Collaborative Data Exploration and Discussion Supported by Augmented Reality

Exploration et Discussion Collaborative des Données Grâce à la Réalité Augmentée

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Informatique
Unité de recherche: Université Paris-Saclay, Inria, Inria Saclay-Île-de-France, 91120, Palaiseau, France
Référent: Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris-Saclay, le 14 Décembre 2021, par

Mickaël SERENO

Composition du jury:

Caroline APPERT	*Présidente*
Directrice de Recherche, Université Paris-Saclay, CNRS, Inria, LISN	
Tim DWYER	*Rapporteur & Examinateur*
Professeur, Monash University	
Michael SEDLMAIR	*Rapporteur & Examinateur*
Professeur, University of Stuttgart	
Ferran ARGELAGUET	*Examinateur*
Chargé de Recherche (CRCN), Inria Rennes	
Jeanne VÉZIEN	*Examinatrice*
Ingénieur de Recherche, LISN, CNRS, Université Paris-Saclay	

Direction de la thèse:

Tobias ISENBERG	*Directeur de thèse*
Directeur de recherche, Université Paris-Saclay, CNRS, Inria, LISN	

Thèse de doctorat

NNT: 2021UPASG106

SYNTHÈSE

J'étudie les avantages et limitations des casques de réalité augmentée (RA) pour l'exploration collaborative de données 3D. Avant d'entamer mes travaux, je voyais dans ces casques des avantages liés à leurs capacités immersives : ils fusionnent les espaces interactifs, de visualisation, de collaboration et physique des utilisateurs. Plusieurs collaborateurs peuvent voir et interagir directement avec des visuels 3D ancrés dans le monde réel.

Ces casques reposent sur une vision stéréoscopique 3D qui fournit une perception de profondeur accrue par rapport aux écrans 2D, aidant les utilisateurs à mieux comprendre leurs données 3D. Laissant les utilisateurs se voir les uns les autres, il est possible de transitionner sans effort d'une phase de discussion à une phase d'exploration. Ces casques permettent aux utilisateurs d'interagir au sein de l'espace de travail de manière directe, rapide et intuitive en 3D, et donnent des indices sur les intentions d'une personne aux autres. Par exemple, le fait de déplacer un objet en le saisissant est un indice fort sur les intentions de cette personne. Enfin, en n'occultant pas le monde réel, les outils habituels mais importants tels que les postes de travail et les cahiers de notes restent facilement accessibles dans cet environnement.

Cela étant, et bien qu'ils soient étudiés depuis des décennies, la puissance de calcul de ces casques avant la récente sortie de l'HoloLens en 2016 n'était pas suffisante pour une exploration efficace de données 3D telles que des données océaniques. De plus, les chercheurs précédemment étaient plus intéressés par *comment rendre la RA possible* que par *comment utiliser la RA*. Malgré toutes leurs qualités, il y a donc peu de travaux qui traitent de l'exploration de jeux de données 3D. Finalement, les casques de RA ne fournissent pas d'entrées 2D qui sont couramment utilisées avec les outils d'exploration actuels tels que ParaView et les logiciels de CAO, avec lesquels entre autres scientifiques et ingénieurs sont déjà efficaces.

Je théorise donc dans cette thèse les situations où ces casques sont préférables. Ils semblent préférables lorsque l'objectif est de partager des idées, d'explorer des modèles ensemble et lorsque les outils d'exploration peuvent être minimaux par rapport à ce que les postes de travail fournissent, et où la plupart des travaux et simulations préalables peuvent être effectués à l'avance. J'associe alors les casques de RA à des tablettes tactiles. J'utilise ces casques pour fusionner la visualisation, certaines interactions 3D et les espaces de collaboration dans l'espace physique des utilisateurs, et les tablettes pour la saisie 2D et l'interface utilisateur graphique habituelle que la plupart des logiciels proposent. J'étudie ensuite l'interaction de bas niveau nécessaire à l'exploration de données. Cela concerne la sélection de points et de régions dans des données 3D à l'aide de ce système hybride. Comme cette thèse vise à étudier les casques de RA dans des environnements collaboratifs, j'étudie également leurs capacités à adapter le visuel à chaque collaborateur pour un objet 3D ancré donné, similairement au "What-You-See-Is-What-I-See" relaxé qui permet par exemple à plusieurs utilisateurs de voir et modifier simultanément différentes parties d'un document partagé. Enfin, j'étudie en ce moment l'utilisation de mon système pour l'exploration collaborative en 3D des jeux de données océaniques sur lesquels travaillent mes collaborateurs du Helmholtz-Zentrum Geesthacht en Allemagne.

Pour résumer, cette thèse fournit un état de l'art de la RA à des fins collaboratifs, fournit un aperçu de l'impact de la directivité de l'interaction 3D sur l'exploration de donnée 3D, et donne aux concepteurs un aperçu de l'utilisation de la RA pour l'exploration collaborative de données scientifique 2D et 3D, en mettant l'accent sur le domaine océanographique.

ABSTRACT

I studied the benefits and limitations of Augmented Reality (AR) Head-Mounted Displays (AR-HMDs) for collaborative 3D data exploration. Prior to conducting any projects, I saw in AR-HMDs benefits concerning their immersive features: AR-HMDs merge the interactive, visualization, collaborative, and users' physical spaces together. Multiple collaborators can then see and interact directly with 3D visuals anchored within the users' physical space.

AR-HMDs usually rely on stereoscopic 3D displays which provide additional depth cues compared to 2D screens, supporting users at understanding 3D datasets better. As AR-HMDs allow users to see each other within the workspace, seamless switches between discussion and exploration phases are possible. Interacting within those visualizations allows for fast and intuitive 3D direct interactions, which yields cues about one's intentions to others, e. g., moving an object by grabbing it is a strong cue about what a person intends to do with that object. Those cues are important for everyone to understand what is currently going on. Finally, by not occluding the users' physical space, usual but important tools such as billboards and workstations performing simulations are still easily accessible within this environment without taking off the headsets.

That being said, and while AR-HMDs are being studied for decades, their computing power before the recent release of the HoloLens in 2016 was not enough for an efficient exploration of 3D data such as ocean datasets that my collaborators at Helmholtz-Zentrum Geesthacht, Germany, are working on. Moreover, previous researchers were more interested in *how to make AR possible* as opposed to *how to use AR*. Then, despite all those qualities one may think prior to working with AR-HMDs, there were almost no work that discusses the exploration of such 3D datasets. Moreover AR-HMDs are not suitable for 2D input which are commonly used within usual exploratory tools such as ParaView or CAD software, with which users, such as scientists and engineers, are already efficient. I then theorize in what situations are AR-HMDs preferable. They seem preferable when the purpose is to share insights with multiple collaborators and to explore patterns together, and where exploratory tools can be minimal compared to what workstations provide as most of the prior work and simulations can be done before hand. I am thus combining AR-HMDs with multi-touch tablets, where I use AR-HMDs to merge the visualizations, some 3D interactions, and the collaborative spaces within the users' physical space, and I use the tablets for 2D input and usual Graphical User Interfaces that most software provides (e. g., buttons and menus).

I then studied low-level interactions necessary for data exploration which concern the selection of points and regions inside datasets using this new hybrid system. As this PhD aims at studying AR-HMDs within collaborative environments, I also studied their capacities to adapt the visual to each collaborator for a given anchored 3D object. This is similar to the relaxed "What-You-See-Is-What-I-See" that allows, e. g., multiple users to see different parts of a shared document that remote users can edit simultaneously. Finally, I am currently (i. e., is not finished by the time I am writing this PhD) studying the use of this new system for the collaborative 3D data exploration of ocean datasets.

This PhD provides a state of the art of AR used within collaborative environments. It also gives insights about the impacts of 3D interaction directness for 3D data exploration. This PhD finally gives designers insights about the use of AR for collaborative scientific data exploration, with a focus on oceanography.

Publications

Accepted Papers:

1. **Mickael Sereno**, L. Besançon, and T. Isenberg, "Point specification in collaborative visualization for 3d scalar fields using augmented reality," *Virtual Reality*, pp. 1–19, 2021, to appear. doi: 10.1007/s10055-021-00614-2

2. **Mickael Sereno**, X. Wang, L. Besançon, M. J. Mcguffin, and T. Isenberg, "Collaborative work in augmented reality: A survey," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–20, 2020, in-press. doi: 10.1109/TVCG.2020.3032761

3. X. Wang, L. Besançon, D. Rousseau, **Mickael Sereno**, M. Ammi, and T. Isenberg, "Towards an understanding of augmented reality extensions for existing 3D data analysis tools," in *Proc. CHI*. New York: ACM, 2020, pp. 1–13. doi: 10.1145/3313831.3376657

4. L. Besançon, **Mickael Sereno**, M. Ammi, L. Yu, and T. Isenberg, "Hybrid touch/tangible spatial 3D data selection," *Computer Graphics Forum*, vol. 38, no. 3, pp. 553–567, Jun. 2019. doi: 10.1111/cgf.13710

Papers In Submission:

5. **Mickael Sereno**, S. Gosset, L. Besançon, and T. Isenberg, "Hybrid touch/tangible spatial selection in augmented reality," in *Proc. EuroVis*. Los Alamitos: IEEE, 2022, pp. 1–12, in submission

Workshop Papers:

6. X. Wang, L. Besançon, F. Guéniat, **Mickael Sereno**, M. Ammi, and T. Isenberg, "A vision of bringing immersive visualization to scientific workflows," in *Workshop CHI-IA*, Glasgow, UK, 2019

Extended Abstracts:

7. **Mickael Sereno** and T. Isenberg, "Subjective views in co-located augmented reality — initial design," in *Proc. Vis*. Los Alamitos: IEEE, 2020, extended Abstract

8. S. Gosset, **Mickael Sereno**, L. Besançon, and T. Isenberg, "Tangible volumetric brushing in augmented reality," in *Proc. Vis*. Los Alamitos: IEEE, 2020, extended Abstract

9. **Mickael Sereno**, L. Besançon, and T. Isenberg, "Supporting volumetric data visualization and analysis by combining augmented reality visuals with multi-touch input," *Computer Graphics Forum*, 2019, extended Abstract

ACKNOWLEDGEMENTS

Here comes the acknowledgements section, which is to me pretty hard to write as many people helped me directly or indirectly in this silly PhD project, and that I may not be really good on expressing feelings. For the beauty of it, and because those people deserve to be named, let me do this exercise and let me start with the professional aspects of this long series of acknowledgments.

This PhD, financed by Inria, started back in 2018 where I just finished to graduate from my engineering computer scientist school and my Human-Computer Interaction master. Without those diplomas, I would not get the opportunity to perform this PhD. Thanks a lot to all my previous teachers at my former high school, my engineering school Polytech Paris-Saclay, and the University of Paris-Saclay. They mostly taught me how to scientifically think, give me joy towards science in general, and provided me opportunities that I recognized and grabbed, and without whom this PhD would not even have started. I may not be good in writing and languages, but if I have the scientific background and mechanics, in terms of thinking, that were required to perform this PhD, it is mostly thanks to them.

I also thank specifically all my previous partners and teachers with whom I shared projects and with whom I had the opportunities to step in academia. I name Philippe Roncin and his colleagues at ISMO, Orsay, France, with whom I participated, during my 2nd under graduating year, in a research project to measure and mechanically fix the effects of the tilt angle between a surface normal and its rotation axis, that could not be manually aligned, during an atomic diffraction process in a ultra-high vacuum chamber [266]; Benjamin Köhler and his colleagues at the visualization group of Magdeburg University, Germany, with whom I worked in a research project, during my last year of under graduation, concerning the denoising of velocity vector fields of blood vessels captured by MRI scans [260]; and Lonni Besanoņ for whom I will provide a dedicated paragraph. Those were my first experiences in academia, which gave me taste for research in general. Those projects were also very useful to get my PhD grant.

Few moments before I got the grant, when people asked me if I would perform a PhD, I always answered no as I did not foresee myself into making a PhD. I could not be more wrong. I enjoyed most of my time at AVIZ (and maybe all of my time if the COVID-19 did not strike in), and I would like to thank each and every one in AVIZ. My first thoughts go to my supervisor Tobias Isenberg, who was always supportive and provided really good feedbacks along this journey. He also framed a PhD project that suited me and which convinced me to do this journey. I am also thinking about all the other permanents. Their scientific approaches, experiences, and good mood in general were more than welcomed. I name Petra Isenberg, always in a good mood and smiling, with whom I played some boardgames before the pandemic kicked-in; Pierre Dragicevic who helped me a lot about statistics and was always helpful in a kind manner; Jean-Daniel Fekete, the head of the team, who always helped as much as possible his team members; and Steve Haroz for some insights about open science. I am also thinking about all the PhD students I met at AVIZ. I name Xiyao Wang with whom I shared the topic of immersive visual analytics using Augmented Reality as a support; Sarkis Halladjian for his supportive attitude and his crazy singings; Marie Destandau with whom I shared only few months but was always smiling and positive; Alexis Pister and Mohammad Alaul Islam with whom I shared really good moments among the team, either by drinking, eating, or just passing time; Natkamon Tovanich with whom I spent a lot of time drinking, eating, especially Thaï food, but also with whom I spent really relaxing time in the middle of days; Jiyai Hong who, at the time I am writing this thesis, shared my office and was always supportive; and Lijie Yao and Yuheng Feng who, unfortunately, I did not had a lot of time to discuss with them due to the pandemic. I then would like to thank all the Post-Docs I met during those past three years. I name Paola Valdivia and Tanja Blascheck who were next to my office and were always supportive and with whom I shared few memorable drinks; Gaëlle Richer with whom I also played a lot of boardgames during my freetime and breaks. My next thoughts go the Stéphane Gosset, an intern at AVIZ under my supervision from March 2020 to September 2020. Supervising an inter gave me insights about my abilities to explain to others, supervise others, but also structure some of my work, as his internship was based on my software. I would also like to thank Raphael James and Eugenie Brasier, two PhD students near my office that work on Augmented Reality research questions with ILDA, Inria, France. We were able to share a lot of good moments but also scientific reflections and helps concerning Augmented Reality in itself as a field, or the HoloLens that we were using. Finally, my last thoughts for what concerns AVIZ go to Katia Evrat, our team assistant. Her dedication and work were absolutely welcomed, she managed all alone most of our administrative tasks with an astounding efficacity. Her smiling mood and efficacity have to be praised. What a hell would the french administrative system be without Katia. I thank you a lot.

A special thank also goes to Lonni Besançon, with whom I did my first Human-Computer Interaction research project with Tangible Brush [38, 261] while I was under-graduading and post-graduading. We shared multiple projects together, and while we do not always agree on some open-science questions, he had good advice and was really helpful. We did share some good drinking too, but I may have already emphasized too much that I drunk a lot!

I would like also to thank all the people that participated to the experiments this thesis is based on. Due to anonymity issue, I cannot name them, but I hope they will recognize themselves in this work. I would also like to thank all the jury members for reading and commenting on this thesis. Those comments, especially from my two reviewers, helped at enhancing this document to reach its best.

Next, I would like to give special thanks to my friends. They were supportive, always there when I needed them (thankfully for them, I did not need desperately to talk with people every time). The moments I shared discussion, walking, playing table-top roleplay games like Dungeons and Dragons and Call of Cthulhu, boardgames, video games, eating, drinking, sharing laughs, and doing sports (I can continue like this for a long moment) were not only welcomed moments, but they were also necessary and always cheered me up. Special thanks go to my daily group of table-top role players, which I name Mathieu Louvet, Tristan Hermant, Rémi Massart, Paul Meunier, Thibault Lasou, Pierre-Antoine Locatelli, Bryan Vigee, Vincent Protois, and Mélanie Monard. Special thanks go also to my friends from high school which I name Muriel De Paris, Cristophe Cardoso, Than Tao Ta, Yasmine Guy, and Mathias Mouchrik. I do not count how many foods and nights we shared together. Final thanks go to my usual drinking and boardgame group, which include my tabletop role playing group, but also Pierre Testart and Florent Volmer to whom I want to say thank you.

How can I continue this long series of acknowledgments without mentioning my family? Life was not always easy, with a lot of ups, and surely even more of downs. A huge, and that term is still not enough to express what I feel, thank goes to my mother Françoise Ribeiro, who managed, and I do not know how truth be told, to give me an environment just good enough for me to blossom. Then follow my brothers, Kévin Sereno and Danny Sereno, with whom I surely shared most part of my life and knew how to give me enough motivations and proudness in what I am currently doing. And finally, I name José Ribeiro, my stepfather, for the supports he gave to this family. Pour ma famille, qui ne lit pas anglais, je le redis : merci beaucoup[1].

And finally, I would like to thank Frédérique Thonon, with whom I share my life for more than a year. Sha gave me taste to a lot of things, including hiking, running, making me travel hundreds of kilometers for whole weeks using bicycles. While this sound like torture, it was really not. Going out while time has stopped was priceless. She was very supportive and gave

[1]For my family, who does not read english, I say it again: Thanks a lot.

me the final strengths I needed to wrap and finish this thesis. Moreover, what is better than an epidemiologist during a pandemic?

Merci beaucoup[2] for reading this thesis, and welcome into the AR multiverse!

[2] Thanks a lot

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

In this thesis, I am studying the benefits and limitations of Augmented Reality (AR) in the specific case of collaborative 3-Dimensional (3D) data visualization and exploration. Specifically, I am interested in datasets that contain an explicit spatial 3D structure. The visualization and the exploration of such datasets is part of the *scientific visualization* research area [55], which relies mostly on analytical tools.

Analytical tools support users, including scientists, at decision making and at understanding their data. Those tools mostly support exploratory tasks in which interactive data visualization is one key component. The research area of interactive data visualization is called *Visual Analytics* [269]. Such systems are important for many scientific domains and need to continuously be improved and adapted to the users' needs. Indeed, since the past years, users are dealing with datasets increasing in storage size. Those datasets can have an inherent spatial structure (*scientific* datasets) or not (*abstract* datasets). As those data cannot be understood just by looking at their raw values, data visualization aims at supporting humans at extracting, both qualitatively and quantitatively, patterns (e. g., the eye of a hurricane extracted from physical properties of the atmosphere such as temperature, wind speed, and pressure) and any trends those users are interested in their respective datasets using the human's visual system. As one visualization is unlikely to highlight all the patterns the users are interested in, those systems generally need to be highly interactive and configurable to let users **explore** their datasets.

However, there are multiple and diverse types of data, which make the creation of one unique solution almost impossible as the respective problems of those datasets might not be related at all. For example, some scientists might be interested, in their meteorological datasets, in how the flow of water or air evolves across time [154], whereas scientists studying dense social networks might be interested in clusters of related users for a specific property (e. g., clusters of co-authors for scientific publications [214]). In this PhD, I am focusing on scientific volumetric datasets, i. e., datasets that contain values at every point of a specific region in space, such as Magnetic Resonance Imaging (MRI) scans used in medicine, or meteorological simulations.

