participants needed several trials to understand that they can guide their partner using the copy, in addition of the original dataset. This introduces a delay which might contradict **H3**, as reported in Sec. 5.5.2. *P00*, *P101*, *P41*, and *P60*, when helping their partner, first looked at and guided the partner through the original dataset, before focusing on the copy. This may be due to the original dataset always being visible, compared to the copy that has to be summoned. Its sudden disappearance may also be uncomfortable, as pointed out by *P5* who wanted to have more time to check the results. *P1* and *P5* reported that, sometimes, they did not have time to check the results.

When working only on the copy, participants tended to work similarly to the *Manual* technique. *P90* and *P111* asked if they can interact with the original in the solo tasks for better accuracy, instead of with its copy. *P801*, at one point, just forgot about the original dataset, not looking at it anymore. They summoned the copy and worked only on it, without paying attention to the rest of the environment. *P11*, *P1001*, *P110* did the same in the solo tasks. This specific behavior made **WIM** and **Manual** similar, which might explain why results reported in Sec. 5.5.2 may contradict **H4**.

*P001*, *P601*, *P91*, and *P101* just used only the original dataset after few trials. They reported that the existence of two coordinate systems (one per HMD) was too obvious when manipulating the copy and that the 3D cursor was too

big to be accurate. The original dataset, in comparison, is bigger and the hand tracking is not disturbed by the guider's pointing. When users looked at the original and not its copy, one may assume that users lose the directions axis (e.g., the left axis). However, guiders managed to guide their partners using these directions in an efficient way and without ambiguity, saying sometimes "closer to me." Strangely, *P11* performed half the trials using the original (sometimes starting with the copy and then heading toward the original) and the remaining trials only using the copy. Participants sometimes looked at both dataset copies simultaneously, which can be confusing as *P10* pointed out and may hinder deictic communication. *P00* and *P101* said that they cannot see their partner's pointing when manipulated the copy. In comparison, *P90* commented that he looked at both versions of the dataset to both target the point in the **WIM** and see his partner's pointing in the original dataset using the replicated 3D cursor. One idea would be to render the guider's 3D cursor to support pointing actions such as "just here."

## 5.6 Second Experiment—Solo Tasks

We are not aware of any previous comparisons of *Manual*, *Go-Go*, and **WIM** in the VR or the AR literature. To provide a more complete evaluation and understanding of these interaction techniques, we thus conducted a second experiment with the same participants to measure their performance in individual tasks. We asked each participant to anchor eight annotations per condition. However, only the person performing the action could now see the target, while the other only observed. We made this choice to get possible feedback from users about their understanding of their partner's actions. The roles switched after each trial. At the end of each condition, we asked participants to take a break and to answer the corresponding *raw NASA-TLX* questionnaire (Hart, 2006) to measure the users' cognitive load; see Fig. 8. By measuring how easy or not users can use the three techniques, this questionnaire is a good complement to the "social presence" questionnaire that measures the efficiency of the awareness cues of the techniques. The *raw NASA-TLX* is thoroughly tested and is widely used in HCI (Hart, 2006). This allows researchers to compare multiple related user studies. However, it does not assess the usability of a system (Longo, 2017).

While our main experiment studied collaboration, the pure technique performance from this second experiment makes our results comparable with exocentric metaphors with direct manipulations or with other ray+cursor techniques.

### 5.6.1 Pre-registered Results

We now analyze our questionnaire responses and logged data of this second experiment the same way as in Sec. 5.5. The computed pairwise comparisons are again stable and our

main conclusions remain unchanged regardless of outlier treatment, for both speed and accuracy.

**Accuracy.** Fig. 5b shows strong evidence that *Manual* is about 4× as accurate as *Go-Go*, *WIM* about 2× as accurate as *Go-Go*, and *Manual* about 2× as accurate as *WIM*. *AR Go-Go* is then the least accurate technique, supporting **H1**, and *Manual* the most accurate one, supporting **H5**. The results are also reflected in the ranking questionnaire results (Fig. 9).

Because of the mentioned gain factor issues, we saw participants  $P1_0$ ,  $P3_1$ ,  $P4_1$ ,  $P6_{0,1}$ ,  $P7_0$ ,  $P8_{0,1}$ ,  $P9_0$ ,  $P10_{0,1}$ , and  $P11_{0,1}$  going closer to the dataset to perform the tasks in the *AR Go-Go* condition, hoping for the technique to be more stable regarding the gesture, which may reduce the utility of its remote nature.  $P2_1$  proposed to use a better tracking that does not rely on a depth camera (e.g., using a sensor placed on the hand).  $P1_0$  and  $P11_0$  proposed to use another way to “click,” one that does not rely on the dominant hand, e.g., using a clicker in the non-dominant hand, yet this was used to hold the touch tablet. While this device was not mandatory for our experiments, we foresee that similar devices could be a prerequisite for collaborative 3D data exploration tasks as explained in Sec. 5.2.  $P6_0$ , by seeing that the Microsoft’s head-gaze cursor looked more robust, proposed to use it as the ray, while keeping the hand to adjust the position. Moreover,  $P0_1$ ,  $P7_1$ , and  $P9_{0,1}$  reported issues when determining the 3D cursor’s depth position.  $P0_1$  said that it was easier in the guidance tasks where his partner—who can better perceive the 3D cursor’s position—guides him. For the solo tasks,  $P0_1$  then relied on the target object’s occlusion as a depth cue.

**Task Completion Time (TCT).** Fig. 6b shows strong evidence that *AR Go-Go* is the slowest technique in solo tasks, supporting the first aspect of **H1**, but we found no strong evidence that *WIM* is the fastest, contradicting **H3**.

**Subjective Workload.** Fig. 8 shows strong evidence that *AR Go-Go* feels the least performant to participants, the most difficult and stressful, and the most physically and mentally demanding technique. Fig. 8 shows also weak evidence for *AR Go-Go* creating a greater time pressure compared to the other techniques. Finally, it shows weak evidence for *WIM* feeling less performant than *Manual*. We found no evidence regarding the stress, difficulty, time pressure, or physical and mental demand metrics between *WIM* and *Manual*.

## 5.7 Discussion

We now discuss the results extracted from logs, media recordings, and questionnaires.

### 5.7.1 Co-presence

While otherwise we did not find evidence for differences, the weak difference between *Manual* and *WIM* can be explained

by the fact that *WIM* creates another object of interest, reducing the possibilities of natural deictic communication. Overall, all techniques provide good presence among participants, likely due to them being in the same physical environment.

### 5.7.2 Message Understanding

The low self-estimated performance with *Go-Go* (Fig. 8) may influence the results of the “Message Understanding” metric. We do not know why there is weak evidence regarding the “co-presence” metric but not in the “message understanding” metric between *WIM* and *Manual*. Indeed, since some guiders did not always face their partner in *WIM*, a natural deictic communication was not always possible, which likely significantly reduced the mutual understanding provided by the technique in the same way the “co-presence” did.

### 5.7.3 Accuracy

**Manual.** One participant grabbed his partner’s arm to guide him to the correct spot in the guidance tasks. Both *Manual* and *WIM* permit such behavior, which can increase the accuracy and be beneficial for educational purposes or for disambiguation. This can be beneficial when co-workers are discussing about close, highly-relevant points (e.g., islands in a geographical dataset). By being the most accurate technique ( $\approx 1\text{cm} \pm 1\text{cm}$ ), *Manual* may be used for tasks which require precise tagging processes, e.g., tagging spatial features in scientific data. For a better accuracy, however, we think that users should be redirected their tablets, either to scale up the dataset, or to use other 2D interaction techniques (e.g., Benko and Feiner, 2007; Benko et al., 2006).

**AR Go-Go.** The Heisenberg effect (Bowman et al., 2001) amplified by the ray metaphor may explain why *AR Go-Go* is the least accurate technique. Even if the displacement is still acceptable ( $\approx 3\text{cm} \pm 1\text{cm}$ ), it creates frustration as  $P5_1$  pointed out. This may have strongly impacted the perception of the tracking accuracy reported in Sec. 5.4. Indeed, the technique was accurate enough to point at the target when adjusting, but the gesture degraded the accuracy, similarly to Looser et al.’s experiment (2007). This technique may then be good only for pointing, e.g., to show remote locations. While a better tracking system may reduce this issue, however, the accuracy of *AR Go-Go* will always be lower than that of *Manual* as the spatial jitter is converted into an orientational one. Moreover, it is known that orientational jitter affects users’ behavior and speed, even with high-quality tracking system such as the Lighthouse system of the HTC Vive (Batmaz and Stuerzlinger, 2019). Finally, this technique has two more limitations. First, controllers do not change shape when clicking compared to hands, impacting AR interaction techniques which tend to rely on hand tracking. Second, by playing with another degree of freedom compared to usual ray-casting

technique, the selected position is more susceptible to errors. Therefore, the same behavior should be expected for this interaction technique if another tracking system would have been used based on mid-air gestures to confirm positions. As we argue that a tablet may be a prerequisite for usual 2D interactions and is highly useful for exploratory tasks, however, we cannot rely on a person's second hand to confirm 3D cursor moved by mid-air gestures.

Moreover, users reported that *AR Go-Go* does not provide good depth perception. Indeed, the ray direction is a function of the head position, weakly supporting the motion parallax depth cue. Some users reported that they first tried to occlude the target before selecting it, giving them the needed depth information (occlusion cue). This confirms Cutting and Vishton's claim (1995) that monoscopic depth cues are more important than binocular ones. A ray origin not defined by the user's head, however, may improve the depth perception.

**WIM.** The accuracy of *WIM* is closer, in percentage, to that of *Manual* in the guidance tasks, compared to the solo tasks. In the former, guiders sometimes used the original dataset, so we hypothesize that the 3D cursor's size has a strong effect on accuracy. We expected *Manual* to be  $2 \times$  as accurate as *WIM* in both conditions as the dataset's copy was  $\frac{1}{2}$  its size, yet did not find such results in the guidance tasks. The accuracy being a function of the copy size, parameterization may then be needed for remote actions that need a precision close to that of *Manual*.