One typical example I am using for non-experts to understand the complexity of visualization concerns weather forecasts (see Figure 1.1). Using visualization projected on 2-Dimensional (2D) maps, people can understand without effort the weather conditions evolutions across time (e.g., sunny vs. rainy vs. windy) over a living area. As there are multiple properties to see, weather forecasts usually comprise multiple animated visualizations (e. g., precipitations and anticyclones representations). The difficulties come from merging different properties (e. g., pressure, velocity, temperature, humidity) for a single data point (e. g., the point above the Eiffel Tower, Paris, France) to derive a complex property (e. g., level of precipitations) and then to a color component (e. g., gradient of blues to know how much it will rain) that users would understand with a glance. Animations, if performed correctly, also ease users to understand how the raining situation will evolve across time. In contrast, non-domain expert users would not understand weather predictions by looking at the raw values of pressure, wind velocity,

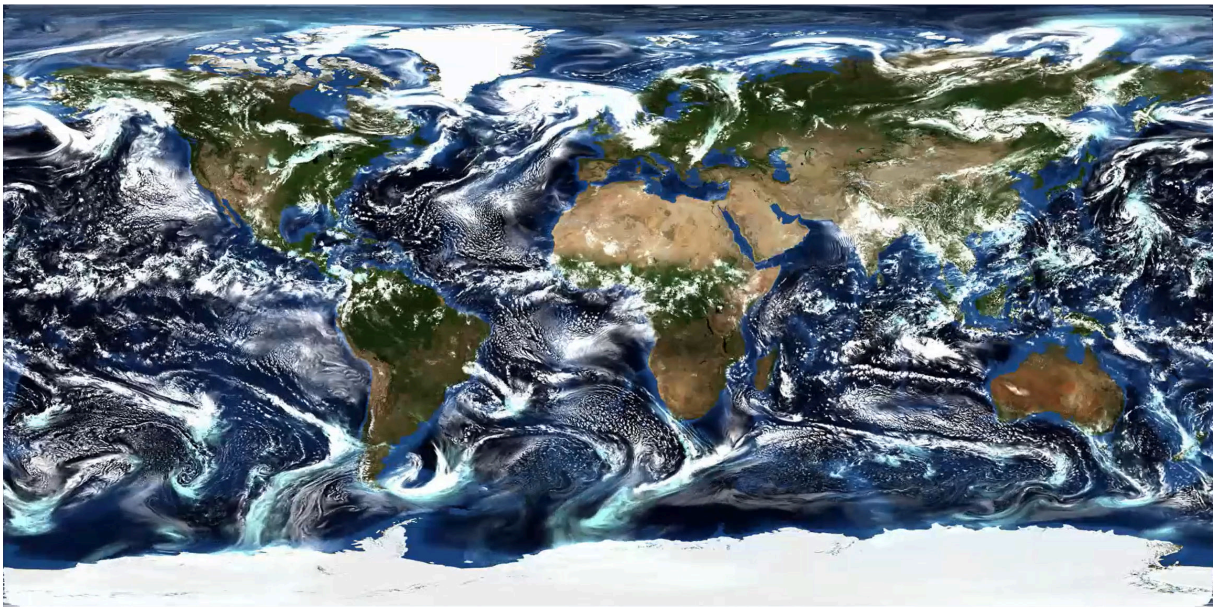


Figure 1.1: Screenshot of a meteorological forecast which visualizes clouds charged in rain and ice. Use with permission. ©DKRZ/MPI-M.

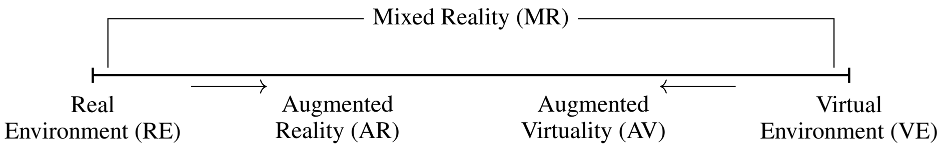


Figure 1.2: Milgram and Kishino's Reality-Virtuality Continuum [189].

temperature, and humidity, for every single measured point (weather station) of a region of interest. While this example is 2D, the scientific visualization of 3D datasets shares the same purpose about easing the understanding of scientific phenomena using the visual channel.

3D datasets, however, are generally much larger than 2D ones due to the additional depth axis. In this PhD, the 3D volumetric datasets that I am interested in contain millions of data points. Specifically, my collaborators at Helmholtz-Zentrum Geesthacht (HZG, Germany) work on low-resolution oceanographic datasets of the Cabo Verde area that comprise more than five million of data points per timestep, with each point comprising multiple physical properties of the ocean such as velocity, temperature, and salinity. By encapsulating multiple parameters, it is not rare that multiple people gather and discuss about the same datasets. Indeed, such datasets can interest multiple expertise (e. g., civil engineers, crisis agents, and meteorologists for possible hurricane events described in weather forecasts) and may involve different publics (e. g., one speaker against multiple listeners such as a teacher against students). The research area that focuses on the collaborative aspects of computer systems is called *Computer-Supported Collaborative Work (CSCW)* [1, 115, 231].

Before discussing about my main research question which concerns the benefits and limitations of AR for collaborative 3D data explorations, I would like to first explain why I considered immersive technologies, which some AR devices are part of, by looking at their properties.

1.1 Immersive Devices as 3D Data Exploratory Tools

But before talking about their properties, I first need to explain what an immersive device is. There are two main categories of immersive devices: Virtual Reality devices (VR) and Augmented Reality (AR) devices. Both can be explained in terms of a *Reality-Virtuality Continuum* (RVC,

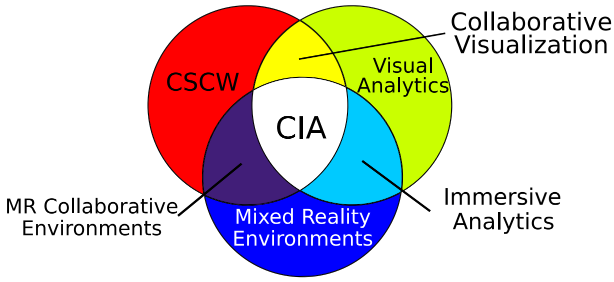


Figure 1.3: CIA relationship, adapted from Billinghurst et al.'s [42] diagram.

Figure 1.2) which covers different ways of mixing computed and physical objects. At one extreme is the real (physical) environment, and at the other is a purely Virtual Environment (VE) where Virtual Reality (VR) belongs. In a pure Virtual Reality environment, the user is totally detached from the real environment. However, this detachment would mean that the user can no longer receive feedback from the real environment, which also includes haptic and kinaesthetic feedback. As it seems impossible to remove kinaesthetic feedback for instance, researchers preferred to classify some devices (e. g., the CAVE [73]) as VR devices based on how "close" they transport the user into a new virtual environment. In contrast, Augmented Reality *augments* the real environment with virtual entities.

Mixed Reality (MR) systems are located between these two extremes. Users may receive feedback about objects through any sensory channel, most commonly visual, audio, and haptic. To better understand this continuum, Roo and Hachet [233] proposes to cut it into six navigable levels: the purely physical world, augmented surfaces (i. e., digital information placed on physical objects), mid-air digital contents, object decoupling (i. e., manipulating virtual objects representing real ones), body decoupling (i. e., virtually navigating in the real world), and a fully virtual world. Finally, systems that *remove* physical objects instead of *adding* virtual contents relate to the *Diminished Reality* [193] concept. AR is a special case of MR and is closely connected to the real world [15]. Azuma [15] and Azuma et al. [14] suggested the following requirements for AR systems: (1) **the merging and alignment** of real and virtual information, (2) **real-time rendering** through all the sensory channels, and (3) a **real-time interactive** environment.

The general concept of immersion relates to a psychological state users experience in an environment that continuously streams stimuli [292]. For MR systems, immersion also relates to the technological features of the devices (e. g., field-of-view, resolution, user's position and orientation tracking) [52], as those features use the most of the users' senses. As I am interested in visualization, I will use the latest definition and focus particularly on the graphical information.

This degree of immersion possesses multiple benefits for 3D visualizations. Indeed, those devices usually provide a stereoscopic screen and track the user's head which improves the motion parallax depth cue. By moving their heads, humans perceive closer objects moving, with respect to their retinal images, faster than objects placed far away. Both of those immersive features give a high degree of depth cues, i. e., passive information parsed by the human's visual system to mentally place in 3D the objects. Cutting and Vishton [77] categorized those depth cues as monocular or binocular. Monocular depth cues include *occlusion*, *aerial perspective* —the color shift of distant objects toward the color of the background (e. g., sky)—, the relative *size*, *density*, and *height* of objects in the user's retinal image, and *motion parallax*. Binocular depth cues include *binocular disparity*, which is the main concept of stereoscopic 3D screen, *convergence*, and *accommodation*. Depth cue conflicts hinder performance and comfort [173].

As I am interested in exploratory analytical tasks revolving around volumetric 3D data, the possible use of immersive devices came quickly to my attention for the stated benefits. The research area relying on immersive devices for visual analytics is called *Immersive Analytics*

Table 1.1: The presentation/placement awareness display techniques matrix [118]. *Literal* information is displayed as it is captured, and *Symbolic* information has been codified. *Situated* information is displayed in the workspace, and *Separate* information is displayed outside.

		Information Placement	
		Situated	Separate
Information Presentation	Literal		
	Symbolic		

(see Figure 1.3), which was first defined by Chandler et al. [65] as *"an emerging research thrust investigating how new interaction and display technologies can be used to support analytical reasoning and decision-making."* While multiple immersive technologies (e. g., Virtual Reality Head-Mounted Displays, AR Head-Mounted Displays (AR-HMDs), VR CAVEs [73], Spatial Augmented Reality systems [225]) exist, I am focusing in this thesis on AR Head-Mounted Displays for collaborative immersive analytics scenarios as I explain next.

1.1.1 Which Technologies for Collaborative Immersive Analytics?

Multiple devices allow for AR experiences; see also Section 2.2.3 for more details. Those devices mainly include Hand-Held Displays which rely on cameras (e. g., smartphones and multi-touch tablets), Spatial Augmented Reality [225] (SAR) systems which are composed of multiple cameras and projectors scanning the closed environment they are placed in and rendering on its surfaces, and finally Head-Mounted Displays. Moreover, AR devices can be coupled with additional devices to propose to users more complex interfaces.

In my PhD, I am relying on AR Head-Mounted Displays and multi-touch tablets to support each collaborator in their analytical tasks as I explain next.

AR Head-Mounted Display as the Main Device

All the stated AR devices do not propose the same kind of immersion and 3D depth perception. Hand-held devices are the least immersive ones. SAR are limited in the number of users they can track, notably because each projector needs to render one to two images for each user to correctly render a (stereoscopic) perspective viewpoint, which increases the needed refresh rate and graphical computing power. For example, for 4 users and a standard refresh rate of 60 Hz per user, projectors need to render the scene at $4 \times 2 \times 60 = 480Hz$ if stereoscopic rendering is needed, which is not yet commercially available. Due to those projectors, SAR have a high cost in maintenance. Head-Mounted Displays, however, are cheaper solutions compared to SAR systems, while having better immersion and handling an undetermined number of users.

Compared to VR headsets, e. g., the Oculus Rift, Oculus Quest, and the HTC Vive,[1] AR Head-Mounted Display technologies propose immersive experiences while merging the real and virtual worlds. This merging allows users to use their common analytical tools while visualizing in stereoscopy and using the "huge" (e. g., room-scale) workspace AR provides [283, 284].

CSCW applications rely a lot on communication, which AR-HMDs do not hinder since users can see each other directly and use their own tools, such as billboards or the combination of a pen and a piece of paper. While those tools seem rudimentary, it is not rare that co-workers need to take notes to consider for the next sessions and to structure the communication of the current work session using sketches. Moreover, by not occluding the real world, users can also see their collaborators without any artifact, allowing them to see facial expressions and rely on gestural communication. Communication then follows common social rules.

[1] When I am refering to VR headsets, I am refering to their VR capabilities and not their AR ones by, e. g., proposing video see-through AR experiences.

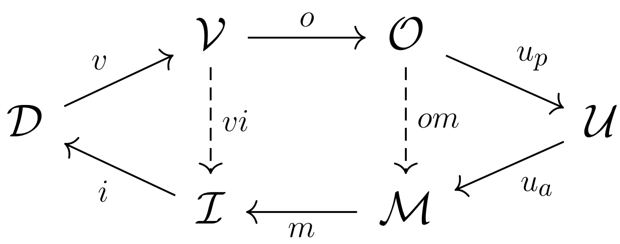


Figure 1.4: Bruckner et al.'s [57] model of spatial directness. $\mathcal{D}$: Data; $\mathcal{V}$: Visualization; $\mathcal{O}$: Output; $\mathcal{U}$: User; $\mathcal{M}$: Manipulation; $\mathcal{I}$: Interaction. Arrows represent transitions from one space to another.

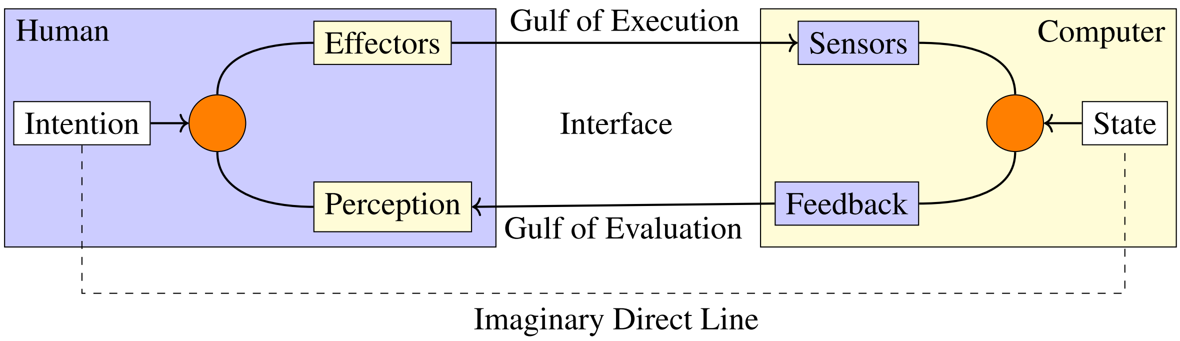


Figure 1.5: A closed loop control process based on Limerick et al. [171], itself based on Norman [197] and Williamson et al. [290]. Humans interact with the system based on what they perceive.

This leads us to the concept of *Workspace Awareness* (WA), which Gutwin and Greenberg [118] define as the *"knowledge created through interaction between an agent and its environment."* They also define Workspace Awareness as *"the up-to-the-moment understanding of another person's interaction with the shared workspace,"* which they derived from Adams et al.'s [2] *"up-to-the minute cognizance required to operate or maintain a system."* In sum, awareness cues allow users to understand and anticipate the interactions between their collaborators and the system. Those are powerful concepts that allow users to synchronize with their collaborators by knowing their intentions and actions.

Gutwin and Greenberg [118] described three sources of WA: bodies and consequential communication, workspace artifacts, and conversational elements (e. g., speech, gestures). They defined from these sources the *Presentation and Placement awareness display techniques* matrix (Table 1.1). Each awareness cue can be placed inside (*Situated*) the workplace, e. g., a 2D/3D cursor representing the current user's location, or outside (*Separate*) the workplace, e. g., it being accessible through a personal screen like a personal pad. Each awareness cue can also be rendered in the same way as captured (*Literal*) or explicitly (*Symbolic*), e. g., displaying the 2D/3D user's cursor as a 2D/3D spherical cursor as opposed to just highlight which graphical element the user's cursor currently is targeting at.

The situated-literal approach is, according to Gutwin and Greenberg [118], perhaps the most natural one and provides effective awareness information, with embodiment and expressive artifacts as two critical elements. They defined embodiment as *"the visual representation of a person's body in the workspace,"*, e. g., the real body, a video record, or a 3D avatar model representing the user with more or less awareness cues (e. g., gaze direction, positions of the limbs). Embodiment is a strong way to provide both consequential communication and conversational resources (e. g., orientation, position, gestural communication, and users' facial expressions) and is important in immersive collaborative environments. While VR immersive devices can render avatars to represent users within the collaborative space, AR devices such as AR-HMDs propose de-facto the strongest embodiment awareness cues as those devices do not occlude the real environment and, by extension, the collaborators.

Finally, AR-HMDs allow to place and interact with 3D visuals within the users' physical environment. I would like to emphasis that aspect using Bruckner et al.'s *"Model of Spatial*

Directness in Interactive Visualization" [57] (see Figure 1.4). This model shows us that, with immersive devices, users can reach the most direct form of interactivity when they interact within the same space as where the 3D visuals representing 3D datasets are. This most direct interactivity (categorized as Class 1 in Bruckner et al.'s [57] categorization[2]) might propose the experience with the lowest users' cognitive load ($\mathcal{U}$). I would like to consider in this model the place of collaboration. While I do not propose connection mappings as Bruckner et al. did between all their introduced sub-spaces, I would like to emphasize that, with AR-HMDs, all the interactive ($\mathcal{I}$ and $\mathcal{M}$), visualization ($\mathcal{V}$ and $\mathcal{O}$), and collaborative (which is not part of the model) spaces can be merged into one single 3D space which is the users' real environment. I did not adapt Burckner et al.'s model because, similarly to Norman's and Williamson et al.'s models (see Figure 1.5), it is based on an interaction loop control process: users interact with the system based on what they perceive. Brucker et al.'s model [57] can be understood as such: Manipulating the input device ($\mathcal{M}$) is interpreted by the system ($\mathcal{I}$) which modifies the state of the system ($\mathcal{D}$) and updates the visuals ($\mathcal{V}$) rendered by the output display ($\mathcal{O}$) that the user looks at and needs to understand ($\mathcal{U}$) before manipulating again the system.

However, while I am not adapting this model, we can apply this model per user ($\mathcal{U}$) to see how collaboration relates to the system ($\mathcal{V}, \mathcal{O}, \mathcal{M}$, and $\mathcal{I}$). Indeed, we can consider that the output display ($\mathcal{O}$) is not the AR-HMDs but the physical environment where users are (collaborative space), as users perceive the visualizations ($\mathcal{V}$) inside this physical environment, and that the manipulation space ($\mathcal{M}$) is tightly linked to it which can lead to 1:1 interaction ($\mathcal{I}$).

By merging potentially all those different spaces ($\mathcal{I}, \mathcal{M}, \mathcal{V}, \mathcal{O}$, and the collaborative space), AR-HMDs then allow users to see each other without artifacts and to interact, at least partially, within their mixed environment as they would do with the real environment, e. g., pointing towards objects, moving closer to see details and moving back to get overviews, etc. AR-HMDs also propose immersive technological features useful for 3D data exploration and visualization.

This most possible direct approach, however, does not necessarily lead to better performance and productivity for immersive technologies compared to traditional workstations [16]. McIntire et al. [186] examined 184 experimental comparisons of Stereoscopic 3D Displays (S3D) to non-stereo ones. They classified those comparisons into 6 categories: *Judgment of position and/or distances*; *Finding/identifying/classifying objects*; *Real/virtual spatial manipulations of objects*; *Navigation*; *Spatial understanding, memory, recall*; and *Learning/training/planning* tasks. In each of the first three categories, 57% or more of the experiments found that a S3D yields better performance. In each of the other three categories, between 36% to 52% experiments found a definite benefit with S3D. But only rarely did S3Ds perform worse than non-stereo screens. One explanation of AR not always performing better than traditional tools might come from that users are more used to, and have more experience with, those traditional tools which mostly rely on mouse+keyboard interactions.

As traditional tools are powerful for multiple tasks, I propose to rely for immersive analytics applications on a hybrid approach to not hinder users' performance. Specifically, for CIA applications relying on AR-HMDs, I propose to augment each AR-HMD with a dedicated multi-touch tablet as I explain next.