#### 5.7.4 Speed

The lack of strong evidence for *AR Go-Go* being slowest in collaboration may indicate (1) that the tasks's "guiding" step takes more time than the "clicking," (2) that the ray metaphor better supports direction instructions (Sec. 5.5.3), and (3) that depth perception is less essential for the actor than for the guider, the actor trusting and following the guider's instructions. Our observations, however, showed that participants took a lot of time trying to perform the click gesture, which the system had difficulties to recognize. Participants tended to reduce the click amplitude due to accuracy issue, but it became too low for the system to recognize the gesture. The lack of evidence for *AR Go-Go* being the slowest in such scenarios (e.g., education) partially contradicts the first part of **H1** but supports **H6**. The lack of strong evidence of differences between *WIM* and *Manual* is also reflected in people's rankings, where they marked both techniques as equally fast (see Fig. 9), not supporting **H3**. We emphasize that our tracking remained stable, with a mean accuracy of  $\approx 0.5$  cm for a TCT of  $\approx 10$  s for *Manual*, the direct interaction technique.

#### 5.7.5 Parallel Work (Hypothesis)

We asked participants to state their intuition about the parallel work potential of each technique. Because we did not

design the tasks to study such a scenario, we only report our hypothesis about this metric based on these responses and our observations. We think that *Manual* may not be suitable for parallelism because of workspace obstruction and tracking conflicts when both users would manipulate the environment, as *P6<sub>1</sub>* reported. In comparison, even if *AR Go-Go* does not obstruct the view, it is not accurate enough for parallel tasks which usually require precise input. Thus, most participants stated that *WIM* may allow them to work in parallel as it does not obstruct the view, while it has enough accuracy. Nonetheless, questions remain about whether the appearance and disappearance of the *WIM* disturb participants, and whether the 3D cursor should also appear inside the original dataset.

#### 5.7.6 Performance and Fatigue

We hypothesize that by being the slowest and least accurate technique, *AR Go-Go* is the most stressful, time-, mentally-, and physically-demanding technique. Participants had to pay attention to their gesture amplitude, which reduced the gesture recognition efficiency, inducing participants to repeat multiple times the gesture to correctly place the 3D cursor. Moreover, the hand movements may not feel natural as pointed out by *P8<sub>0</sub>*. In contrast, we hypothesize now that *WIM* is the least physically demanding technique because users do not need to walk to the dataset of interest, and that the copy version is, by design (see Sec. 4.3), smaller and generally placed at a comfortable 3D position, which reduce arm fatigue. *P6<sub>0,1</sub>* and *P9<sub>0</sub>*, however, wanted to modify the copy's position by dragging its edges, and *P9<sub>0</sub>* also wanted to parameterize the *WIM* size for more flexibility. Some other participants sometimes just stepped back because the copy was created "too close" to them, which can be frustrating. The *Manual* technique, however, was preferred and users felt more performant (see Fig. 8) with it, which supports **H2**.

Some participants in *Manual* put down their tablets on a desk for the solo tasks, as they rarely interfered with the tracking system and were not required for the interactions the tasks demanded. However, no participants reported any fatigue regarding their use.

## 6 Limitations

We asked our participants to use a tablet. While it was not fundamental for our tasks, we do not envision a complete explorative tool to rely solely on mid-air gestures and speech interfaces. As past work stated (Wang et al., 2020; Surale et al., 2019; Bornik et al., 2006), a hybrid interface is more likely to be accepted in a real use-case scenario by scientists than a standalone AR-HMD. By enforcing the use of a tablet, we thus saw the effects of such a tool used in a more comprehensive system that requires multiple and flexible forms of explorative tools. Yet, carrying a tablet may have affected the

performance of the studied techniques in isolation by adding a constant delay to start an anchoring process and being a cause of fatigue. However, no participant reported to be tired by the device, and we can assume the delay to be constant for all conditions. We thus do not expect our results, based on pair-wise comparisons, to change without the use of tablets.

Moreover, the “Tap” gesture slightly modified the detected palm position. We used a low-pass filter to reduce the noise, yet it was still evident in *AR Go-Go* where the noise is multiplied by the gain factor. However, it is well known that ray casting techniques are not well suited for selecting small objects far from the user (Bowman and Hodges, 1997; Forsberg et al., 1996; Poupyrev et al., 1998). Moreover, occasionally the arm was detected instead of the hand during the gesture which induced huge distance errors. These errors are normally removed either through the participants noting that a miss-click occurred, or by outlier removal. They count for less than 5% of the total number of measures. The HoloLens gesture recognizer did also not detect all gestures or produced false-positive, modifying slightly the speed metric (which was particularly true for *AR Go-Go* where participants tried to stay still as much as possible) and generating frustration.

In addition, doing the guidance experiment as first experiment may have introduced noise in the corresponding data. We saw, however, that, while the collaborative part began with a supervised training, most improvements happened not due to repeated execution but due to changes in approach during the collaborative trials (e.g., what words are efficient). This order also allowed us to study accurately the individual condition where users are well trained to give insights to researchers not interested in collaborative scenarios.