A Hybrid Approach

While AR-HMDs provide stereoscopic perspective views parameterized based on the human anatomy, which is proven useful for visualizing scientific datasets relying on 3D spatial components, these devices generally rely only on speech, mid-air gestures, and eye-gaze interactions. This is limiting considering the high number of tools and functionalities scientists need in their

[2]Bruckner et al. [57] classified interactions as belonging to Classes 1–6, where Class 1 interactions are the most and Class 6 ones the least "direct."

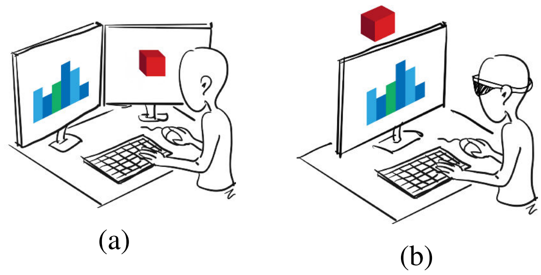


Figure 1.6: Wang et al.'s proposition [284] to transition from a traditional workstation environment (a) to an AR-HMD hybrid setup (b) that can seamlessly be used by scientists such as particle physicists.

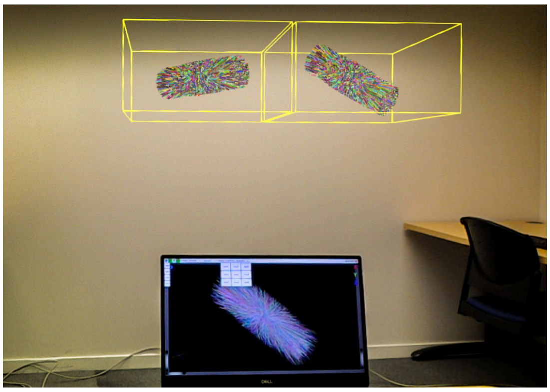


Figure 1.7: Screenshot showing a particle physicist using the hybrid setup Wang et al. [284] developed. The laptop handles all the interaction with its keyboard+mouse interaction paradigms, and the AR-HMD provides 3D visualizations of paths generated by particle collisions.

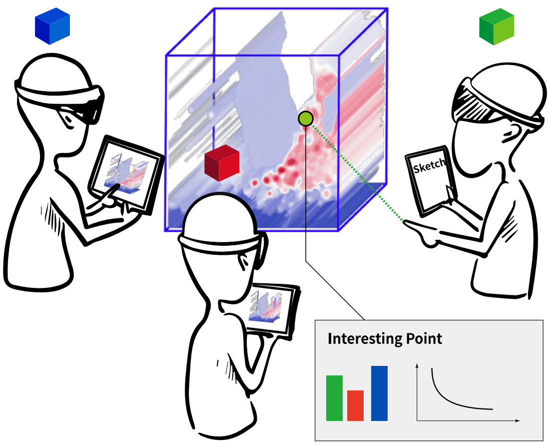


Figure 1.8: Sketch resuming my designed setup for immersive collaborative analytics environments relying on AR-HMDs. Each user wears both an AR-HMD and a multi-touch tablet to (1) see in stereoscopy the 3D visuals, and (2) interact with ease with the visualizations. Each user can see each other without artifacts. The 3D glyphs above each user represent their identity (color) and current action (shape). Those glyphs aim at giving awareness cues to collaborators about the activity of each other.

exploratory tasks, e. g., browsing the transfer function domain, filtering the values, looking at different values, rotating, scaling, translating, adding annotations, performing side-by-side comparisons, browsing the history of changes, adding additional metadata, etc.

In an ACM Human Factors in Computing Systems (ACM CHI) workshop paper I co-authored with Xiyao Wang, Lonni Besançon, Florimond Guéniat, Mehdi Ammi, and Tobias Isenberg [283], we envisioned that immersive technologies should enhance but not replace traditional tools for scientific workflows, mostly because some necessary tools (e. g., scripting tools such as MATLAB and Python) are not available in, or are hard to transpose to, immersive 3D spaces [283, 284] (see Figure 1.6). Later, with Xiyao Wang, Lonni Besançon, David Rousseau, Mehdi Ammi, and Tobias Isenberg [284], we studied how particle physicists might use an AR-HMD as an extension of a workstation; see Figure 1.7. As it is not rare for scientists to work currently with two 2D screens put side-by-side, and since AR-HMDs do not occult the real environment, we replaced one of those 2D screens with a 3D AR view using an AR-HMD. This study showed that, while each device might need their own dedicated input devices, this particular combination of AR+workstation would enhance, without replacing, the current workflow of those scientists.

However, we focused in this project on single-user environments [284] which was the main focus of my colleague Xiyao Wang in his thesis [281]. As I am focusing on collaborative tasks involving multiple users, I cannot rely on workstations as this would break the merging between the users' analytical space (which combines $\mathcal{I}$, $\mathcal{M}$, $\mathcal{V}$, and $\mathcal{O}$) and the collaborative space.

Indeed, users will require active focus transitions between the 2D screen of the workstation and the collaborators, as users (1) cannot see both their collaborators and the screen simultaneously and (2) as usually workstations propose only a limited set of input devices (usually keyboards and mice) that are not meant to be used by multiple users simultaneously. Relying on multiple workstations would possibly solve the issue related to the input devices, but it would also create as many *views* inside the virtual world the workstations are rendering, would not solve the issue about the transitions of focus, and will not encourage users to move around as they need to stay in front of their respective workstation.

Indeed, with Wang et al. [284], we saw that users would like to move around their visuals and spend more time at understanding what they see. I then enhance, in this PhD, AR-HMDs no with workstations but with multi-touch tablets as additional interactive devices as this would (1) give usual graphical user interfaces (GUI) users are accustomed to, (2) give tangible interactivity such as using the tablets as cutting planes or brushing devices [252] (see also Chapter 5) which would then act as 3D dedicated interactive devices for the AR environment where mouse+keyboard might not be the best, and (3) ensure mobility and keep this merging of all the stated spaces when users focus on the views the AR-HMDs render. Moreover, as interacting in 3D using hand gestures induce fast fatigue (see the *Gorilla-arm* syndrome [128, 142]), users can instead interact with the touch capabilities of the tablet which can be put at a more comfortable position (e. g., can be lowered), limiting the effort put on the users' shoulders (which depends on how high the users' arms are lifted). Still, if 3D input are necessary, a multi-touch tablet can be used as a tangible device or be held in one hand while the other hand is used for 3D mid-air gesture interactions. I summarize the basic idea of the system in its whole in Figure 1.8.

1.2 Research Questions and Structure of this PhD

While I am interested in collaborative immersive analytics scenarios, there is a lack of fundamental research for AR immersive analytics scenarios, them being collaborative or not. Indeed, AR as a field can be traced back to the late 1960s (e. g., [253]) and the term was first used in the early 1990s by Caudell and Mizell [63]—former Boeing engineers seeking to display simple information (e. g., text) in the users' physical 3D space to enhance manufacturing processes. However, for what concerns immersive devices, the field mostly have focused on *how to make AR possible* and not on *how to use AR* until approximately the release of the Microsoft's HoloLens 1^{st} gen. in 2016. Moreover, the main academic conference focusing on AR, IEEE International Symposium on Mixed and Augmented Reality (ISMAR), has paid little attention to CSCW, with only $\approx 1.7\%$ of papers published between 2008 and 2017 that discuss collaborative work in MR or AR [148]. The intersection of AR and CSCW, which I call AR-CSCW, possesses its own challenges to overcome but has not been studied much by researchers. In this PhD then, I focus primarily on the basic computer-human interaction (exploration) and then on the visualization of scientific datasets (visualization) while keeping collaborative aspects in perspective.

Moreover, before 2016, AR-HMDs—the devices I am focusing on during this PhD—were not standardized and had not enough computing power to meet the high requirements of scientific visualization. This is, however, changing with the release of the Microsoft's HoloLens 2^{nd} gen. in 2019 priced at US$ 3500 which supports Unity as an accessible SDK. This headset possesses a SoC Snapdragon 850 which is similar to the SoC of the Samsung Galaxy Tab S4 (a Snapdragon 835) which is a high-end multi-touch tablet released in 2018. Based on past work, these new technologies allow for fast and precise anchoring of virtual objects in the real environment, and embed a computing power comparable to recent high-end smartphones and multi-touch tablets which can handle scientific visualizations, yet on a smaller scale compared to desktop machines.

Because of the lack of strong literature, and because there were recent technological breakthroughs, I decided, in collaboration with Xiyao Wang, Lonni Besançon, Michael J. McGuffin, and Tobias Isenberg, to perform a survey on the field of AR-CSCW to better understand its current status and the multiple strong challenges to tackle. This survey was accepted and is accessible in early-access at the IEEE Transaction on Visualization and Computer Graphics (TVCG) journal [267] since 2020. We first surveyed work on AR-CSCW in general, and then gave a discussion about the special case of collaborative immersive analytics using AR as a support. With a deeper understanding of the AR-CSCW field, we saw, similarly to Fonnet et al. [99], that few work studied immersive analytics using AR as a support, should collaboration be involved or not.

After we have performed the survey that was influenced by my knowledge on the topic and the purposes of this PhD, I chose to pursue multiple projects in such a way that I tackle multiple fundamental aspects of AR-HMDs for scientific visualizations to give general and fundamental insights to readers and researchers interested in this topic.

To study the limitations and the benefits of AR-HMDs for the collaborative exploration of 3D data—my PhD topic—I created a coherent immersive analytics environment. While I developed this environment, I always kept in mind a real use case scenario and its possible uses in industry and production, in order to have the most impact as possible. This requirement is also shared with my former colleague Xiyao Wang's PhD research project which had a similar topic [281, 284]. Hence, all the functionalities I developed for the different research projects I pursued are available simultaneously and in a coherent way.

This application relies on three main distinct components: multiple AR-HMDs (currently HoloLens, either first or second generation), multiple multi-touch tablets (currently Galaxy Tab S4), and a remote server which handles the communication between all the devices. Because current commercial AR-HMDs cannot track multi-touch tablets yet, I decided to also use an external tracking component (the VICON) to track the users' tablets to enable tangible interactions. This system leads to my first well-defined research question:

RQ1 Compared to standard existing approaches, what are the advantages and disadvantages of the combined use of multi-touch tablets and AR-HMDs in collaborative 3D data exploration scenarios?

However, as it is for my PhD topic, this research question is large and may not be answered directly. Since the research area at the junction of CSCW, AR, and scientific visualization is not well studied, I focused first in my thesis on low-level interactions users need in their (collaborative) analytical processes, in order to provide a foundation for future researchers to explore higher-level interactions and tasks. This is a personal choice and someone else with the same topic might have tackled it differently.

1.2.1 Low-Level Interactions: 3D Point Specifications and Volumetric Selections

As my first project, I focused on 3D point specifications (Chapter 4), i. e., the selection of any spatial point in the 3D space. This interaction allows for instance users, using mid-air gestures, to probe data at specific locations, to show this point to a collaborator, and to anchor objects such as drawn or textual 2D annotations. Using three main interaction techniques borrowed from the Virtual Reality (VR) literature, I studied, in collaboration with Lonni Besançon and Tobias Isenberg (see Chapter 4), those three interaction techniques both in collaborative and non-collaborative scenarios. Depending on the level of directness of each technique and the perceived awareness cues by collaborators, some techniques might be better than others for particular tasks. These lead to my second research question:

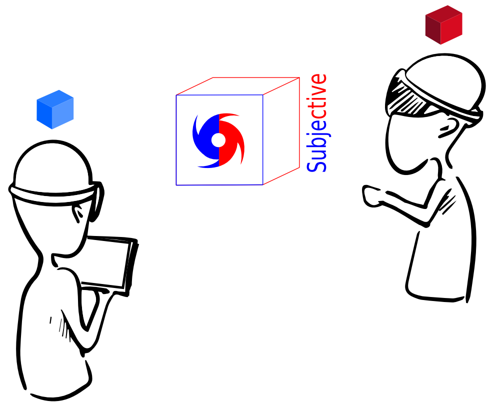


Figure 1.9: Concept of subjective views. User "blue" visualizes the 3D data computed with the blue visual parameters. User "red" sees this same spatial referenced (i. e., the object has the same 3D location/rotation/scaling) 3D data but the visual relies on the red visual parameters. Both users then do not see the same features inside the data but look at the data at the same 3D position, which can lead to a break inside the workspace coherency.

RQ2 To specify points in a volumetric dataset, what are the implications of the different available metaphors on the users' understanding, co-presence, and performance, in a collaborative AR environment?

I then focused on the selection of regions of interest based on a past work I co-authored [38]—Tangible Brush—which uses the combination of one 2D screen and a tangible multi-touch tablet (see Chapter 5). In Tangible Brush, the user draws a 2D lasso on the tablet that he or she extrudes by moving the tablet around, creating a 3D shape used to select a subset of the 3D space. The tablet, because the user needs to be accurate, displays the environment using an orthographic view. A second screen was then necessary to provide depth cues to users using perspective views. As this work showed promises (high accuracy and versatility), but showed limitations concerning the users' mental workload, I wanted to study the effects of placing the tangible multi-touch tablet inside an AR environment where its absolute 3D position now has meanings in the users' visualization, interactive, and physical spaces. Moreover, this new project also tackled the use of a multi-purpose tangible device as a dedicated 3D input that might give answers to the specific question Wang et al. [284] has raised concerning the dedicated 3D interfaces. These lead to the third research question which the survey also mentions (Section 2.7.4):

RQ3 To specify regions of interest, what are the implications of a tangible multi-touch tablet where its 3D position has meanings in the AR space compared to the original Tangible Brush which decouples the input and output spaces?

Because volumetric selection was not addressed for AR technologies to the best of my knowledge, and because there are more fundamental differences for tangible interactions compared to 3D point specifications between AR and VR, I focused on non-collaborative scenarios. At the end of my PhD, by looking back at those two projects, I realized that they both discuss the direct interactiveness (remote vs. direct 3D interaction) that AR-HMDs propose. I then discuss the implications of the use of tangible tablets for collaborative scenarios (Chapter 7), which leads to the fourth research question:

RQ4 As a side effect (post-hoc research question) of the interaction modalities I studied, what are the main benefits and limitations of direct interaction mappings compared to remote ones for one-user and collaborative environments?

1.2.2 Subjective Views: Same Objects, Same Positions, Different Visuals

For this PhD, I also wanted to dive further into one of the particular features of AR-HMDs I found interesting. It concerns their faculty to render custom contents per user while keeping co-workers in sight and working within the same 3D registered environment. This concept is called subjective views for immersive environment, i. e., the capacity of personal screen to alter the rendering of a given spatial object per user (Figure 1.9). This is similar to the relaxed "What-You-See-Is-What-I-See" concept that customizes collaborative objects and 2D screen rendering and layout based on users' roles and current status. For example, for shared document editing, users' viewports may diverge, as may the zoom or the layout of the application.

Multiple past researchers discuss about such possibilities in virtual environments [4,207,243] and showed some potential issues regarding workspace coherency. Notably, Smith and Mariani [243] proposed to classify the modifications along two dimensions: *appearance* and *modifier*. The first dimension considers the geometry of the objects, while the second one considers global modifications such as transparency and render mode. In the papers we surveyed in Chapter 2, only Nilsson et al.'s [196] and Prytz et al.'s systems [224], which focus on the collaboration between crisis agents (e. g., emergency, police, firemen), proposed subjective views to their users based on their expertise and roles that rely on different communication protocols.

However, compared to shared documents or communication protocols, the field of scientific visualization focuses on extracting, using visualization techniques and algorithms, specific features users can visually understand, e. g., an eye of a hurricane. For a given 3D dataset, multiple visualizations focusing on multiple properties might strongly diverge, as they, in addition to altering the color per data point, alter also the transparency of the data points (*modifier* dimension). Displaying a property to some users and not others might weaken the workspace coherency and lead to confusion between users, as one may see a feature at a 3D position which is perceived as being missing for another user. The selection of regions of interest (RQ3) affects as well strongly the visuals and should be considered as well in the *modifier* dimension. These lead to my final research question:

RQ5 What are the advantages and disadvantages of subjective views during the collaboration following the *modifier* and *appearance* dimensions for volumetric scientific datasets? Which interaction techniques and visualizations support them best with regard to the users' understanding, co-presence, and performance?

As this topic may depend on the type of data users are working on, my co-authors and I performed this project in collaboration with Helmholtz-Zentrum Geesthacht (HZG, Germany). This was also an opportunity to study a real use case scenario to give more insights to my main and fundamental research question (RQ1) that concerns the benefits and limitations of AR-HMDs for collaborative 3D data exploration. However, due to the COVID-19 pandemic, I can only give primary hypotheses, design considerations, and insights based on the comments of Burkard Baschek, my domain expert collaborator. To have strong insights, a study with multiple end-users is required.

1.2.3 Contributions

My contributions to the literature are then multiple. The survey gives newcomers a global picture of the AR-CSCW research area and is the entry point into this PhD (Chapter 2). The design considerations I made to create my own AR-CSCW system gives researchers in the field general insights concerning collaborative immersive analytics systems relying on AR-HMDs (Chapter 3). While I restricted myself to oceanography, it is worth noting that multiple other domains, e. g., biology and medicine, require the exploration and the visualization of volumetric 3D datasets. At this effect, two research projects were conducted regarding fundamental interactions that were barely studied in AR before to the best of my knowledge. Both projects focused on abstractions of real use cases to be as general as possible. The first project focuses on basic

3D hand gestures to specify 3D positions (Chapter 4; **RQ2**), and the second one focuses on the tangibility properties of the tablets to define regions of interest (Chapter 5; **RQ3**), while keeping collaborative tasks in sight. Finally, the concept of subjective views (Chapter 6; **RQ5**) might be one key component of AR Head-Mounted Display-based systems, as these technologies are the only technology I know that allow to merge all the users' collaborative, interactive, and visual spaces together while proposing adaptable contents for each of these users. I hope you will enjoy the reading of this immersive PhD as much as I immersed myself into. Next teleportation station: The State-of-the-art (STAR) hallway of the AR collaborative multiverse.

CHAPTER 2

SURVEY OF THE AR-CSCW RESEARCH AREA

This chapter is based on a survey I published at IEEE Transaction on Visualization and Computer Graphics [267] with my co-authors Xiyao Wang, Lonni Besançon, Michael J. McGuffin, and Tobias Isenberg. The survey concerns the AR-CSCW area. Any use of "we" refers to the aforementioned authors and concerns the survey side of this chapter. Since its first publication in 2020, I updated the survey in an opportunistic way by incorporating papers [60, 159, 227] published in 2020 and 2021 which discuss collaborative immersive analytics using AR as a support. I am also linking some parts of this survey with the projects I performed throughout this PhD, giving some piece of answers to previous research questions my co-authors and I asked. Any use of "I" therefore concerns either the part of the survey I updated, or the discussion about this PhD and its structure I am bringing along.

2.1 Introduction

Since the first introduction of AR technologies, researchers have first continued to improve AR hardware, algorithms (e. g., tracking algorithms), and then the user interfaces, with several communities showing interest in augmenting physical reality with virtual contents. AR has been applied in many disciplines, including medicine [23, 232], education [5, 210], archaeology [22, 28], games [211], remote expert guidance [92, 162, 200], industry [63, 68, 126, 204], crisis response [196, 220], information visualization [25, 60, 159, 227], and now scientific visualization (Chapter 3). Over the same period, smartphone applications (e. g., the Pokémon Go game) have increased the public's knowledge of, and interest in, Augmented Reality. But as stated in this thesis' introduction (Section 1.1.1), few work investigated the potential use of Augmented Reality in collaborative settings. However, collaborative settings play an important role among researchers and industrials. Within this work, we thus aim to provide both software designers and researchers with an up-to-date overview and a research agenda for the use of AR in CSCW contexts and a new taxonomy of AR-CSCW systems to (1) summarize what happened in the field, and (2) propose discussing about some research areas we found interesting to investigate.