Another possible limitation is that we did not train our participants for how to collaborate with others. We purposefully did not do that because, in real life, users often collaborate with changing partners. Our results then reflect such scenarios, and may not apply to people who frequently collaborate with the same partners, for longer periods of time.

Also, while we motivated our work with visualization scenarios, we did not use such a setting in our experiment. Our rather generic setting, however, allowed us to study the general interaction and is directly applicable to visualization work. Here, people need to specify arbitrary positions within point clouds or volumetric datasets, for example scientists who usually filter their datasets to find patterns, generally resulting in sparse visualizations.

Our parameter choices may also have affected the results. A smaller/bigger *WIM*, e.g., may significantly affect the collaboration, fatigue, and accuracy, while an *AR Go-Go* with a lower gain-factor may change people’s preferences. We chose our specific parameters to satisfy meaningful constraints, such as being able to see the *WIM* entirely and to have a suitable range for *AR Go-Go*.

Our specific choice of questionnaires also affected the

results because terms such as difficulty, mental/physical demand, or being rushed are subjective categories. Nonetheless, the questionnaires we used are common in HCI experiments, which allows one to compare our results with others.

Finally, some participants reported a small but visible mismatch of the shared coordinate system created by the HoloLens’ API. This may affect the measured awareness of *WIM* because the dataset replica was small, which makes the mismatch more perceivable.

## 7 Conclusion

In this work we adapted three 3D pointing techniques from the VR literature and studied them in an isolation and collaborative contexts using AR-HMDs as means for specifying points in scientific data. Compared to other work, we used a markerless tracking system which seems to be the method that standalone HMDs will continue to use in the foreseeable future. While our accuracy results are not surprising, we saw that participants do not behave in the same way with the remote virtual hand (*AR Go-Go*), the exocentric (*WIM*), and the local virtual hand (*Manual*) metaphors. Our results and discussion are thus of interest for designers who need to understand the consequences of human behaviors for each of these metaphors, as they may not be ideal for every scenario. *AR Go-Go* may be suitable for pointing-only tasks (i.e., no specification performed), the exocentric technique to parallelize work among collaborators (which still needs to be verified in a separate experiment), and the *Manual* technique for users working on the same objects. Indeed, users may not be able to use *Manual* for parallel work due to the optical hand tracking and potential body collisions. For non-parallel work, the sudden appearance of proxy objects may disturb collaborators who do not expect these. We then saw weak evidence that *Manual* provided a better co-presence and supported users’ comprehension better than *WIM*, even if the co-presence results are satisfying in all cases.

## References

- Adcock M, Feng D, Thomas B (2013) Visualization of off-surface 3D viewpoint locations in spatial augmented reality. In: Proc. SUI, ACM, New York, pp 1–8, DOI 10.1145/2491367.2491378
- Amrhein V, Trafimow D, Greenland S (2018) Inferential statistics as descriptive statistics: There is no replication crisis if we don’t expect replication. Am Stat DOI 10.1080/00031305.2018.1543137
- Amrhein V, Greenland S, McShane B (2019) Scientists rise up against statistical significance. Nature 567(7748):305–307, DOI 10.1038/d41586-019-00857-9