We have systematically examined contributions from the IEEE ISMAR and ACM CSCW conferences between 2008 and 2019, the most important conferences respectively for AR and CSCW. We also included work from other venues that we found to be relevant in an opportunistic approach. This results in a total of 68 surveyed papers. As we come from the visualization and human-computer interaction communities, we wanted to analyze those work along their users' interfaces. We thus consider this aspect in our organization of these papers in a taxonomy that we based on the following dimensions: space, time, role symmetry (whether the roles of users are symmetric), technology symmetry (whether the hardware platforms of users are symmetric), and output and input devices.

Our work extends previous surveys related to AR-CSCW. Some of these surveys focused on more limited domains, such as education [5,210,293] or medical training [23]. Other surveys have discussed AR and CSCW in a more general way. In chronological order, these are Billinghurst and Kato [43], now 20 years old; Alem and Huang [8] and Billinghurst and Thomas [45], who focused on mobile collaborative work; Lukosch et al. [176], whose work still predates recent HMDs such as the Microsoft's HoloLens, and unlike the present survey, did not discuss psychological aspects of AR nor interaction techniques nor provide design guidelines; Irlitti et al. [134], who focused on asynchronous AR collaboration; Billinghurst et al. [42], who discussed the emerging field of Collaborative Immersive Analytics but without a specific focus on AR; and Ens et al. [94] whose taxonomy focused more on the purpose of the systems while we focus more on their output and input interfaces (see Section 2.3). In contrast to their taxonomy, we analyze different dimensions, examine different application areas—we discuss general CSCW topics and, in particular, the field of visualization—, and propose a summary of design considerations that we extracted from the surveyed papers. Looking at research focusing in AR and not on CSCW, Zhou et al. [300] and later Kim et al. [148] surveyed work presented at IEEE ISMAR, categorizing and analyzing papers by research area (e. g., tracking, displays).

The present chapter, and the article linked to it, are thus unique in surveying AR and CSCW, focusing in particular on user interfaces. This chapter is organized into three major parts. Section 2.2 reviews additional fundamentals of CSCW and AR, and can be skipped by practitioners and experts. The subsequent sections discuss the papers we selected for review, present design considerations, identify research areas to further investigate, and propose a discussion on visual analytics which we added to increase the relevance of the survey to the visualization community. This last part is strongly influenced by my PhD topic and embeds ideas and research questions I am pursuing in this PhD project.

2.2 CSCW concepts and Augmented Reality Technologies

When a Mixed Reality (MR) environment, which encapsulates AR, are coupled with CSCW features, we speak of a *Mixed-Reality Collaborative Environment*, with *Studierstube* [239, 240, 255] being among the first systems that were created. As my main field of research concerns Human-Computer Interaction and Visualization, and since we targeted the Visualization community, we thought it would be interesting to review, for our readership, the fundamental CSCW psychological aspects and what are their impacts on AR technological aspects. One should also note that the Introduction already gave a definition of AR (Section 1.1) and introduced the concept of Workspace Awareness (Section 1.1.1), where both are core concepts of this thesis report. As previously stated, this section targets newcomers to this field of research.

2.2.1 Psychological Aspects

In AR systems, users may feel that virtual objects are transported to them or that they are transported to a remote place. With users experiencing a 3D augmented world mimicking the real world, designers can improve the users' sense of presence, and their performance [292], by considering how users behave in real life for a particular task. Still, past research [196] has shown that the choice of devices and interfaces influences the user's level of engagement, immersion, and presence, which in turn influences their performance. We therefore discuss the concepts of *presence and immersion* as well as *engagement*.

Presence and Immersion

Presence is a matter of focus that users experience when they feel involved and immersed in their tasks [100, 292]. Witmer and Singer [292] found a weak but consistent positive correlation between presence and performance. They theorized several factors contributing to the sense of presence:

Control Factors refer to ways in which users can control the system and environment. Interactivity should happen in real-time and users should be able to foresee system feedback. Yuill and Roger [299] similarly identified controllability as important for the success of multi-user systems.

Sensory Factors refer to the richness of feedback to the user's senses. Inconsistent stimuli engender illness, called *cybersickness* [160].

Distraction Factors are determined by the user's isolation. Presence increases with greater isolation.

Realism Factors are determined by the realism and the consistency of the environment.

Witmer and Singer also defined immersion as a psychological state users experience in an environment that continuously streams stimuli. For MR environments, their concept of immersion relates to the technological features (e. g., field-of-view, FoV) [52] which positively influence (medium-size effect) the users' sense of presence [76]. Slater [242] defined *immersion* and *place illusion* as a set of *sensorimotor contingencies* that approximate those of physical reality, and the *valid actions* users can perform in their Virtual Environment (VE). He also introduced the concept of *plausibility illusion* as *"the illusion that the scenario being depicted is actually occurring,"* which relates to the realism factor. Immersion is usually greater in MR environments than in regular workstations, as long as users do not experience *cybersickness* [160].

Finally, presence can be categorized into three categories:

Physical presence A cognitive state in which users experience the virtual stimuli as real [167]. Strong physical presence means the MR environment is expected to exhibit plausible physics, with the user's imagination often interpolating incomplete stimuli. In AR environments, physical presence is related to the *object presence* which refers to how realistically a virtual object is anchored in the real environment [145, 251].

Social presence *"The sense of being with another"* [47] is one of the main concepts in remote AR collaboration. According to Biocca et al. [47], it mainly influences the productivity of collaborative environments.

Self presence A cognitive state where the virtual self is experienced as the real self [167]. This relates more to virtual environments where virtual avatars represent users.

As it is for traditional relaxed What-You-See-Is-What-I-See CSCW systems, AR and VR CSCW applications can display custom content per user (albeit of their viewpoint) because users wear their own rendering system. These *subjective views* [4] (see also Chapter 6) can then alter the users' social presence if the workspace loses its spatial coherency [207]. Finally, the use of AR, as compared to traditional tools, can improve user performance and learnability [210] by engaging its users more. This leads to the concept of *engagement* as discussed next.

Engagement

Attfield et al. [13] defined engagement as *"the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource."* Potentially, this produces a state of *flow* [74] which is always enjoyable.

O'Brien and Toms [198] characterized the level of engagement of an application by its *aesthetic appeal, attention, challenge, endurability, feedback, interactivity, perceived user control, pleasure, sensory appeal*, and *novelty*. AR tends to engage users more than traditional tools and sometimes to pure VEs. The control and richness of sensory feedback (e. g., stereoscopic vision, spatial audio), and the ubiquitous interfaces increasing both perceived user control and novelty, tend to be higher in MR systems than in regular workstations. Moreover, the virtual and real-world objects and tools are blended together, making focus transitions seamless [43]. Finally, social presence is higher in AR systems by being close to real-world rules and by preserving social protocols as compared to regular workstations and VEs [48, 122].

Scenarios with users not sharing the same role exist. Some displays (e. g., AR-HMDs, traditional 2D screens) may then be more suitable for some users than others, which relates to Isenberg et al.'s [136] CSCW visual applications categorization based on users' engagement:

Viewing occurs when a majority is spectator of a minority's actions. Traditional 2D screens might be considered for the spectators while the actors use AR devices or vice-versa.

Interacting/Exploring occurs when the group is exploring, discussing, and understanding the same data.

Sharing/Creating occurs when participants of the group can create, upload, and share new content. The creation can happen in personal rendering space (e. g., a personal multi-touch tablet) which can later be shared to a space accessible by all (e. g., a public AR 3D space).

The use of different render devices leads us to the Mixed-Space Collaborative Work concept.

2.2.2 Mixed-Space Collaborative Work

Mixed-space collaborative setups include any collaborations where users are in multiple shared-spaces, defined as a space placed at one point on the Reality-Virtuality Continuum.

After defining shared-spaces, Benford et al. [26] introduced Mixed-Space Boundaries, i. e., how one can see, view, or interact from one shared-space to another *adjacent* one which have common properties and purposes. Misunderstandings of two separate spaces (e. g., two remote users having each their own AR space) can hinder performance. For example, users should be aware of each other in distributed collaborations. Strong embodiments support this goal, as do awareness cues, e. g., augmented pointing actions or displaying a user's FoV to others. One common mixed-space boundary is between one *exocentric* space (i. e., view of the data from outside) and one *egocentric* space (i. e., view of the data from inside). Following Kiyokawa et al. [151], Shared Augmented Environments, which appear in an exocentric way, lead to richer awareness than Shared Virtual Environments (SVE) and are more suited for co-located work. In contrast, SVEs better support parallel activities, e. g., visualizing from different perspective (e. g., scales, positions) and can be used in an egocentric *private view*. However, if your co-workers can see you and do not understand what you are doing (e. g., manipulating your private view), they can be confused: how can one understand whether the other is manipulating the private or the public view? Benford et al. [26] characterized Shared-Space Technologies with:

Transportation How much are users transported to another place or vice-versa? While feeling not transported in co-located systems is usually expected, it may be a requirement for remote ones.

Artificiality What is the degree of artificiality of the shared-space? This can be related to the Reality-Virtuality Continuum (Figure 1.2).

Spatiality How is the user's shared-space technology spatially defined? Is its coordinate system shared? Can others determine this user's position and orientation relative to themselves?

The transportation and artificiality axes rely mostly on the system's output and input modalities, which we review in the next sections and include in our taxonomy.

2.2.3 Displays

Workspace Awareness (WA), an important research topic in AR-CSCW, can be impacted by the type (and positioning) of a display. We consider Head-Mounted Displays (HMDs), Hand-Held Displays (HHDs), and spatial displays anchored in the physical environment. We also consider optical see-through (OST) displays, video see-through (VST) displays, and projectors. Finally, we consider stereoscopic 3D displays (S3Ds). McIntire et al. [186] summarized, across 184 experiments, where and when S3Ds outperform traditional ones. They asserted that S3Ds improve performance when monocular depth cues are weak, the objects of interest are *close* for a significant binocular disparity, and when the tasks are *difficult* (see also Bowman and McMahan [52]). S3Ds better support novices than expert users. S3Ds rely on depth cues which Cutting and Vishton [77] categorized as monocular or binocular; see Section 1.1.

HMDs support all the technologies cited above and are mostly stereoscopy-enabled. VST displays usually have higher FoV than OST displays due to physical constraints but have lower refresh rate due to camera capture rate and latency, strongly influencing cyber-sickness. HHDs are generally based on a VST display as part of a smartphone or a multi-touch tablet. Raskar et al. [225] introduced Spatial Augmented Reality (SAR) as virtual content displayed within or on the user's physical space, which can contain complex colors and geometries [46]. They are based on VST stationary screens and/or room-space projector devices. In SAR settings, users do not wear any display devices, except for shutter glasses when stereoscopy is required. Without glasses, users can still have a sense of perspective [30] which can be coupled with stereoscopy with worn glasses [31]. Multiple users can be in the AR space using time-multiplexed shuttered glasses to display contents individually. However, the minimum refresh-rate of the displays (projectors or VST) when time-multiplexed shuttered glasses are used is linear with the number of users, limiting the collaborative environment to only few users.

Each display device has its own benefits and limitations. A SAR is non-intrusive but lacks mobility and may have a strong limit regarding the number of simultaneous users. HHDs are commonly found in the consumer market, with smartphones and tablets being ubiquitous, but are not stereoscopy-enabled. Head-Mounted Displays are stereoscopic enabled, but are usually embedded systems with limited computing power, which, compared to HHDs, render the scene twice due to the stereoscopy requirement and handle the tracking of the 3D environment. Color fidelity (e. g., black cannot be rendered with AR-HMDs) should also be considered. Regarding depth cues, designers should focus first on monocular ones and should avoid depth cue conflicts.

We next consider input modalities which the employed output modalities can influence.

2.2.4 Interaction Techniques

Data exploration relies on exploratory tools and thus by extension on interaction. Input modalities include touch, tangible, and mid-air gesture input. When collaboration is an important aspect, not every modality might be useful. For example, it may be hard to understand what one is doing on a private multi-touch tablet that is used as the input device.

Touch Devices

Touch devices like smartphones or tablets can be used in multiple ways. First, they provide a good way to interact with Graphical User Interface (GUI) components (e. g., buttons, menus) [168, 236]. They can be used as VST AR devices which provide seamless interaction in both the GUI and the AR views. However, the view is limited by finger occlusion during interaction and low FoV, especially for small devices like smartphones. Finally, they can be used as a complement to another AR display, such as an AR-HMD [252] (see also Chapter 3).

The *Studierstube* system [239, 240, 255] used a lightweight Personal Interaction Panel (PIP) to interact—via a WIMP (Window, Icon, Menu, Pointer) interface—with data visualized in AR-HMDs. The touch device can be used to draw graphical (e. g., [245]) or textual annotations, a key feature of their collaborative system. The 3D pen allows direct manipulation in the 3D space. López et al. [175] showed how a multi-touch tablet facilitates 3D data transformations visualized in a S3D. Büschel et al. [61] explored and studied multiple aspects of a tracked multi-touch device as a way to pan and zoom 3D datasets visualized through an AR-HMD.

The benefits of touch over other forms of interaction have been deeply studied for a variety of settings. The literature explains [57] why touch can be considered to be a direct form of interaction, usually improving speed at the cost of accuracy [7, 101]. Its speed advantage in 2D interfaces, however, does not seem to translate to 3D scenarios [37].

Tangible Interface

Tangible User Interfaces (TUIs) [139] augment the physical world by linking digital information with real objects. TUIs thus extend the graspable interface [97] concept, which uses physical objects to directly control digital objects. Most VR games use two hand-held controllers, which can serve as both classical controllers and tangible interfaces that represent, e. g., light sabers.[1] The buttons can be used as *triggers* of the virtual object to which the controller is attached. Tangible interaction facilitates fast and precise interaction [37, 66, 130] that mimic the real world [96, 139], foster collaboration [182, 203], and can provide entertainment [37, 295].

SARs allow users to directly manipulate physical objects. Piper et al.'s [213] *Illuminating Clay* allows users to interact with landscape data using AR and depth cameras on 3D landscape models. Wilson and Benko's [291] *LightSpace* is a luminous room that uses several depth cameras and projectors to track multiple users and let them tangibly interact with immobile rectangular augmented surfaces. Roo et al.'s [232] *Inner Garden* allows users to meditate by building a mini-world using a SAR tangible sandbox and exploring it with a VR-HMD. The SAR content changes based on the sand shape (e. g., display snow), with some effects (e. g., cloud shadows) derived from the user's breathing and heart rate.

In AR, users may need their hands free to, for instance, repair an aircraft engine. Henderson and Feiner [125] imagined a system where the AR device selects special physical elements to anchor Augmented Widgets (e. g., buttons). This design leaves the physical environment unmodified, while providing the user with passive haptic feedback and tangible interfaces. They called these interfaces *Opportunistic Control*.

Gestural Interfaces

Many forms of gestures exist, e. g., touch/pen gestures (i. e., drawing), mid-air gestures, eye gestures, and gestures using props (e. g., conducting). Gestures in general may not be correctly understood by the system, while mid-air gestures can also fatigue users (e. g., *Gorilla-arm* syndrome [128, 142]). In co-located collaborative setups where users use body language, mid-air gestures can be confusing: it is unclear if gestural actions are meant as system input or

[1] See the Beat Saber game, `https://beatsaber.com/`.

deictic communication with a collaborator [28]. In addition, techniques relying on virtual arm modifications (e. g., the *Go-Go* technique [222] which extends virtually the user's arm to remote select objects; see Chapter 4 for an AR adaptation) may not be a good fit for AR because users can see both their real and their virtual hands, introducing a perceptual conflict. Mid-air gestures, however, may be a requirement to provide users with a way to keep their hands free as it could be crucial in, e. g., sterile medical environments [89, 202]. They can usually be used with HMDs without additional input devices and may improve visual data exploration. Keefe and Isenberg identified challenges faced by using gestures in 3D visualization [147].

Hybrid Interfaces

Hybrid Interfaces use multiple kind of inputs. Bolt's [49] system, e. g., allows users to draw shapes at specific locations using both voice and gesture commands. Speech input can also be used to set the system mode (e. g., scaling, translating, rotating), while a gesture performs the action [178]. Touch devices can be used as tangible objects (e. g., 2D cutting planes [36, 245]). A tablet can be used to seamlessly grasp and release virtual objects. Surale et al. [252], e. g., proposed to use a tablet in VR space both as a touch device and as a tangible one. Their design space also includes mid-air gestures to interact with 3D content in a coarser way. The PIP discussed in Section 2.2.4 supports 3D interactions with AR data using a 6 DoF pen. Dedual et al.'s [81] hybrid interface associated a 2D multi-touch tabletop with an AR-HMD. The tabletop displayed footprints (2D projections) of 3D augmented objects that are visible through the HMD. Other collaborators can see the footprints and interact with them in the same coordinate system.

In general, hybrid interfaces are employed to propose complex interactions. Each interface then can fix the weakness of the others, e. g., using the accurate 2D input of a multi-touch tablet with efficient 3D movements [38], or use a non-AR multi-touch tablet for 2D input and an AR-HMD for 3D input; see Chapter 3.

2.2.5 Multiple Rendering Windows

Finally, I would like to raise that, compared to 2D screens, the AR space is (1) in 3D, and (2) can be unlimited. To help to structure and layout the 2D and 3D contents, those contents can be grouped as contexts. Those contexts can be in 2D or 3D depending on the nature of the contents, and be private, i. e., only visible to a single user, and public, i. e., visible to all. Effective context representations may increase WA about what is publicly shared and what is not.

Multiple 2D Windows

Designers can add 2D content information in AR such as charts [150], web pages [43], 2D windows [95], 2D sketches, and virtual cameras. The latter render real-world content from a different point-of-view to show occluded or remote information.

Ens et al.'s [93] *Ethereal Planes* framework categorizes 2D (extendable to 3D) elements along seven dimensions:

Perspective Is the information *exocentric* (relative to the physical world) or *egocentric* (relative to the user)? The authors suggest that users prefer exocentric views for public content and egocentric views for private data.

Movability Is the content movable in its reference frame?

Proximity can be *on-body*, *near* (within arm's reach), and *far* (outside of arm's reach).

Input mode Is the input mode *direct* or *indirect*? Indirect mode is more precise and less tiring, while direct mode is more user-friendly.

Table 2.1: Papers selected for review. References in parentheses refer to discarded papers.

		Total
IEEE ISMAR 2008–2019	[117, 149, 166, 169, 172, 178, 179, 196, 199, 201, 224, 268] ([170])	12 + 1
ACM CSCW 2008–2019	[20, 83, 105, 208, 220, 221]	6
Additional 2008–2019	[3, 30, 40, 68, 84, 87, 108–110, 131, 133, 135, 146, 155, 162, 165, 191, 194, 200, 205, 215, 217–219, 234, 244, 258, 259, 278–280, 291], ([134, 216])	32 + 2
Additional 2020–2021	[60, 159, 227]	3
Additional before 2008	[26, 28, 41, 43, 44, 62, 151, 177, 206, 213, 228, 239, 247, 254, 255]	15
	Total	**68 + 3**

Tangible Is the information bound to a tangible object?

Visibility Is the information fully visible, partially visible, or does it rely little or not at all on the visual channel?

Discretization Is the information continuous or discrete?

These dimensions are not orthogonal (i. e., independent), e. g., a distant object is difficult to directly manipulate, and egocentric objects are usually not outside the arm's reach. Some CSCW applications permit participants to have a private view where they manipulate private data, a public view visible by all, and a customized view merging all content. Using egocentric views at arm's reach for private elements allows users to not disturb the workflow of other collaborators, by not manipulating the data that is displayed for all workers. Custom views, however, can be either egocentric or exocentric, depending on their nature, e. g., some pathlines for 3D flow dataset can be visible to all or only to some users. Other private layouts may also be considered following the nature of the collaborative environment, e. g., adaptive layouts [158].

Multiple 3D Windows

Schmalstieg et al. [239] defined a 3D window as a context that contains its own scene graph—limited in space and distinguishable from others using decorations. 3D windows can be resized, moved, rotated, minimized, reopened, and have multiple graphical effects. When a user interacts with a 3D window, their system highlights its boundaries, which provides additional awareness cues. In other systems, empty 3D windows could also highlight others' private contexts while giving a frame to understand what one is doing in his or her private space.

2.3 Method

We now describe how we selected the papers we reviewed and discuss the dimensions of our taxonomy.

2.3.1 Selection Process

Our survey covers 68 papers related to AR-CSCW that we chose in two phases. The first phase was more systematic, while the second one added additional papers through an exploratory,

Table 2.2: Traditional Time-Space Matrix [17].

Time \ Space	Co-located	Remote
Synchronous	Co-Located synchronous	Remote synchronous
ASynchronous	Co-Located asynchronous	Remote asynchronous

opportunistic search. Our criteria for selecting papers were that they should focus on collaborative systems with at least a partial AR module, and they should be at least the length of a short paper, excluding extended abstracts.

In the first phase, we surveyed the two main conferences focused on AR and CSCW, respectively: IEEE ISMAR and ACM CSCW. We started with Kim et al.'s [148] survey, which classifies all ISMAR papers from 2008 to 2017, including nine papers related to collaboration. After further investigation, we added one more paper related to collaboration [199]. We then checked the proceedings of IEEE ISMAR 2018 and 2019, and added three additional relevant papers. Next, we used the ACM digital library to find papers within the ACM CSCW conference published between 2008 and 2018 and matching the keywords "Augmented," "Mixed," "AR," or "MR," which yielded 1001 papers. We filtered these papers by looking at titles and keywords, narrowing the selection down to a subset of 62 possibly relevant papers. We further filtered out extended abstracts and demos. We carefully read the remaining papers, and finally decided to add three of them to our survey based on our criteria. We compared our set of papers to those in Ens et al.'s recent survey [94], adding one additional paper from ACM CSCW. In the course of our research, the CSCW 2019 proceedings were published, from which we added two more papers to our survey. Thus, our first phase resulted in selecting 19 papers on AR-CSCW, as shown in the first two rows of Table 2.1.

In the second, exploratory phase, we selected papers we already knew to be relevant, and we used other search tools such as Google Scholar, identifying 34 papers from the 2008–2019 period and 15 more published before 2008. In order to have this chapter as updated as possible, I also added three more papers that concern collaborative immersive analytics for the 2020–2021 period. We did not place the focus on a particular venue. We used the other surveys we mentioned in Section 2.1 to find highly relevant papers in the AR-CSCW field. Finally, to be up-to-date with the recent literature, we mostly focused on papers published during the last decade. We mostly used keywords including "collaborative AR/MR," "remote AR/MR," and "co-located AR/MR." In total, we found 71 papers related to AR-CSCW, out of which we discarded 3 papers, because they focused on algorithms or theory, without much detail about a concrete system. We list the results in Table 2.1.

2.3.2 Taxonomy

A categorization of collaborative AR techniques cannot rely only on AR or only on CSCW aspects. We thus base our taxonomy on a combination of past classifications from AR, CSCW, and AR-CSCW work. The most commonly used dimensions to classify CSCW systems are time and space (Table 2.2), which are fundamental for any CSCW system.

Time is split into **synchronous** and **asynchronous** work.

Space is split into collaborators who are **co-located** (working in the same place) or **remote**.

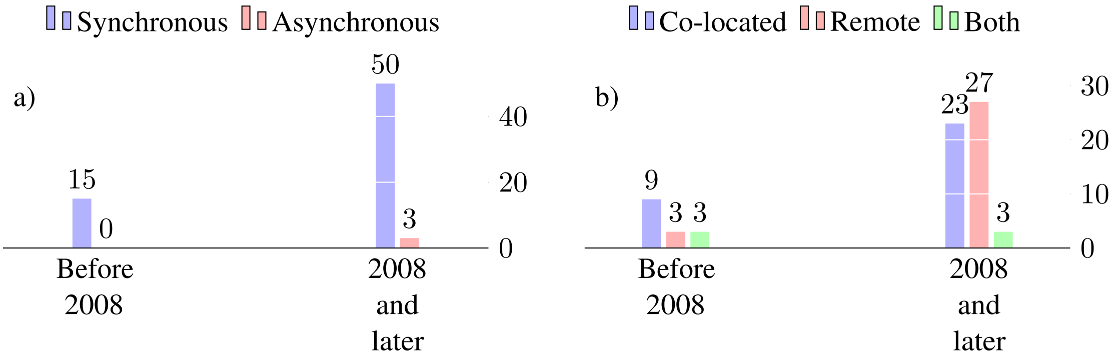


Figure 2.1: Surveyed paper distribution along a) the time and b) the space dimensions.

We further assess aspects of symmetry considered by Ens et al. [94], which we split into two dimensions for a more fine-grained analysis. We also decided to exclude their dimensions of scenario and focus because they involve too many possible settings to be useful in our context.

Role Symmetry refers to the role users have: **symmetric** if users are performing the same kind of tasks; **asymmetric** if, for instance, one user is assisting another one.

Technology Symmetry is **symmetric** if users use the same hardware; **asymmetric** otherwise.

In these last two dimensions, a value of **both** indicates symmetry between some users and asymmetry between others.

In visualization applications, the Role Symmetry dimension relates to Isenberg et al.'s [136] *level of engagement* dimension: users may not have the same level of engagement based on their roles. For example, symmetry can occur when all users are exploring their data, while asymmetry occurs when one user drives the analysis and others observe.

We also include two dimensions related to the used hardware devices, which are tightly linked to the technology symmetry dimension. As explained in Section 2.2.1, these influence the users' *immersion*, *presence*, and their *performance*.

Output Devices refer to the type of visual output devices being used. In the surveyed papers, we found that **CAVEs**, **VR-HMDs**, **AR-HMDs**, **(AR) Hand-Held Displays**, **SAR devices**, **large displays**, and **traditional screens** were used.

Input Devices refer to the type of input used to interact with the virtual contents. We found in the papers we surveyed the use of **hand tracking**, **tracked controllers**, **hand mid-air gestures** (which involves hand tracking), **touch**, **head gaze/orientation**, **eye gaze**, **tangible**, **non-tracked controller**, **speech**, and regular **keyboard & mouse** input modalities.

The dimension of "Output Devices" also relates to Kiyokawa's [151] artificiality dimension and Gupta et al.'s [117] classification of wearable collaborative systems. Moreover, designers generally know in advance the type of output they want to use due to physical, marketing, or maintenance constraints. On the other hand, the "Input Devices" dimension indicates how designers chose to empower users of AR, whether it is in a CSCW context or not. Note that the input and output devices are not independent dimensions, e. g., a touchscreen coupled with standalone AR-HMD would be an unusual and often technically awkward combination.

We did not consider other sensory output (haptic, audio) because the visual channel is typically the most important one, and because none of the work we surveyed discussed such alternative forms of output.

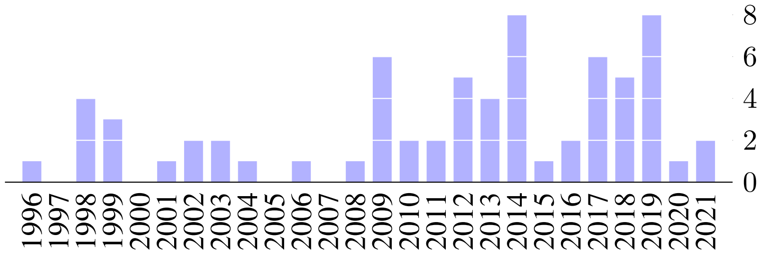


Figure 2.2: Surveyed papers by year of publication.

2.4 Paper Survey

We begin by reviewing papers that do not implement a complete AR system in Section 2.4.1, before discussing the majority of papers with complete AR systems with respect to our taxonomy, as classified in Table 2.3. In our set, most papers were published after 2008 (see Figure 2.2) due to our semi-systematic selection and, as Ens et al. [94] stated, due to the technology becoming more affordable and easier to use (e. g., smartphones in the consumer market).

2.4.1 Quasi-AR Systems

Quasi-AR systems are those setups that do not respect Azuma et al.'s [14] conditions (see Section 1.1). Such setups, however, can still exemplify aspects relevant for full AR systems. For instance, Benford et al.'s [26] virtual theater was a Mixed-Space Collaborative Work. It demonstrated that *being aware* of another space is primordial if one wants two groups to pay attention to each other. Also, to better support social presence, a common *spatial frame of reference* should be created, which allows users to better measure relative collaborators' positions and orientations and to better *feel together*. Benford et al.'s work [26] is thus a clear example of why *presence* should be considered in AR-CSCW systems.

Dunleavy et al.'s [87] spatially-aware educational system supports social presence and engagement from students who were solving riddles using location-aware smartphones. Their system, however, is not a fully AR one since it does not align 3D virtual contents in the 3D space. Maimone and Fuchs [179] proposed an algorithm and a proof-of-concept of 3D teleconferencing using several RGB-D cameras to capture a remote 3D scene. They aimed to display the remote scene on a 3D stereoscopic screen, which is rendered correctly by tracking the viewer's eyes. While they are not full AR systems, these systems still increase immersion for users.

2.4.2 Time–Space Collaborative Matrix

The time-space matrix is a well-known CSCW taxonomy [17]. We now discuss its implications for AR-CSCW approaches.

Asynchronous Collaboration

Little work explored asynchronous AR contexts [134]. We only categorized three surveyed papers [135, 146, 191] as such, that discuss the asynchronous creation and visualization of annotations. Irlitti et al. [135] studied how physical 3D tokens guide co-workers faster and with fewer errors to virtual anchored tags. The use of an AR-HHD with a projector, however, may have biased the results due to its low FoV compared to SAR. Kasahara et al. [146] allows users to collaboratively tag an outdoor environment using AR-HHDs. Finally, Mora et al.'s *CroMAR* [191] allows users to visualize geolocalized tags (e. g., tweets) in an outdoor environment, to share viewpoints via emails (which are visualized with traditional tools), and to rate these tags.

Irlitti et al. [134] suggested to divide collaborators into (1) *producers* generating information to be visualized by (2) *consumers*. Because asynchronous collaboration often revolves around annotations, they proposed three research questions: (1) what are the time effects on collaborative work, (2) how to capture and later visualize AR annotations, and (3) what are the effects of the lack of communication cues.

The visualization of annotations may also depend on the output dimension. Indeed, SAR environments are usually bound to a specific room, which reduces their compatibility for asynchronous distributed contexts, while AR-HMDs may work everywhere. Moreover, systems relying on multiple output devices may also diverge a lot from standalone systems because some devices may not be AR-compatible (like the HHD used for input) or because they propose different AR views and functionalities, e. g., one output device used for annotations only. We can also separate AR applications displaying virtual objects tightly bound to real objects (e. g., AR situated analytics [90]) from those not proposing such bounding (e. g., climate visualization software). The later can rely on traditional tools [283, 284] which should be considered as well.

Finally, as we explain in Section 2.4.3, AR-CSCW applications may use both an exocentric and an egocentric view, which may impact how the registered annotations should be rendered. Indeed, we can imagine an asynchronous application where users are always, for each work session, in pair in a synchronous manner. In these work sessions, users may need to get an overview in both view modes, which impacts the visualization of asynchronous annotations.

Synchronous Collaboration

Figure 2.1 highlights that synchronous remote setups were more studied than co-located ones during the past decade. We hypothesize that distributed setups require more research because awareness, one key functionality of CSCW, is more difficult to provide in a distributed than in a co-located setup, where co-workers directly see each other.

Co-located Collaboration As embodiment is natural in co-located systems, awareness of others is better perceived since the *bodies and consequential communication* cues [118] follow social rules. For instance, in Rekimoto's [228] system, users could easily interpret others' deictic gestures. Rekimoto also found that users are mostly absorbed in the workspace. The cues, if any, should thus reach the workspace and not stay on users. This absorption may explain why Nilsson et al. [196] and Prytz et al. [224] did not notice a reduction of the communication quality by occluding the users' eyes. The communication was instead supported by other cues still visible without artifacts, such as head orientation and body gestures. Sometimes it can be problematic, however, to have physical bodies as two users may want to have physical access to the same locations. In such cases, one strategy is for one user to create virtual clones of the target object [212] and manipulate the data [294] or to create slightly different virtual coordinate systems in some specific cases [201].

Distributed Collaboration Many fields are interested in AR-CSCW as discussed before. However, co-workers are not always physically able to be at the same place, hence the need to collaborate remotely. Billinghurst et al.'s tool [41], for instance, was one of the first teleconferencing systems to use AR. The main focus of such a system is to improve the sense of presence (e. g. [82, 149, 280]). They also rely mostly on asymmetric technologies (see Section 2.4.3), with remote guidance and video conferencing as common scenarios.

Some surveyed papers focused on remote guidance using annotations, as body language was partially or not-at-all conveyed. Annotations are thus a key component of CSCW systems which tend to improve the overall communication. For instance, Stafford et al.'s [247] god-like interaction technique allows outdoor users to see objects placed by an indoor user (e. g., cans). The indoor user has an exocentric god-like view and places objects using a table surface augmented with cameras, supporting navigational cues such as "go here" by annotating landmarks.

For teleconferencing, Kim et al. [149] showed that pointing cues and annotations created by a remote person allow all users to feel more *together*, *connected*, and facilitate a better *mutual understanding*. They argued that video streams are better from a user's eye perspective than hand-held views and that virtual pointers are better than virtual annotations if users can vocally communicate. Users could draw annotations directly into the moving video stream, and to freeze it to draw world-stabilized annotations. However, the authors argue that both modes may be inconvenient. Gauglitz et al.'s [108] world-stabilized technique allows a remote user to draw in a 3D reconstructed world, first constructed by panorama mappers, and then by SLAM algorithms [110]. The remote user can navigate in this reconstructed world with a stabilized view which is continuously updated by the local user's viewpoint. Both experiments showed that their technique performed better than video-only. The authors did not find evidence of a difference between video stream-stabilized and world-stabilized, which users preferred, however. The authors later improved the remote user's user interface (UI) [109], introducing spray-like annotations (i. e., each point p_i is assigned a depth d_i) and planar (dominant or not) annotations (i. e., all p_i are on a plane). Users preferred the dominant plane method. Lien et al.'s algorithm [170] may improve this anchoring, hence the call for another study. Later, Huang et al. [131] created *HandsInTouch* to capture and display the remote user's hands on a multi-touch tablet and investigated sketching features. They found that the combination of both remote hand rendering and sketching may be better for some specific tasks, at the cost of a higher mental demand, which may be influenced by their UI design. Ou et al.'s system [206], finally, allowed remote users to draw normalized (i. e., recognized patterns) or free-hand strokes on a video stream. Their study showed that strokes improve the performance and that their auto erasures after a defined duration improves it even more. Normalized strokes can be understood once and then be cleared as soon as the local user grasps the idea. The authors suggested it to be configurable.

In remote expert guidance scenarios, the cited systems implemented the annotation feature in another set of output and input devices accessible only by remote users, creating a technological asymmetry between local and remote users.

2.4.3 Technology and Role Symmetry

Table 2.3 shows a strong relationship between the role and technology dimensions, with 48 out of 68 papers ($\approx 70.5\%$) sharing the same values for these two dimensions. Out of all 36 papers that discuss remote collaboration, 28 ($\approx 78\%$) use partially or fully asymmetric technologies. This can be explained by expert guidance (i. e., an asymmetric role) being the most studied remote scenario (see also Ens et al.'s survey [94]), which typically relies on asymmetric technologies to suit both the local users performing the task and the remote experts giving instructions.

Technology Asymmetry

Technology asymmetry concerns systems where users benefit not from the same types of output devices simultaneously. Because the physical world is visible with AR devices, it is hard to keep spatial coherence between all remote users. In effect, all but two papers [3, 172] that discuss distributed asymmetrical setups rely on remote users using traditional screens or VR technologies. For instance, Komiyama et al.'s [155] *JackIn Space* allows seamless transitions between first-person views (i. e., remote users' viewpoint) and a third-person view (i. e., bird's view). Accessing both views was preferred by users over only having access to first-person views. Lehment et al.'s [169] merging algorithm may be appropriate for two remote users both using

AR devices, yet their solution is limited to two separate rooms. Still, a study might be required to measure the efficiency of such consensus regarding the immersion and social presence metrics.

One of the main research questions in such setup concerns awareness conveyance between remote VR experts and local AR users. Oda et al.'s [200] system allows a remote expert, either via VR or video-streamed AR, to guide a local AR user with virtual replicas. They showed that direct manipulations were faster for assembly tasks than drawing annotations. Piumsomboon et al.'s [217] *Mini-Me* displays an adaptive remote VR user's avatar always visible to an AR user, either as a mini or as a life-size avatar. The avatar conveys pointing actions, is surrounded by a halo notifying the mini avatar entering of leaving the local user's FoV, has a blowing ring location cue for the actual life-size avatar, and has a graphical distinction between the mini and the life-size avatars. Piumsomboon et al.'s [219] *CoVAR* facilitates remote collaboration between AR and VR users. The AR environment was captured, reconstructed, and sent to the VR user, who can control his/her scale (i. e., magnified or minified). FoV and head, hand, and gaze orientations were captured and shared. They later studied these cues [215] and found that a 3D frustrum denoting the remote user's FoV improves performance, which increases even more when coupled with head or eye gaze cues. Surprisingly, they found that eye gaze alone is worse than head gaze, which can be explained by the eyes tendency to look everywhere. They also found that the balance of actions between the VR and AR users is not symmetrical, mostly due to the larger FoV the VR user has, improving his/her efficiency. Piumsomboon et al.'s [218] remote collaboration system allows a local AR user to manipulate, using a tangible device, the remote VR user's position. The study showed that the control a local user has over a remote user's orientation should depend on scenarios, and that displaying the FoV in addition to an avatar was preferred among participants. Adcock et al. [3] studied three techniques to track a remote collaborator's point-of-view. They found that, even if wedges were not valued in users' qualitative responses, they performed well in the quantitative results in spotting 3D locations via 2D cues. The authors suggested to couple wedges with shadows cast from the location, as local users felt more comfortable with them due to their everyday life similarity.