- Argelaguet F, Andujar C (2013) A survey of 3D object selection techniques for virtual environments. *Comput Graphics* 37(3):121–136, DOI 10.1016/j.cag.2012.12.003
- Baguley T (2009) Standardized or simple effect size: What should be reported? *Br J Psychol* 100(3):603–617, DOI 10.1348/000712608X377117
- Baloup M, Pietrzak T, Casiez G (2019) RayCursor: A 3D pointing facilitation technique based on raycasting. In: Proc. CHI, ACM, New York, pp 101:1–101:12, DOI 10.1145/3290605.3300331
- Batmaz AU, Stuerzlinger W (2019) Effects of 3D rotational jitter and selection methods on 3D pointing tasks. In: Proc. VR, IEEE CS, Los Alamitos, pp 1687–1692, DOI 10.1109/VR.2019.8798038
- Benford S, Greenhalgh C, Reynard G, Brown C, Koleva B (1998) Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Trans Comput-Hum Interact* 5(3):185–223, DOI 10.1145/292834.292836
- Benko H, Feiner S (2007) Balloon selection: A multi-finger technique for accurate low-fatigue 3D selection. In: Proc. 3DUI, IEEE CS, Los Alamitos, pp 79–86, DOI 10.1109/3DUI.2007.340778
- Benko H, Wilson AD, Baudisch P (2006) Precise selection techniques for multi-touch screens. In: Proc. CHI, ACM, New York, pp 1263–1272, DOI 10.1145/1124772.1124963
- Besançon L, Dragicevic P (2019) The continued prevalence of dichotomous inferences at CHI. In: Proc. CHI Extended Abstracts, ACM, New York, pp alt14:1–alt14:11, DOI 10.1145/3290607.3310432
- Besançon L, Ammi M, Isenberg T (2017a) Pressure-based gain factor control for mobile 3D interaction using locally-coupled devices. In: Proc. CHI, ACM, New York, pp 1831–1842, DOI 10.1145/3025453.3025890
- Besançon L, Issartel P, Ammi M, Isenberg T (2017b) Hybrid tactile/tangible interaction for 3D data exploration. *IEEE Trans Vis Comput Graphics* 23(1):881–890, DOI 10.1109/TVCG.2016.2599217
- Besançon L, Sereno M, Yu L, Ammi M, Isenberg T (2019) Hybrid touch/tangible spatial 3D data selection. *Comput Graphics Forum* 38(3):553–567, DOI 10.1111/cgf.13710
- Besançon L, Jansen Y, Cockburn A, Dragicevic P (2021) Definitely maybe: Hedges and boosters in the HCI literature. DOI 10.31219/osf.io/mjg7h
- Billinghurst M, Cordeil M, Bezerianos A, Margolis T (2018) Collaborative immersive analytics. In: Marriott K, Schreiber F, Dwyer T, Klein K, Henry RN, Itoh T, Stuerzlinger W, Thomas BH (eds) Immersive Analytics, Springer, Berlin, chap 8, pp 221–257, DOI 10.1007/978-3-030-01388-2\_8
- Bornik A, Beichel R, Kruijff E, Reitinger B, Schmalstieg D (2006) A hybrid user interface for manipulation of volumetric medical data. In: Proc. 3DUI, IEEE CS, Los Alamitos, pp 29–36, DOI 10.1109/VR.2006.8
- Bowman D, Wingrave C, Campbell J, Ly V (2001) Using pinch gloves™ for both natural and abstract interaction techniques in virtual environments. Tech. Rep. TR-01-23, Department of Computer Science (0106) Virginia Tech, Blacksburg, VA 24061, USA
- Bowman DA, Hodges LF (1997) An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proc. I3D, ACM, New York, pp 35–38, DOI 10.1145/253284.253301
- Bowman DA, Johnson DB, Hodges LF (1999) Testbed evaluation of virtual environment interaction techniques. In: Proc. VRST, ACM, New York, pp 26—33, DOI 10.1145/323663.323667
- Bruckner S, Isenberg T, Ropinski T, Wiebel A (2019) A model of spatial directness in interactive visualization. *IEEE Trans Vis Comput Graphics* 25(8):2514–2528, DOI 10.1109/TVCG.2018.2848906
- Bruder G, Steinicke F, Stürzlinger W (2013) Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. In: Proc. 3DUI, IEEE CS, Los Alamitos, pp 115–118, DOI 10.1109/3DUI.2013.6550207
- Büscher W, Chen J, Dachselt R, Drucker S, Dwyer T, Görg C, Isenberg T, Kerren A, North C, Stuerzlinger W (2018) Interaction for immersive analytics. In: Marriott K, Schreiber F, Dwyer T, Klein K, Riche NH, Itoh T, Stuerzlinger W, Thomas BH (eds) Immersive Analytics, Springer, Berlin, chap 4, pp 95–138, DOI 10.1007/978-3-030-01388-2\_4
- Cockburn A, Gutwin C, Dix A (2018) Hark no more: On the preregistration of chi experiments. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ACM, New York, CHI ’18, p 1–12, DOI 10.1145/3173574.3173715
- Cockburn A, Dragicevic P, Besançon L, Gutwin C (2020) Threats of a replication crisis in empirical computer science. *Commun ACM* 63(8):70–79, DOI 10.1145/3360311
- Coe R (2002) It's the effect size, stupid: What effect size is and why it is important. In: Proc. BERA, London, pp 12:1–12:14
- Cumming G (2014) The new statistics: Why and how. *Psychol Sci* 25(1):7–29, DOI 10.1177/0956797613504966
- Cutting JE, Vishton PM (1995) Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: Perception of Space and Motion, Elsevier, pp 69–117, DOI 10.1016/B978-012240530-3/50005-5
- Czauderna T, Haga J, Kim J, Klapperstück M, Klein K, Kuhlen T, Oeltze-Jafra S, Sommer B, Schreiber F (2018) Immersive analytics applications in life and health sciences. In: Immersive Analytics, Springer, Cham, pp 289–330, DOI 10.1007/978-3-030-01388-2\_10
- Daiber F, Falk E, Krüger A (2012) Balloon selection revisited: Multi-touch selection techniques for stereoscopic data. In: Proc. AVI, ACM, New York, pp 441–444, DOI 10.1145/