Lee et al. [165, 166] showed to a remote VR user a 360° stabilized live panorama view of a local AR user. The VR user saw the AR user's FoV but controlled their own viewpoint to avoid sickness. The VR user's hand positions were shared to support non-verbal communication such as deictic cues. Surrounding halos allowed the local user to constantly track the guest's hands, which improved performance. They found that independent views improved social and spatial presence, require less mental effort, and were preferred by most participants, who also appreciated the visual cues.

In addition to awareness cues, annotations may improve further users' co-presence and performance. Teo et al. [259] showed that annotations, coupled with pointing action augmentations, further improve the local AR user's understanding about what the remote VR user intends to do.

From these examples we can learn that awareness improves communication and co-presence, hence increases users' performance. Not all awareness cues, however, should be continuous, or the high amount of information can lead to distraction [64] and cognitive overload [299].

All papers discussed so far address one remote user collaborating with one local user. Thanyadit et al. [268] proposed an educational system where multiple students are in a VR environment, supervised by an AR professor. This is the only paper we found which focuses on one-to-many scenarios using AR. The authors listed awareness cues, views, and algorithms that place the students in the environment to maximize the professor's efficiency, depending on the scenario (helping individuals vs. supervising the group).

Finally, as AR is increasingly used in mobile scenarios, altering the environment where other people live raise questions [221]. This research area about ethics is, however, not well explored yet. Only two papers we surveyed [146, 221] relate to it; see also Ens et al.'s survey [94].

Mixed Outputs in Co-Located Setup Some co-located setups rely on multiple output modalities. Kiyokawa et al. [151] introduced one of the first mixed-space collaborative systems, in which users switch between AR and VR with one HMD. In VR mode, users can change their position and scaling without physically moving. The *MagicBook* [44] is a physical book augmented with 3D scenes to illustrate the text using embedded markers. Users could either be in an exocentric AR view or in an egocentric VR view where they are in the book scene and can be viewed by AR exocentric users as 3D avatars. Several years later, Roo et al.'s [234] system allowed users to be either in an exocentric SAR or to wear a VR-HMD to get access to more complex virtual contents and being able to enter in the mock-up objects in an egocentric way. The exocentric users see the egocentric ones as 2D arrows. Clergeaud et al.'s [68] system allowed a user to visualize and manipulate, in an egocentric way using a VR-HMD, an engineering prototype, while the others saw the prototype and the VR user in an exocentric SAR manner.

Through these examples we see a distinction between a *collaborative mode* where users share the same space, which appears mostly in AR, and a *semi-private mode* where users are immersed in an egocentric view to get more insight regarding the object of interest, which mostly appear in VR. VR modalities are, however, not required. The *SHEEP* management game [177] allowed users to control sheep during their lifetime, rendered through several output devices: AR-HMDs, AR-HHDs, laptop screens, and SAR. While the UI was adapted to each display, the authors did not demonstrate scenarios where such a number of displays would be useful. Butz et al.'s [62] *EMMIE* incorporated multiple devices, which were associated to their own input paradigms, to exchange graphical object informations. This environment accommodated HHDs, traditional computers, wall-sized displays, and AR-HMDs, all sharing the same database. When needed, a user could transfer an object from the AR-3D space to a laptop for further analysis.

Technology Symmetry

Compared to distributed systems, co-located ones rely more on symmetrical technology, with 24 out of 32 co-located papers ($\approx 75\%$) categorized as such. Technology symmetry leads to the use of fewer output device types. We saw only four papers with multiple output devices, including three which used traditional screens in addition to an AR output.

Role Asymmetry Six papers we surveyed rely on asymmetric role and symmetric technology between users. Three of them concern awareness. Oda and Feiner [199] studied object selections in a 3D space captured by depth cameras. Sodhi et al. [244] used AR-HHDs and several depth cameras to allow remote users to point at, annotate, and manipulate virtual objects in a local user's environment. Le Chénéchal et al.'s [162] technique stretches a remote expert's arms from a local user's shoulders for guidance tasks. We clearly distinguish here active from passive users.

Subjective views allow users to see the workspace more adapted to their roles. Nilsson et al. [196] and Prytz et al. [224], e. g., studied crisis agents (e. g., police, military) using an AR-CSCW system to support discussion, with each agent having a customized view. Similarly, players see their opponent's tiles as blank in Szalavari et al.'s [254] Mah-Jongg AR game.

An expert user can also be the only one manipulating the AR space to empower the discussion this user has with, e. g., a customer or trainee [83].

2.4.4 Output and Input Devices

The output display usually drives how users interact with the system. For example, a mouse and a keyboard do not fit an AR-HMD well when users move in the 3D space. We thus group the different approaches by their main display.

The choice of the output device can significantly alter the users' performance. For instance, Müller et al. [194] showed that landmarks increase the sense of presence by creating a common ground between remote users. The use of HHDs, however, may have biased the speed results: Compared to HMDs, the HHDs' narrow FoV and lack encouragement for users to look around may have introduced the delay Müller et al. [194] noticed at the end of their study.

While some systems do not enforce the same technology to every user (see Section 2.4.3), others give users the same set of output and input devices. Butz et al.'s [62] system allowed users to benefit from different input paradigms (touch-based for HHD, tangible for AR-HMD, keyboard & mouse for PC), these outputs possibly being complimentary (e. g., using a tracked HHD to see complementary objects through an AR-HMD). Via these sets of output devices, users could use personal devices for private information (e. g., HHD). They also used a red spotlight as a metaphor to indicate private properties of a 3D object. Later, Benko et al.'s [28] archeology visualization software combined multiple output devices. Due to the low resolution of AR-HMD resolution at the time, they used a tabletop display and a tracked HHD to show high-resolution documents, while keeping AR-HMDs for 3D-stereoscopic views and 3D input.

These displays still have some features in common. They usually provide all users with an active device which can display private and subjective contents per user, the latter being reviewed in Section 2.4.3. Users found the distinction between public and private views useful in Billinghurst and Kato's [43] informal trials of an AR web browser.

AR Head-Mounted Display (AR-HMD)

AR-HMDs have been extensively used in the past, with 10 ($\approx 15\%$) papers relying only on, and 29 ($\approx 43\%$) partially on, AR-HMDs, out of all 68 AR-CSCW papers we surveyed.

By merging the users' physical environment with the visualization and the interaction spaces, AR-HMDs seem promising for co-located settings. Büschel et al. [60] investigated the use of AR-HMDs for co-located spatio-temporal data analysis. The use of AR-HMDs allows people to see collaborators in sight (compared to VR-HMDs), but also to reuse the space where the spatio-temporal data were captured, which gives a strong spatial context to the data being visualized. For example, one can see trajectories of people inside the same room as where the data was captured. The authors argued that such a system engages their users more compared to traditional tools (e. g., 2D web-based collaborative environments) and allows users to grasp the general patterns of the data in an efficient manner. This discussion also relates to our arguments in Section 2.7.3.

Moreover, AR-HMD users can share their viewpoint to others which might be useful in distributed setups (see Section 2.4.3). Because AR-HMDs are still expensive, because not everyone feels comfortable with them, and because there may be too many passive users, Franz et al. [105] allowed these passive users to grasp, via a traditional screen set on a wall, the active AR user's experience who visualizes an augmented 2D landscape map rendered on a tabletop. Franz et al. [105] proposed both an *over the shoulder* view and an abstract data representation of the object of interest. They showed that the *over the shoulder* view forced the active user to constantly look at the object of interest for the passive users to see it as well. Displaying the intersection with the AR user's head gaze on the tabletop improved the communication, the sense of presence, and the usability of the system.

Display Characteristics AR-HMDs tend to fit the users' anatomy, which might be useful in engineering and educational scenarios relying on 3D data. Dong et al.'s [84] *ARVita* used an AR-HMD to augment a tabletop displaying markers. The tool's goal was to improve students' learning of engineering processes. For the authors, using a tabletop allows users to have a better discussion, when compared to traditional tools. Wang and Dunston [278] demonstrated that users were faster in the civil engineering field using AR, compared to traditional perspective sketches. However, participants were uncomfortable with the HMD used at that time, which may change with modern AR-HMDs (e. g., Microsoft's HoloLens). They also found that navigation cues (e. g., a compass) are useful because multiple 3D transformations such as rotations and scaling can confuse users who are trying to orient themselves. Similarly, Wang and Dunston's study [279] showed that AR-HMDs drastically increase the speed in error detection tasks in civil engineering field, when compared to traditional tools.

Input Modalities Not all input modalities are available by default on AR-HMDs (e. g., touch, keyboard & mouse) due to relative position and portability constraints. They often rely on voice, tangible, and gestural interfaces. Ong and Shen's [205] CAD system allowed co-located users to manipulate and model common 3D shapes using tangible devices. For consistency, only one user could modify the 3D model at a time. When finished, users committed their changes that others could then (in)validate. The *CoVar* system [219] combined VR and AR-HMDs to benefit from both technologies. Users were able to move objects closer or further away with mid-air gestures. The *GARDEN* system [199] studied, for the selection of 3D objects, how a user can refer to an object, highlight the selected object, and how others acknowledge it. This technique was more accurate than laser pointers, virtual arrows coming out from the controller, and video sharing techniques. But their technique was also slower, which can be explained by the higher number of steps required. Mahmood et al.'s system [178] allowed users to remotely collaborate and visualize geospatial data using AR-HMDs. Their user interface, allowing history recording, 3D transformations, and data filtering, relied on both voice and gestures. Their system, in addition to Lehment et al.'s [169] consensus algorithm, is the only surveyed system using solely AR-HMDs in distributed collaboration, which may be due to their application domain that does not require a strong relationship between virtual and real objects.

While AR-HMDs do not propose their own touch input, several systems use touch support by adding other input devices that provide 2D input [239] and/or tangible control [62]. When mobility is less important, AR-HMDs can also be coupled with a tabletop and 2D views [84, 105, 131]. Wang et al.'s [280] teleconferencing system regrouped, per local setup, one camera and one screen for each remote co-worker with whom this local user interacts. This improves workspace awareness and spatial faithfulness. Each user had a tabletop displaying markers augmented by AR, conveying remote collaborators' arms and objects of interest.

AR-HMDs as a Complement While systems such as 2D projectors for, e. g., slideshow presentations (one active user vs. multiple passive users) or large displays are currently used or being investigated for co-located collaborations, they possess inherent flaws. For example, users cannot simultaneously have an overview of what a large display renders, where users need to step back, and interact with it and have high-resolution visualizations, where users need to step in. Similarly, the input modalities of the AR-HMDs are limited. This makes designers want to use hybrid setups where the used devices would solve the flaws of each other. This may also be the reason why we have almost three times more systems ($\approx 43\%$ of the total amount of papers) relying on the partial use of AR-HMDs, compared to them being used alone ($\approx 15\%$ of the total amount of papers).

Reipschläger et al. [227] thus investigated the use of AR-HMDs combined with a large display to have the best of both worlds. The large display renders high-resolution data and acts as a public screen, while AR-HMDs give overviews of the scene even if the users step in. AR-HMDs also allow for private data rendering without disturbing co-workers. Similarly, HHDs have high resolution and provide touch input with haptic feedback, but they are limited in size compared to AR-HMDs which have an unlimited visualization space but are limited in interaction and have lower resolution. Langner et al. [159] then combined these two types of devices to fix the flaws of each other. Their main devices are the HHDs as they are largely available among users and consumers. Each user might then follow and join the discussion using solely the HHDs. Langner et al. [159] then used the AR-HMDs to give an overall context and link information between multiple HHDs. Finally, Franz et al. [105] used AR-HMDs to augment a tabletop as stated previously. To summarize, while systems can rely solely on AR-HMDs, the AR-HMDs can also complement other devices to benefit from each platform.

Hand-Held Display (HHD)

HHDs were also often used, with 21 ($\approx 31\%$) surveyed papers relying partially ($n = 12$; non-AR use-cases: $n = 3$), or fully ($n = 9$; non-AR use-cases: $n = 1$ [87]) on them.

HHD as a Complement HHDs are often used to complement a larger system. Surale et al. [252], e. g., explored the role of a tablet in VR CAD contexts, which can be applied in AR. The *Studierstube* [239], discussed in Section 2.2.4, used a personal pad to add touch modalities without using it as an output device. MacWilliams et al.'s system [177] used HHDs as spatially-aware tangible objects, and Butz et al.'s system [62] used HHDs as a magical lens for users not wearing an HMD, or a magical mirror to capture an object. Benko et al. [28] used an HHD as a magical lens to visualize documents in a higher resolution. AR-HHDs, if coupled, have mostly been used with AR-HMDs in co-located contexts. Only one paper in this survey [177] coupled an AR-HHD with SAR to add tangible input. Future work may investigate how the sensors of such devices, in addition to their high display resolution, improve the collaboration in SAR-CSCW contexts.

Ubiquitous AR-HHD Touch devices such as smartphones are highly present in the consumer market, which can be used as affordable and widespread AR devices. Huynh et al. [133] claimed that AR-HHD collaborative games enable a kind of social play experience not present in non-AR counterparts. AR-HHD games can use the limited viewport as a feature, e. g., to intentionally hide information from the player. Their user study showed that, except for tracking issues, players enjoyed playing this kind of game. Occlusion from one user's device, however, may disturb the visual tracking of the other devices. Hence, multiple sources of tracking (e. g., SLAM and visual) may be needed. Bhattacharyya et al.'s [40] game design iterations showed that AR-HHD are so ubiquitous that users tend to use them in a specific way, e. g., tap-and-hold for selecting and moving objects, instead of gestures which are seen as awkward and cumbersome interactions. Due to the users' experience of 2D applications, some cues should be present to encourage users to physically walk in their 3D AR environment.

Awareness of AR-HHD While we do not imagine to vertically align the users' heads, it may not be the same for hands holding AR-HHDs. Oda and Feiner [201] addressed this collision issue by applying small shifts between the virtual and the real environments for each user, which improved performance without disturbing the users. This should work only, however, when strong spatial coherency between the virtual and the real environments is not mandatory, or users may be confused. Moreover, AR-HHDs occupy the hands and force users to look at their screens, which may not be appropriate for all cases. Nilsson et al.'s [196] iterative study showed that users preferred to use an HMD over an HHD to free the hands, allowing users to interact both with other people (social communication) and with the physical workspace. This may explain why Sohdi et al.'s [244] work is the only surveyed work which tracks users' hands with an AR-HHD, even though the hands are holding the device.

Despite these limitations, the movements of the tablet are visible by all, which everyone understands as manipulating the AR world in co-located contexts. In our survey, the only work focusing on awareness is Rekimoto et al.'s [228], which proposed to add an *eye gaze* equivalent as an awareness cue.

Spatial Augmented Reality (SAR)

Most of the SAR papers in our survey discussed how the whole environment becomes tangible. Therefore, we already mentioned some papers (e. g., [213]) in Section 2.2.4. SAR systems, by typically adding several sensors in a room, can scan the room and enable mid-air gestures by tracking all the users' body. SAR setups are used in distributed and co-located settings.

In distributed settings, Pejsa et al. [208] brought a remote user into a local user's SAR space. This works well when the remote user is virtually close to their anchoring position. They compared their setting with *Skype* and *Face To Face* regarding *speed*, *presence*, and *communication efficiency* for assembly tasks. One further study might compare their settings with remote AR-HMD ones. Lincoln et al. [172] showed how SAR can support remote medical consultations by capturing the doctor's body at one location and using SAR at the other location. They proposed to capture the doctor's head and retransmit it onto a remote animatronic head using projectors.

In local settings, Benko et al. [30] allowed two face-to-face users to see in perspective without wearables. The personal tracking and the perspective facilitate collaboration and, even with depth cue conflicts, users were able to determine distances and pointing directions. Wilson and Benko [291] allowed a maximum of six users to move, in a tangible way, media objects between surfaces presented in their smartroom. It is also important to note that six users are possible because the system does not render the scene from users' viewpoints, otherwise the system would be limited with regards the framerate of the projectors.

Finally, SAR environments can be non-intrusive, which might be useful in contexts where people expect and are used to some pre-defined behaviors (e. g., pencil selling [83].)

Most of the surveyed papers using SAR in all endpoints are symmetric regarding the role axis. The other work focuses on passive vs. active users [83, 172]. One interesting research question could be how SAR can permit parallel work, with users assigned to different roles. This setup would differ from other work using multiple types of output devices suiting the different roles between users.

2.5 Design Considerations

Most of the papers we surveyed for remote settings focused on studying which workspace awareness (WA) cues improve user performance. Others provided insights about the benefits and limitations of output and input devices in local setups. Moreover, some early work showed the capabilities of AR systems applied to specific scenarios. These are usually linked to discussions about how AR influences mutual communication and the cognitive capabilities of users. Finally, a minority of papers discussed techniques to capture users and environments and to merge local and remote settings. Based on these contributions, we now provide high-level design considerations we extracted from these papers.

2.5.1 Private/Shared Views, and Awareness Cues

Embodiment in remote systems provides information about the users' locations. The general focus of a collaborator can be provided through FoV representation, and eye- and head-gaze cues [215,217–219]. FoV cues work best and eye-gaze cues are not sufficient [215]. Embodiment can rely on hands-only [131], head-only [178], exocentric life-size avatars, or egocentric mini-avatars [217] which allow workers to keep the remote collaborators' expressions in sight. In co-located scenarios with a shared exocentric view, most of the listed cues come for free and can even be removed. A weaker eye contact due to wearing an HMD, e. g., can allow users to save their hands for 3D interactions and body gestures [196].

In both local and remote scenarios, WA cues are usually needed. Our survey showed that cues, such as compass or arrows [278], support navigational tasks [247]. To create a *common "virtual-real" space* [194] in distributed settings, systems can render additional virtual objects. Finally, annotations support remote discussion, see Sections 2.4.2 and 2.4.3.

Users may need to customize their views based on their role [196] or use private egocentric spaces [178]. Such a space, however, can complicate the communication with co-located collaborators due to a lack of shared information. If a user manipulates their private space, the physical appearance of their head and body with respect to another user's 3D scene can be misleading. One way to avoid such confusion is to enforce the switch between (1) an AR mode for when users want to share the same exocentric 3D scene and (2) a VR mode for *all* users when private, egocentric views are desired, so that one user's physical head and body will not give misleading cues to other users (e. g., [151]). In the specific case of distributed collaborative VR/AR scenarios, remote VR collaborators should control their own viewpoints for comfort, and FoV/head-gaze cues should be adapted [166].

Awareness cues can also be added in AR or VR to indicate each collaborator's virtual position and orientation [234]. By reflecting on all the surveyed papers, an interesting scenario to investigate another kind of awareness cue may be the following. If Alice and Bob are co-located, and Alice wishes to employ a private egocentric view, but Bob wishes to remain in AR to, e. g., share exocentric views with Carol, Bob's headset may display a virtual mask over Alice's to indicate that she no longer sees the same exocentric 3D scene.

2.5.2 Distributed Work Using Asymmetrical Technology

Most of distributed AR-CSCW systems we surveyed rely on either two similar rooms for both endpoints, or on one remote user using a VR-HMD while the local user relies on AR (Section 2.4.3).

Relying on asymmetrical technology (e. g., AR-HMD + VR-HMD) makes sense when there is a strong relationship between the virtual and the real environments in the local space. When both users have the same role, we do not envision that using AR devices for both users should be considered if virtual objects are tightly related to real objects.

However, as Mahmood et al. [178] showed, using AR devices in both endpoints can be considered to have flexibility in both places when the virtual objects can float around.