- 2254556.2254641
- Dragicevic P (2016) Fair statistical communication in HCI. In: Robertson J, Kaptein M (eds) *Modern Statistical Methods for HCI*, Springer International Publishing, Cham, Switzerland, chap 13, pp 291–330, DOI 10.1007/978-3-319-26633-6\_13
- Dragicevic P, Chevalier F, Huot S (2014) Running an HCI experiment in multiple parallel universes. In: *CHI Extended Abstracts*, ACM, New York, pp 607–618, DOI 10.1145/2559206.2578881
- Fleming PJ, Wallace JJ (1986) How not to lie with statistics: The correct way to summarize benchmark results. *Communication of the ACM* 29(3):218–221, DOI 10.1145/5666.5673
- Forsberg A, Herndon K, Zeleznik R (1996) Aperture based selection for immersive virtual environments. In: *Proc. UIST*, ACM, New York, pp 95–96, DOI 10.1145/237091.237105
- Fu CW, Goh WB, Ng JA (2010) Multi-touch techniques for exploring large-scale 3D astrophysical simulations. In: *Proc. CHI*, ACM, New York, pp 2213–2222, DOI 10.1145/1753326.1753661
- Gigerenzer G (2018) Statistical rituals: The replication delusion and how we got there. *Adv Methods Pract Psychol Sci* p 2515245918771329, DOI 10.1177/2515245918771329
- Grossman T, Balakrishnan R (2005) The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In: *Proc. CHI*, ACM, New York, pp 281–290, DOI 10.1145/1054972.1055012
- Grossman T, Balakrishnan R (2006) The design and evaluation of selection techniques for 3D volumetric displays. In: *Proc. UIST*, ACM, New York, pp 3–12, DOI 10.1145/1166253.1166257
- Gupta K, Lee GA, Billinghurst M (2016) Do you see what I see? The effect of gaze tracking on task space remote collaboration. *IEEE Trans Vis Comput Graphics* 22(11):2413–2422, DOI 10.1109/TVCG.2016.2593778
- Gutwin C, Greenberg S (2002) A descriptive framework of workspace awareness for real-time groupware. *Comput Support Cooperat Work* 11(3):411–446, DOI 10.1023/A:1021271517844
- Harmon R, Patterson W, Ribarsky W, Bolter J (1996) The virtual annotation system. In: *Proc. VR*, IEEE CS, Los Alamitos, pp 239–245, DOI 10.1109/VRAIS.1996.490533
- Harms C, Biocca F (2004) Internal consistency and reliability of the networked minds measure of social presence. In: *Seventh Annual International Workshop: Presence*
- Hart SG (2006) Nasa-task load index (NASA-TLX); 20 years later. *Proc Hum Factors Ergon Soc Annu Meet* 50(9):904–908, DOI 10.1177/154193120605000909
- Helske J, Helske S, Cooper M, Ynnerman A, Besançon L (2021) Can visualization alleviate dichotomous thinking? effects of visual representations on the cliff effect. *IEEE Trans Vis Comput Graphics* 27(8):3397–3409, DOI 10.1109/TVCG.2021.3073466
- Hincapié-Ramos JD, Guo X, Moghadasian P, Irani P (2014) Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In: *Proc. CHI*, ACM, New York, pp 1063–1072, DOI 10.1145/2556288.2557130
- Hinckley K, Pausch R, Goble JC, Kassell NF (1994) Passive real-world interface props for neurosurgical visualization. In: *Proc. CHI*, ACM, New York, pp 452–458, DOI 10.1145/191666.191821
- Hurter C, Riche NH, Drucker SM, Cordeil M, Alligier R, Vuillemot R (2019) FiberClay: Sculpting three dimensional trajectories to reveal structural insights. *IEEE Trans Vis Comput Graphics* 25(1):704–714, DOI 10.1109/TVCG.2018.2865191
- Irlitti A, Smith RT, Von Itzstein S, Billinghurst M, Thomas BH (2016) Challenges for asynchronous collaboration in augmented reality. In: *Proc. ISMAR-Adjunct*, IEEE CS, Los Alamitos, pp 31–35, DOI 10.1109/ISMAR-Adjunct.2016.0032
- Isenberg T (2016) Interactive exploration of three-dimensional scientific visualizations on large display surfaces. In: Anslow C, Campos P, Jorge J (eds) *Collaboration Meets Interactive Spaces*, Springer, Berlin, chap 6, pp 97–123, DOI 10.1007/978-3-319-45853-3\_6
- Issartel P, Besançon L, Isenberg T, Ammi M (2016) A tangible volume for portable 3D interaction. In: *Proc. ISMAR*, IEEE CS, Los Alamitos, pp 215–220, DOI 10.1109/ISMAR-Adjunct.2016.0079
- Jackson B, Lau TY, Schroeder D, Toussaint KC, Keefe DF (2013) A lightweight tangible 3D interface for interactive visualization of thin fiber structures. *IEEE Trans Vis Comput Graphics* 19(12):2802–2809, DOI 10.1109/TVCG.2013.121
- Jansen Y, Dragicevic P, Fekete JD (2013) Evaluating the efficiency of physical visualizations. In: *Proc. CHI*, ACM, New York, pp 2593–2602, DOI 10.1145/2470654.2481359
- Johnson CR (2004) Top scientific visualization research problems. *IEEE Comput Graphics Appl* 24(4):13–17, DOI 10.1109/MCG.2004.20
- Keefe DF (2010) Integrating visualization and interaction research to improve scientific workflows. *IEEE Comput Graphics Appl* 30(2):8–13, DOI 10.1109/MCG.2010.30
- Keefe DF, Isenberg T (2013) Reimagining the scientific visualization interaction paradigm. *Computer* 46(5):51–57, DOI 10.1109/MC.2013.178
- Kim K, Billinghurst M, Bruder G, Duh HBL, Welch GF (2018) Revisiting trends in augmented reality research: A review of the 2<sup>nd</sup> decade of ISMAR (2008–2017). *IEEE Trans Vis Comput Graphics* 24(11):2947–2962, DOI 10.1109/TVCG.2018.2868591
- Klapperstuck M, Czauderna T, Goncu C, Glowacki J, Dwyer T, Schreiber F, Marriott K (2016) ContextuWall: Peer col-