2.5.3 Multimodal Interfaces Improve User Experience

Hybrid systems aim to support a heterogeneous mix of hardware for both output and input (e. g., HMD, HHD, traditional screen; mouse, keyboard, mid-air gestures). The chosen AR hardware may depend on the available devices, the users' roles, and how much effort the users are willing to invest to participate [105]. Mouse + keyboard input, for example, is typically less tiring than mid-air gestures [138], and a 2D screen, projector, or tabletop can be casually used by multiple users [105, 220] if they are mostly "passive."

Through our survey, we found four roles for traditional screens. They, first, permit experts to follow what a remote AR user is doing and allow them to send annotations (e. g., [110]). Second, they allow non-AR users to follow a discussion about what a co-located AR user is visualizing (e. g., [62]). Third, they provide a common ground for 2D information (e. g., video stream [20]). Forth, they allow users to work with their usual tools and profit from high-resolution screens [28].

Hybrid systems using personal 2D screens can provide user-friendly personal 2D views and UIs [62, 196, 228].

2.5.4 Synchronous Workspace Consistency

The ownership of manipulated objects can support workspace consistency because it facilitates a more robust synchronization [205, 228]. If an object is at the center of the discussion, private proxies acting for public counterparts can be visualized with a correct viewpoint and allow users to direct manipulate them [200], e. g., using a private World-In-Miniature [249]. Private objects can also be shared and frozen by the owner of the object [178]. In an AR-CSCW context with more than two users, awareness cues can indicate which user is remotely manipulating or owning a public object, making it easier for others to track the users' actions. This can happen, e. g., through the decorations of the 2D/3D window-contexts [178, 239] introduced in Section 2.2.5.

2.6 Remaining Research Areas

In addition to these design considerations for established environments, we now discuss some remaining research areas based on the survey. We intentionally avoid overlap with Ens et al.'s [94] recent survey which already mentions (1) complex collaboration structures in time, space, and symmetry, (2) convergence and transitional interfaces, (3) empathic collaboration, (4) collaboration beyond the physical limits, and (5) social and ethical implications.

2.6.1 Co-Located Awareness in Interactive Environments

We surveyed approximately as many papers investigating synchronous co-located setups as synchronous distributed setups (Table 2.3 and Figure 2.1). Marai et al.'s [180] study showed that a team is the most productive when they are within the same space. However, questions about providing awareness to evaluate collaborators' internal state (e. g., privacy status), annotations' (e. g., who puts what?), workspace manipulation, and replaying previous actions and handling the undo/redo actions performed by multiple users (which relate to Mahmood et al.'s [178] interactions and designs) are areas that, to the best of our knowledge, are not yet well covered. In my system (Chapter 3), I am providing glyphs floating above users' heads. I then adapts the decorations of the 3D windows on which I am relying to know which co-located user is manipulating these views by matching the color of the glyph with the color of the 3D windows. I am also replacing private scene graphs with 3D question marks to let the other users, who do not own those views, know that those views are private.

How viewpoints are shared so that each user is aware of others' views is not yet answered either. Proxies allow users to see the target objects without occlusion, but they may lose the correct orientation and spatial referencing. Users may need to move around to see the object. Otherwise, if viewpoints are shared via a video stream (i. e., collaborator A directly sees B's point of view), it remains unclear how coherency issues can be handled, because collaborator A will, in this example, see their own body through B's view.

2.6.2 Annotation Rendering in Co-Located Settings

Few surveyed papers [110, 135, 146] supported 2D annotations (canvas), but they were rarely meant for multiple users. Should they be rendered in an egocentric way, exocentric way, or a mix of both? As a first approach, we hypothesize that it may be preferable to render annotations in an exocentric view for *workspace coherency*, while having an egocentric view bound to the exocentric one for better *legibility*. Still, more studies are needed to understand possible approaches in regard to these criteria. In multimodal systems, however, it might be better to anchor annotations with respect to the other output devices when both the AR devices and the other output devices are tightly linked in the 3D space, e. g., a public large display tightly linked with private AR-HMDs [227].

The on-the-fly creation of anchored annotations is also not well explored, as most studies we encountered focused on visualizing already registered annotations. For such tasks, the implications of the interaction techniques depending on the output dimensions remain unclear, i. e., how well these techniques support the collaboration and mutual understandings of all users. Let us consider AR-HMDs as I am focusing on them in my PhD. Arora et al. [12] showed that drawing on a 2D canvas using 3D mid-air gestures suffers from instability even if the system projects the 3D strokes onto a 2D plane. Moreover, writing texts using AR-HMDs is difficult. One can thus rely on additional Hand-Held Displays to draw the 2D strokes and enter texts. However, converting the number of pixels (i. e., the canvas local space) from a given HHD to a specific size in the users' physical environment is challenging, as annotations may not have all the same size. The mapping of the annotations from the HHDs to the AR-HMDs may also lead to challenges. One solution might be to use the HHD as a tangible device; see Chapter 5.

2.6.3 Role and Technology Asymmetry in Remote Work

Most work on expert guidance (see Section 2.4.3) focused on one expert guiding one user, except for Poelman et al.'s system [220], which allows multiple experts to guide a remote user without further exploring that direction, and Thanyadit et al.'s work [268], which allows one professor to supervise multiple students. The management of an undetermined number of experts guiding an undetermined number of local users remains to be explored. Similarly, more work is needed to understand how awareness and annotations could be rendered so that local users understand who is doing or has done what.

Moreover, all surveyed papers use an AR-HMD at one endpoint, no matter what devices (VR or AR) are used for the other endpoint. We summarize that, to follow a local user's position with a VR-HMD, the most common approach is to track the local user's head. We found only one paper that associated the remote VR user's position with a local AR user's tangible device [218]. Head tracking can also be achieved in SAR but reduces mobility. It remains unclear if a remote VR-HMD user can be directly coupled with an AR-HHD to keep the mobility, and what benefits and limitations such coupling entails.

2.6.4 Output and Input Devices in Asynchronous Work

In Section 2.4.2 we discussed asynchronous AR-CSCW systems, which mostly concern the creation and consumption of annotations [134, 135, 146]. Because multiple input modalities can benefit AR, it may be interesting to understand their implications in collaborative environments. Understanding how output devices influence the collaboration, and how coupling several devices changes the creation and consumption of annotations [134] are also research areas interesting to investigate. Moreover, because AR devices can provide both exocentric and egocentric views, it remains challenging to capture and reproduce recorded actions in both views.

2.6.5 Mixed Input Modalities

Mid-air gestures, although considered by many to be natural forms of input, are often not preferred for long-time use because they induce fatigue and have low position accuracy. In 3D selection tasks, for example, existing hand tracking techniques usually result in low precision. This issue is further amplified when one wants to point at a location beyond arm's reach; see also the experiments I ran with my co-authors for the specification of 3D positions (Chapter 4). Oda and Feiner [199] tackled this problem using a special tracked device, but it could be cumbersome for existing systems. Moreover, their selection speed drops a lot. Under this circumstance, hybrid interfaces (e. g., AR-HMD+HHD [159, 262] (Chapter 3), AR-HMD+PC [283, 284], AR-HMDs+large displays [227]) with AR are currently being studied for visualization, both for more effective input and as a combination with traditional work environments. However, the mapping between the different input and view modes (e. g., 2D input with the 3D AR view [175]) remains challenging. In this PhD, I studied the use of a tangible multi-touch tablet coupled with an AR-HMD for volumetric selections (Chapter 5). My co-authors and I showed that users did not prefer direct manipulations compared to remote interactions. Moreover, we showed that, instead of using 6 Degree of Freedom (DoF) manipulations, which are often described to be “natural”, they may yet not always be preferred by people in practice as our participants mostly preferred constraining themselves along 1 DoF manipulations.

More research is also required to understand when and why users switch from one input or output device to another, possibly depending on the action and the needed accuracy. In this PhD, I found evidence that the choice of output devices, e. g., external 2D screens against AR-HMDs, influences greatly on which screen users focus in hybrid environments that combine multiple output and input devices. In the research project I describe in Chapter 5, a user manipulates a multi-touch tablet to perform spatial selections. The user draws on the tablet a 2D shape which he or she extrudes by moving the tablet around. Because of perspective projection issues, the tablet renders an orthographic projection of the visualization for the user to draw accurately, in size, the 2D lasso to extrude. To have a better depth understanding, the tablet was either coupled with an external 2D screen or with an AR-HMD that renders the scene using perspective projections. Hence, we use exactly the same input modality for both conditions. However, my co-authors and I showed that users’ attention, for the specifications of regions of interest, switched from the multi-touch tablet that they manipulated (i. e., the input device) in the “2D external screen” condition to the AR-HMD (i. e., the output device) in the “3D AR” condition.

Finally, these private 2D interfaces provide only poor awareness cues, as the input does not reach the public workspace that is visible to others [228]. More research is then necessary to find suitable awareness cues when one is manipulating such a private 2D interface. While I am providing some awareness cues in my system, I did not study them and cannot give any insight.

2.6.6 Comparison Between AR-HHD, AR-HMD and SAR

Although we are aware that some hardware is unavailable due to physical constraints, in relatively few of the papers we surveyed the authors justified why they selected their used output devices (e. g., [60, 159, 196, 227]). More research is required to understand when and where would a specific display be preferred, which may depend on the tasks and the users. Since several settings combine HHDs with other output devices (e. g., [28]), we wonder when an HMD combined with a multi-touch tablet for 2D input is more preferable than a standalone HHD. I discuss this question within my PhD; see Chapter 3.

One may argue that co-located gestural communication should only use well-established gestures. AR-HHDs may not support those, however, because users may struggle to look at both AR objects and their collaborators, as Rekimoto et al. [228] pointed out. Moreover, by holding an HHD, users cannot use their hands as they would otherwise do (e. g., body language); see Section 2.4.4. Past work [28, 62, 177] and myself within this PhD suggest to couple an AR-HMD with an HHD, which then can be lowered and held into one hand. Only one paper [177] coupled HHDs with SAR displays. As we state in Section 2.4.4, only more research can show how to combine HHDs and SAR devices w.r.t. the sensors and high display resolutions of typical HHDs.

Table 2.3: The selected surveyed papers.

Year	Reference	Space	Time	Role	Tech.	Output	Input
1996	Rekimoto [228]	co-loc.	sync.	sym.	sym.	AR-HHD	non-tracked controllers
1998	Benford et al. [26]	both	sync.	both	asym.	traditional screen	not described [1]
1998	Billinghurst et al. [41]	remote	sync.	sym.	asym.	AR-HMD / traditional screeen	mouse (and trackball)
1998	Szalavari et al. [255]	co-loc.	sync.	sym.	sym.	AR-HMD	hands tracking / tangible / touch
1998	Szalavari et al. [254]	co-loc.	sync.	sym.	sym.	AR-HMD	hands tracking / tangible / touch
1999	Billinghurst and Kato [43]	both	sync.	sym.	both	AR-HMD / traditional screeen	head gaze / mouse (and trackball)
1999	Butz et al. [62]	co-loc.	sync.	both	both	AR-HMD/AR-HHD/traditional screen	tracked controllers/keyboard&mouse/touch/tangible
1999	Kiyokawa et al. [151]	co-loc.	sync.	sym.	both	AR-HMD / VR-HMD	hands tracking / head gaze
2001	Billinghurst et al. [44]	co-loc.	sync.	sym.	both	AR-HMD / VR-HMD	tangible / non-tracked controllers
2002	Piper et al. [213]	co-loc.	sync.	sym.	sym.	SAR	tangible / mouse
2002	Schmalstieg et al. [239]	co-loc.	sync.	sym.	sym.	AR-HMD	hands tracking / tangible / touch
2003	Ou et al. [206]	remote	sync.	sym.	asym.	traditional screen	touch / mouse
2003	MacWilliams et al. [177]	co-loc.	sync.	sym.	both	AR-HMD / AR-HHD / SAR / traditional screen	hands tracking / touch / speech / tangible
2004	Benko et al. [28]	both	sync.	sym.	both	AR-HMD / HHD / traditional screen	hands gestures / touch / tangible
2006	Stafford et al. [247]	remote	sync.	asym.	asym.	AR-HMD / traditional screen	hands gesture / tangible / mouse (trackball)
2008	Wang and Dunston [278]	co-loc.	sync.	sym.	sym.	AR-HMD	tangible
2009	Dunleavy et al. [87]	co-loc.	sync.	sym.	sym.	HHD	touch
2009	Huynh et al. [133]	co-loc.	sync.	sym.	sym.	AR-HHD	tangible
2009	Lincoln et al. [172]	remote	sync.	asym.	asym.	SAR	none
2009	Nilsson et al. [196]	co-loc.	sync.	asym.	sym.	AR-HMD / AR-HHD	tracked controllers
2009	Oda and Feiner [201]	co-loc.	sync.	sym.	sym.	AR-HHD	touch
2009	Ong and Shen [205]	co-loc.	sync.	sym.	sym.	AR-HMD	tangible
2010	Prytz et al. [224]	co-loc.	sync.	asym.	sym.	AR-HMD	tracked controllers
2010	Wilson and Benko [291]	co-loc.	sync.	sym.	sym.	SAR	hands tracking / touch / tangible
2011	Maimone and Fuchs [179]	remote	sync.	sym.	both	SAR / traditional screen	none
2011	Wang and Dunston [279]	both	sync.	sym.	sym.	AR-HMD	tangible
2012	Oda and Feiner [199]	co-loc.	sync.	asym.	sym.	AR-HMD	tracked controllers
2012	Poelman et al. [220]	remote	sync.	both	asym.	AR-HMD / traditional screen	hand gestures /keyboard & mouse
2012	Gauglitz et al. [108]	remote	sync.	asym.	asym.	AR-HHD / traditional screen	keyboard & mouse
2012	Kasahara et al. [146]	co-loc	async.	sym.	sym.	AR-HHD	touch
2012	Mora et al. [191]	both	async. [2]	both	both	AR-HHD / traditional screens [2]	touch /keyboard & mouse [2]
2013	Adcock et al. [3]	remote	sync.	asym.	asym.	SAR	tracked controllers
2013	Ballagas et al. [20]	co-loc.	sync.	sym.	sym.	SAR / traditional screen	touch
2013	Dong et al. [84]	co-loc.	sync.	sym.	sym.	AR-HMD / traditional screen	keyboard & mouse
2013	Sodhi et al. [244]	remote	sync.	asym.	sym.	AR-HHD	hands gestures / touch
2014	Benko et al. [30]	co-loc.	sync.	sym.	sym.	SAR	hands gestures
2014	Gauglitz et al. [110]	remote	sync.	asym.	asym.	AR-HHD / traditional screen	keyboard & mouse
2014	Gauglitz et al. [109]	remote	sync.	asym.	asym.	AR-HHD / traditional screen	touch
2014	Kim et al. [149]	remote	sync.	asym.	asym.	AR-HMD / AR-HHD / traditional screen	mouse
2014	LeChenechal et al. [162]	remote	sync.	asym.	sym.	VR-HMD	tracked controllers
2014	Lehment et al. [169]	remote	sync.	sym.	sym.	AR-HMD	none
2014	Wang et al. [280]	remote	sync.	sym.	sym.	AR-HMD / traditional screen	hands gestures / touch
2014	Irlitti et al. [135]	co-loc.	async.	both	sym.	AR-HHD	tangible
2015	Oda et al. [200]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD / HHD	touch / tracked controllers
2016	Gupta et al. [117]	remote	sync.	asym.	asym.	AR-HMD / traditional screen	eye gaze /keyboard & mouse
2016	Pejsa et al. [208]	remote	sync.	sym.	sym.	SAR	none
2017	Clergeaud et al. [68]	co-loc.	sync.	both	both	VR-HMD / SAR	hands tracking / tangible
2017	Komiyama et al. [155]	both	sync.	both	both	CAVE	non-tracked controllers
2017	Lee et al. [165]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands tracking
2017	Muller et al. [194]	remote	sync.	sym.	sym.	AR-HHD	touch
2017	Piumsomboon et al. [219]	remote	sync.	sym.	asym.	VR-HMD / AR-HMD	hands gestures / head gaze / eye gaze
2017	Roo et al. [234]	co-loc.	sync.	both	both	VR-HMD / SAR	hands tracking / tracked controllers
2018	Lee et al. [166]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands tracking
2018	Piumsomboon et al. [217]	remote	sync.	both	asym.	VR-HMD / AR-HMD	hands tracking / hands gestures / head gaze
2018	Poretski et al. [221]	co-loc.	sync.	asym.	asym.	AR-HHD	touch
2018	Huang et al. [131]	remote	sync.	asym.	asym.	AR-HMD / traditional screen	hands tracking / touch
2018	Teo et al. [259]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands gestures
2019	Bhattacharyya et al. [40]	co-loc.	sync.	sym.	sym.	AR-HHD	touch / tangible
2019	Teo et al. [258]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	hands gestures
2019	Piumsomboon et al. [218]	remote	sync.	both	asym.	VR-HMD / AR-HMD	tracked controllers / hands gestures / tangible
2019	Piumsomboon et al. [215]	remote	sync.	asym.	asym.	VR-HMD /AR-HMD	hands gestures / head gaze / eye gaze
2019	Mahmood et al. [178]	remote	sync.	sym.	sym.	AR-HMD	hands gestures / speech / head gaze
2019	Thanyadit et al. [268]	remote	sync.	asym.	asym.	VR-HMD / AR-HMD	none
2019	Franz et al. [105]	co-loc.	sync.	sym.	asym.	AR-HMD / traditional screen	head gaze / non-tracked controller
2019	Dolata et al. [83]	co-loc.	sync.	asym.	sym.	SAR	tangible / touch
2020	Reipschläger et al. [227]	co-loc	sync.	sym.	sym.	AR-HMD / large displays	touch
2021	Büschel et al. [60]	co-loc	sync.	sym.	sym.	AR-HMD	hands gestures
2021	Langner et al. [159]	co-loc	sync.	sym.	sym.	AR-HMD / HHD	touch

[1] The paper does not describe how they can interact with the virtual world. We suppose it is through hand tracking and mouse/keyboard but we are unsure.
[2] Traditional tools are included because their system allows the sending of emails visualizable "anywhere". We did not consider its synchronous case because it relies on video conferencing without AR support.

2.7 Collaborative Immersive Analytics (CIA)

Quite a few of the papers we surveyed tackle *analytics and visualization* tasks (our domain expertise) using AR. For this reason, we provide an overview of CIA next.

2.7.1 Immersive Analytics

Immersive Analytics (IA) is at the junction of MR Environments and Visual Analytics [269] (Figure 1.3). AR can be a medium to support users in their analytical tasks [25, 91, 185], collaborative or not. Such systems need high computational power for both simulations and rendering to provide quick feedback. IA projects use VR platforms more often than AR ones, and only a few projects investigated IA using AR [99]. This lower interest for AR in IA may be explained by the fact that VR platforms are more mature and easier to develop for and by the higher constraints of AR devices, notably regarding tracking. We also hypothesize, regarding hardware platforms for IA, that

SAR is similar to the CAVEs used by VR researchers which, due to their size capacity, allow users to see each other and to use personal devices, e. g., laptops;

HHDs have limited value for 3D data visualization because they are non-stereoscopic and, compared to a desktop machine, are less convenient to use because they are often less powerful for rendering, have smaller screens, lack a keyboard & mouse (useful for scripting), and have less mature software tools for analytics; and

AR-HMDs have lower computing power than VR-HMDs by being embedded systems, while having more constraints than HHDs to bring immersion (e. g., higher resolutions and frequencies, lower weight). This is, however, changing with the arrival of new powerful AR-HMDs, such as the Microsoft's HoloLens 2 with its Snapdragon 850. Indeed, four immersive analytics research papers we surveyed [60, 159, 178, 227] have been published after 2019 and use a Microsoft's HoloLens (either 1^{st} or 2^{nd} gen.). The only other system that supports immersive analytics not relying on the HoloLens is the *Studierstube* system [239, 240, 255]. Compared to VR-HMDs, moreover, AR-HMDs might be considered for situated analytics (e. g., medical data), such as Büschel et al.'s system [60].