- laboration using (large) displays. In: Proc. BDVA, IEEE CS, Los Alamitos, pp 7–14, DOI 10.1109/BDVA.2016.7787047
- Klein T, Guéniat F, Pastur L, Vernier F, Isenberg T (2012) A design study of direct-touch interaction for exploratory 3D scientific visualization. *Comput Graphics Forum* 31(3pt3):1225–1234, DOI 10.1111/j.1467-8659.2012.03115.x
- Krzywinski M, Altman N (2013) Points of significance: Error bars. *Nat Methods* 10(10):921–922, DOI 10.1038/nmeth.2659
- Le Goc M, Dragicevic P, Huron S, Boy J, Fekete JD (2016) A better grasp on pictures under glass: Comparing touch and tangible object manipulation using physical proxies. In: Proc. AVI, ACM, New York, pp 76–83, DOI 10.1145/2909132.2909247
- Lee GA, Teo T, Kim S, Billinghurst M (2018) A user study on MR remote collaboration using live 360 video. In: Proc. ISMAR, IEEE CS, Los Alamitos, pp 153–164, DOI 10.1109/ISMAR.2018.00051
- Lenne D, Thouvenin I, Aubry S (2009) Supporting design with 3D-annotations in a collaborative virtual environment. *Res Eng Des* 20(3):149–155, DOI 10.1007/s00163-009-0071-8
- Longo L (2017) Subjective usability, mental workload assessments and their impact on objective human performance. In: Proc. INTERACT, Springer International Publishing, Cham, pp 202–223
- Looser J, Billinghurst M, Grasset R, Cockburn A (2007) An evaluation of virtual lenses for object selection in augmented reality. In: Proc. GRAPHITE, ACM, New York, pp 203–210, DOI 10.1145/1321261.1321297
- Marriott K, Schreiber F, Dwyer T, Klein K, Henry RN, Itoh T, Stuerzlinger W, Thomas BH (2018) Immersive Analytics. Springer, Berlin, DOI 10.1007/978-3-030-01388-2
- McIntire JP, Havig PR, Geiselman EE (2014) Stereoscopic 3D displays and human performance: A comprehensive review. *Displays* 35(1):18–26, DOI 10.1016/j.displa.2013.10.004
- McShane BB, Gal D (2017) Statistical significance and the dichotomization of evidence. *J Am Stat Assoc* 112(519):885–895, DOI 10.1080/01621459.2017.1289846
- Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. *IEICE Trans Inf Syst* E77-D(12):1321–1329
- Nilsson S, Johansson B, Jonsson A (2009) Using AR to support cross-organisational collaboration in dynamic tasks. In: Proc. ISMAR, IEEE CS, Los Alamitos, pp 3–12, DOI 10.1109/ISMAR.2009.5336522
- Oda O, Feiner S (2012) 3D referencing techniques for physical objects in shared augmented reality. In: Proc. ISMAR, IEEE CS, Los Alamitos, pp 207–215, DOI 10.1109/ISMAR.2012.6402558
- Pierce JS, Stearns BC, Pausch R (1999) Voodoo dolls: Seamless interaction at multiple scales in virtual environments. In: Proc. I3D, ACM, New York, pp 141–145, DOI 10.1145/300523.300540
- Piumsomboon T, Dey A, Ens B, Lee G, Billinghurst M (2019a) The effects of sharing awareness cues in collaborative mixed reality. *Front Rob AI* 6, DOI 10.3389/frobt.2019.00005
- Piumsomboon T, Lee GA, Irlitti A, Ens B, Thomas BH, Billinghurst M (2019b) On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In: Proc. CHI, ACM, New York, pp 228:1–228:17, DOI 10.1145/3290605.3300458
- Poupyrev I, Billinghurst M, Weghorst S, Ichikawa T (1996) The Go-go interaction technique: Non-linear mapping for direct manipulation in VR. In: Proc. UIST, ACM, New York, pp 79–80, DOI 10.1145/237091.237102
- Poupyrev I, Ichikawa T, Weghorst S, Billinghurst M (1998) Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Comput Graphics Forum* 17(3):41–52, DOI 10.1111/1467-8659.00252
- Rheingans P (2002) Are we there yet? Exploring with dynamic visualization. *IEEE Comput Graphics Appl* 22(1):6–10, DOI 10.1109/38.974511
- Ro H, Chae S, Kim I, Byun J, Yang Y, Park Y, Han T (2017) A dynamic depth-variable ray-casting interface for object manipulation in AR environments. In: Proc. SMC, IEEE CS, Los Alamitos, pp 2873–2878, DOI 10.1109/SMC.2017.8123063
- Rosa DAW, Nagel HH (2010) Selection techniques for dense and occluded virtual 3D environments, supported by depth feedback: Double, bound and depth bubble cursors. In: Proc. SCCC, IEEE CS, Los Alamitos, pp 218–225, DOI 10.1109/SCCC.2010.51
- Saalfeld P, Glaßer S, Beuing O, Preim B (2017) The FAUST framework: Free-form annotations on unfolding vascular structures for treatment planning. *Comput Graphics* 65:12–21, DOI 10.1016/j.cag.2017.03.003
- Sauro J, Lewis JR (2010) Average task times in usability tests: What to report? In: Proc. CHI, ACM, New York, pp 2347–2350, DOI 10.1145/1753326.1753679
- Schmalstieg D, Fuhrmann A, Hesina G, Szalavári Z, Encarnaçāo LM, Gervautz M, Purgathofer W (2002) The *Studierstube* augmented reality project. *Presence: Teleoper Virtual Environ* 11(1):33–54, DOI 10.1162/105474602317343640
- Sereno M, Wang X, Besançon L, McGuffin MJ, Isenberg T (2021) Collaborative work in augmented reality: A survey. *IEEE Trans Vis Comput Graphics* DOI 10.1109/TVCG.2020.3032761, in press.
- Sherbondy A, Akers D, Mackenzie R, Dougherty R, Wandell B (2005) Exploring connectivity of the brain's white matter

- with dynamic queries. *IEEE Trans Vis Comput Graphics* 11(4):419–430, DOI 10.1109/TVCG.2005.59
- Springmeyer RR, Blattner MM, Max NL (1992) A characterization of the scientific data analysis process. In: Proc. VIS, IEEE CS, Los Alamitos, pp 235–242, DOI 10.1109/VISUAL.1992.235203
- Stoakley R, Conway MJ, Pausch R (1995) Virtual reality on a WIM: Interactive worlds in miniature. In: Proc. CHI, ACM, New York, pp 265–272, DOI 10.1145/223904.223938
- Surale HB, Gupta A, Hancock M, Vogel D (2019) Tablet-InVR: Exploring the design space for using a multi-touch tablet in virtual reality. In: Proc. CHI, ACM, New York, pp 13:1–13:13, DOI 10.1145/3290605.3300243
- Szalavári Z, Schmalsteig D, Fuhrmann A, Gervautz M (1998) “Studierstube”: An environment for collaboration in augmented reality. *Virtual Real* 3(1):37–48, DOI 10.1007/BF01409796
- Teo T, Lee GA, Billinghurst M, Adcock M (2018) Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In: Proc. OzCHI, ACM, New York, pp 406–410, DOI 10.1145/3292147.3292200
- Top A, Hamarneh G, Abugharbieh R (2011) Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3D image segmentation. In: Proc. MCV, Springer, Berlin, pp 204–213, DOI 10.1007/978-3-642-18421-5\_20
- Tory M, Möller T (2004) Human factors in visualization research. *IEEE Trans Vis Comput Graphics* 10(1):72–84, DOI 10.1109/TVCG.2004.1260759
- Valentine JC, Aloe AM, Lau TS (2015) Life after NHST: How to describe your data without “p-ing” everywhere. *Basic Appl Social Psychol* 37(5):260–273, DOI 10.1080/01973533.2015.1060240
- VandenBos GR (ed) (2009) Publication Manual of the American Psychological Association, 6th edn. American Psychological Association, Washington
- Wang X, Besançon L, Guéniat F, Sereno M, Ammi M, Isenberg T (2019) A vision of bringing immersive visualization to scientific workflows. In: Workshop CHI-IA, Glasgow, UK
- Wang X, Besançon L, Rousseau D, Sereno M, Ammi M, Isenberg T (2020) Towards an understanding of augmented reality extensions for existing 3D data analysis tools. In: Proc. CHI, ACM, New York, DOI 10.1145/3313831.3376657
- Yu L, Svetachov P, Isenberg P, Everts MH, Isenberg T (2010) FI3D: Direct-touch interaction for the exploration of 3D scientific visualization spaces. *IEEE Trans Vis Comput Graphics* 16(6):1613–1622, DOI 10.1109/TVCG.2010.157
- Yu L, Efstathiou K, Isenberg P, Isenberg T (2012) Efficient structure-aware selection techniques for 3D point cloud visualizations with 2DOF input. *IEEE Trans Vis Comput Graphics* 18(12):2245–2254, DOI 10.1109/TVCG.2012.217
- Yu L, Efstathiou K, Isenberg P, Isenberg T (2016) CAST: Effective and efficient user interaction for context-aware selection in 3D particle clouds. *IEEE Trans Vis Comput Graphics* 22(1):886–895, DOI 10.1109/TVCG.2015.2467202
- Zhou F, Duh HBL, Billinghurst M (2008) Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In: Proc. ISMAR, IEEE CS, Los Alamitos, pp 193–202, DOI 10.1109/ISMAR.2008.4637362