2.7.2 AR Collaborative Immersive Analytics (AR-CIA)

Collaborative Immersive Analytics, which Billinghurst et al. [42] define as *"the shared use of new immersive interaction and display technologies by more than one person for supporting collaborative analytical reasoning and decision-making,"* is a field intersecting CSCW, Visual Analytics [269], and Immersive Environments (Figure 1.3). At the intersection of CSCW and visual analytics lies *collaborative visualization*, which Isenberg et al. [136] define as *"the shared use of computer-supported, (interactive) visual representations of data by more than one person with the common goal of contribution to joint information processing activities."* At the intersection of CSCW and MR Environment lies *MR Collaborative Environments*. *MR environments* intersect both visual analytics and MR environments.

Although researchers studied many scenarios in the past (e. g., security [196], archaeology [28], engineering [278]), visualization has thus far been largely neglected over the last decades. This trend is now changing as we found recent research projects focusing on AR-CIA. This new trend appeared at the same time I started my PhD, i. e., in 2019 [60, 159, 178, 227]. However, it seems that researchers, for the moment, are focusing on abstract data visualization and not scientific data visualization, which I am focusing on within this PhD.

2.7.3 Should 3D Visualization be Immersive?

Past research (e. g., [102,132,157]) showed that immersive technologies can significantly enhance the way people understand 3D data compared to traditional interfaces.

AR technologies seem to be promising for 3D data visualization [88,181]. By merging the real and virtual worlds, users can use their common analytical tools while visualizing in stereoscopy and using the "huge" (e. g., room-scale) workspace AR provides [283, 284]. As discussed in several education papers (e. g., [84]) and through this survey, students understand concepts better using stereoscopic 3D visualizations which reduce their cognitive workload. Networks [286] and 3D vector fields [102] are better perceived using stereoscopy and motion cues which AR relies on. AR could also be appropriate in 3D visual analytical tasks [278, 279], with AR reducing the time collaborators take to detect errors in 3D designs compared to traditional tools.

As stated in Section 1.1.1, however, immersive technologies are not always better than traditional workstations. While my PhD gives some thoughts about the potential of immersive technologies for (collaborative) visual analytics, more research is required to better understand immersive technologies for scientific workflow, as Immersive Analytics, especially for AR, is really new. In this PhD, I provide insights about whether 3D visualizations should be immersive or not within my system (see Chapter 3) I am developing in collaboration with HZG.

2.7.4 CIA Research Questions

Because the AR-CIA field is still new, we propose in this section some relevant research areas to be investigated. While this section is part of the original TVCG publication, it was mostly driven by my PhD topic in addition to the taxonomy used in this work. Most of these research questions were considered in the system I developed throughout this immersive PhD journey.

AR Devices?

The five papers we reviewed concerning immersive analytics used AR-HMDs as a support. Reipschläger et al. [227] used those devices to augment large displays and Langner et al. [159] to augment multi-touch tablets. Büschel et al. [60] used AR-HMDs to put 3D spatio-temporal data where they were captured. While Mahmood et al. [178] also used AR-HMDs for remote data analysis collaboration, they did not give strong reasons of why they have used AR-HMDs over other forms of immersive devices (e. g., VR-HMDs). The Studierstube [239, 240, 255] used AR-HMDs which allowed users to keep their interactive devices (personal pads) and their collaborators in sight.

Because perspective projections have multiple parameters (e. g., far/near clipping plane, FoV), users often rely on orthographic views for some 3D tasks, e. g., drawing 2D lassos being extruded in 3D [38] or measuring distances in CAD software. AR-HMDs, however, lower this limitation by having perspective and stereoscopic parameters matching those of the human vision system. However, such wearable devices currently have lower resolutions and computing power than workstations and scientists still need to use workstation tools and software [283,284]. While a hybrid AR–workstation system might solve each other's issues, however, it may remove the AR interaction benefits due to people following social rules and their limited mobility in their environment; see also Wang et al.'s system [284] targeting particle physicists which I co-authored. Questions regarding which AR device or which hybrid interface to use for AR-CIA remain to be explored, and surely depend on the domain.

Some researchers are investigating the use of AR-HMDs combined with large displays [227] or multi-touch tablets [159]. The collaborative system I developed for this PhD relies on AR-HMDs and multi-touch tablets (Chapter 5). But compared to Langner et al. [159] and Reipschläger et al. [227], I am augmenting the 3D data rendered on AR-HMDs with additional devices instead of augmenting 2D data rendered on 2D screens with AR-HMDs. My method decouples the visualization space that the AR-HMDs provide from the interaction space the tablets provide. Users can then, if required, only follow the discussion by just focusing on the AR-HMDs without being disturbed by the interactive capabilities of the system, while other collaborators can manipulate the environment using the multi-touch tablets for which they are trained to use, which is not the case of, e. g., visitors at museum. This can be similar to slideshow presentations where only the active user sees the interactive environment while the passive users follow the discussion based on the slides where the interactive interface is hidden. This new trend from the data visualization community and this PhD provides some indications about this general research question.

The 3D Coordinate System, AR Display + HHD

Lopez et al. [175] combined a S3D and a multi-touch tablet to propose user-friendly 9 DoF (3D translations, 3D rotations, 3D scaling) transformations. However, in CSCW contexts, there are multiple users, hence multiple local 3D coordinate systems. Should the system select a particular collaborator's coordinate system for pointing actions? Should one visualize a dataset with the same relative position that another user has to understand the discussion? To track users, should the tablet represent the others' relative position? What features may be required when using this mobile device as a rendering, a touch input, and a tangible input device? In this PhD, I provide insights about the use of this mobile device as a rendering, touch input, and tangible input device simultaneously for a one-user environment (Chapter 5). The specific use of those mobile devices in a collaborative environment, however, remains to be studied, where awareness of others is important, and where parallel work should be considered.

Spatial Selections

Filtering Regions-of-Interest (ROIs) and specific features are common interactions in 3D visualization software and were studied a lot; see Besançon et al.'s taxonomy [38]. They were not studied, however, in collaborative contexts, with users being either in AR or VR. How can we use these techniques in AR-CSCW where collaborators must understand the actions of others? Should a selected region be displayed only in private or subjective views? If yes, referencing can be challenging in scalar fields as discussed next. Moreover, compared to VR-based systems, users can easily move around, which can be a rich source of awareness.

In this PhD, I first studied collaborative point specifications to understand which paradigms, e. g., direct manipulations or remote raycasting, have what effects on the collaboration (see Chapter 4). I then studied in a one-user environment the ROI selections using a multi-touch tablet as a tangible object (see Chapter 5). I could not provide such a study for collaborative work as there were no basic understanding of ROI selections in immersive environments to begin with. However, both studies are linked together, as tangible tablets can be used in a remote or direct fashion similarly to pointing actions using hand gestures. I hypothesize then that awareness cues are similar in both cases and provide a joint discussion based on both studies (Chapter 7).

Subjective Views and Immersive Analytics

Smith and Mariani [243] proposed two dimensions for subjective 3D: *appearance* and *modifier*.

In our survey, Nilsson et al.'s [196] system proposed subjective object representations based on the users' roles (appearance dimension). The VR literature also contains work on subjective view [4, 243] and showed some potential issues regarding workspace coherency. However, we did not find studies related to 3D cloud points or volumetric datasets which are explored via a set of filters or transfer functions. These filters significantly alter the transparency (modifier dimension) of some part of the data and not the other, reducing the workspace coherency.

The same phenomenon can arise for 2D visualizations. Reipschläger et al. [227], e. g., use some private 2D filters, making the visualizations those filters are attached to subjective. Indeed, their AR-HMDs render new contents or modify the visualizations the large display renders, making those visualizations different per user. However, they gave no answers about the workspace coherency issues when collaborators are working together for, e. g., sharing ideas using such subjective visualizations as the support of the discussion.

It is then unclear if such subjective filters can be efficient and with what visualization and interaction techniques, e. g., AR devices could switch to one of the collaborators' parameter sets when detecting communication cues (e. g., pointing). Similarly, should such a system use a UI that lists all registered parameters (e. g., one parameter per user plus one default) and select one before starting the discussion? In Chapter 6, I discuss and give design considerations about the use and implementation of subjective views for 3D data analysis. My co-authors and I, however, could not study this functionality due to the COVID-19 pandemic and its resulting restrictions.

2.8 Limitations

The work we surveyed is scattered among several venues, explaining why only 18 ($\approx 26.5\%$) papers we discussed are from our systematic search. Our classification might thus not entirely reflect the real breadth and depth of the literature. Hence, the design considerations may be incomplete; and Section 2.6 discusses on-going research areas. We argue, however, that—while not exhaustive—we have surveyed a large part of the literature so that our discussion of research questions and design considerations are relevant. We also note that some of the surveyed papers may be subject to interpretation. For example, while Lehment et al.'s work [169] was tested with HMDs, the authors claimed that their algorithm is applicable to HHDs. Moreover, the difference between *tracked controller* and *tangible* that we used in our taxonomy is rather small and could be confusing at first. Similarly, *CroMAR* [191] allows users to share their viewpoints by emails which can be visualized in traditional screens, hence their inclusion as output devices.

2.9 Conclusion

In this chapter we first reviewed fundamentals regarding AR-CSCW, technologies that are used to provide a coherent and effective AR experience, and psychological aspects that designers should pay attention for. We then reviewed 68 papers (18 through a systematic review and 50 additional ones) regarding collaborative work using AR, alone or not (e. g., combining AR and VR), and categorized them along our six dimensions. Based on this review, we extracted a current research agenda, discussed guidelines derived from the literature, and mentioned remaining research areas that yet have to be covered. In addition, we contemplated collaborative immersive analytics using AR technologies. With this survey we thus aim to give an overview of the field for newcomers, researchers and practitioners alike, and up-to-date information for domain experts.

This survey, especially the discussion and the research questions concerning Collaborative Immersive Analytics (Section 2.7), gave me cues at the beginning of my journey about what projects to tackle, and what I have to consider when developing a collaborative AR software for Immersive Analytics purposes. Next, I change my focus from the AR worlds we surveyed to my own AR space, where I describe which Immersive Analytics scenarios I am targeting at and the associated software I developed, and justify the design choices I made. Based on this software, I worked on three research topics that I describe in the subsequent chapters.

3

CHAPTER

A New Collaborative Immersive Analytics Environment

While the survey targets AR-CSCW in its whole, the discussion around Collaborative Immersive Analytics in Section 2.7 gave me multiple cues for the AR-CSCW application that I created and increased in functionalities as the only author during those past three years. I remind the reader that I am interested in collaborative 3D scientific data visualization and exploration.

In this chapter, I describe the application I created in its whole, and justify its design and my choices based on the literature and the incremental design steps I performed. I do not, however, describe the specific functionalities concerning the respective projects I developed these functionalities for. I instead describe them in their respective chapters which I refer where needed. This chapter then extends the general introduction (Chapter 1) of this thesis that already gives some justifications for my application.

I developed my AR-CSCW application with the help of oceanographers from HZG, Germany. My main contact was Burkard Baschek, director of the Institute of Coastal Ocean Dynamics at the Helmholtz-Zentrum Geesthacht research institute, Germany. From a survey I designed with the help of Burkard Baschek, Petra Isenberg[1], and Tobias Isenberg, we saw that those oceanographers are not efficient with their current tools to share ideas and explore their 3D datasets together. Based on heavy discussion with Burkard, I developed an AR-CSCW system that relies on AR-HMDs and multi-touch tablets. The final purpose of such a system is to support those oceanographers during their meetings.

This application and the design choices I made are a contribution on their own right that I would like to highlight. For instance, many types of hardware combinations exist in the AR-CSCW research area, but, as stated in the survey (Section 2.6.6), too few projects justified their decisions concerning their choices of output and input devices they relied on. Moreover, I also describe and discuss in this chapter the results I got from a pilot study that I ran with two oceanographers using my system. Welcome into my augmented world!

3.1 Introduction and Description of the Problem

My colleagues at HZG are interested in investigating how eddies (Figure 3.1) influence the weather in general. Their work concerns the global weather around the world and weather specific to a local region. For what concerns this PhD, they are mostly interested in oceanic eddies that have a short lifetime, about few hours, with a particular focus on the Cabo Verde area.

To study those oceanic eddies, they first create and study a forecast simulation of the area of interest based on physical models. They then intensively study the results of the simulation to understand what is happening, and whether or not the simulation computed something plausible.

[1]Université Paris-Saclay, LISN, CNRS, Inria, France

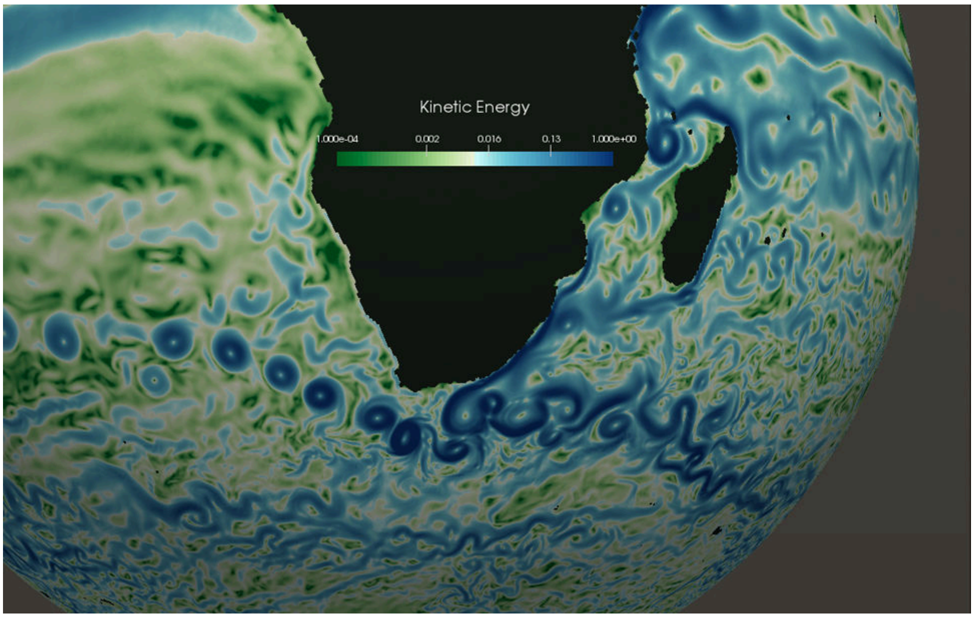


Figure 3.1: A three-dimensional image of the Agulhas current and retroflection in South-Africa. Large eddies can be seen around the Horn of Africa into the South Atlantic which transfer heat and other constituents between the oceans. Image taken from Samsel et al. [237].

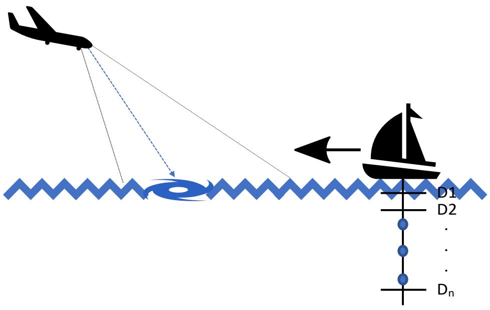


Figure 3.2: A sketch summarizing the expedition processes to measure eddies in-situ. The plane hovers above an area of interest. Using sensors, the plane detects an eddy and gives its position to the ship, which moves toward the eddy and makes measurements along the depth axis with its sensors.

Based on the gathered insights, they then prepare an actual expedition (see Figure 3.2) to measure directly the oceanic properties of the area of interest. The expedition involves two teams. The first team is on a plane and hovers above the area of interest. Using multiple sensors such as IR cameras, the plane measures mainly the velocity and the temperature of the surface-level water. Based on what they measured, the crew on the plane asks the second team, on a ship, to go to the targeted area and make on-site measurements with their equipment. The sensors the ship embeds, placed along a cable, measure multiple properties along multiple depth-levels simultaneously (e. g., the water temperature, speed, and salinity). Based on the measurements, the scientists update the simulation and determine the next place to go.

After the expedition, that can last a day, those scientists discuss about their findings and the difficulties they encountered during the expedition in order to (1) understand better the effects of the oceanic eddies and (2) to improve themselves for the next expedition, e. g., improving how the ship approaches the eddies to depart correctly for the next area of interest, to not lose more time than necessary in maneuvers.

However, as I explain in Section 3.2, these scientists mostly rely on static visualizations shared through slideshow presentations during those meetings. They rely little to not-at-all on exploratory tools. In these meetings then, only the data that is pre-computed is shared. However, as the discussion advances, it is not rare that one wants to look at additional aspects of the data, hence the need for exploratory tools.

In this project, I then rely on AR-HMDs and multi-touch tablets (see Section 3.3) to improve the efficiency of their meetings. More specifically, I aim at supporting users at understanding, discussing, and exploring collaboratively both the forecasting simulations and the actual measurements that were made in-situ during those expeditions. While in this PhD I focused on oceanography, the insights I gathered throughout the design and the evaluation of the application (Section 3.6) might serve other domains such as medicine and particle physics.

3.2 Understanding Users

From previous discussions with Burkard Baschek, it seems that oceanographers at HZG perform a lot of pre-processing steps on desktop machines or even supercomputers before meeting with their colleagues and discussing about their datasets. Those discussions gave me a general picture of the problem that allowed me to start designing a possible solution. Later, to better justified my solution and to refine it as part of an iterative design process, I designed, in collaboration with Burkard Baschek, Petra Isenberg, and Tobias Isenberg, an online survey that we broadcast to the people at HZG working on ocean dynamics to better understand the needs of a broader audience; see Appendix A. We released the survey in May 2021.

At the time I am writing this PhD, three domain experts answered the survey. While three seems a pretty low number, I would like to emphasize that those users are domain experts which possess good understanding of their workflow and associated problems, and that this work is still going on, as I was not able to conduct a complete user study with the complete application due to the COVID-19 pandemic and its associated problems with users' experiments and travels. While I did not pre-register the survey, I originally expected a total of six-to-eight users to fill the questionnaire. Given the difficulties to meet remotely domain experts, I preferred sharing a questionnaire instead of doing remote interviews to better suite their calendar. I might, however, plan private interviews later as the project continues to gain more details from the answers my collaborators and I got.

To simplify the reading of this chapter, I group and describe separately next the results of the survey based on their categorizations, where I extract multiple design goals.

Meetings

These three users meet usually once a week on average, with multiple other persons. Those include students and Postdoc researchers, co-authors, and sometimes other non-scientist persons from the department (including administrative persons) who do not all possess all the required scientific background to understand a specific project, but still need to understand some principles and insights to, e. g., make communication supports. The meetings seem to rely mainly on sketching on whiteboards and slideshow presentations. Throughout the meetings, it seems that these users mainly take notes and make multiple sketches throughout the meetings to keep a trace of what have happened. I am purposefully ignoring results that say that those domain experts currently rely on virtual platforms (e. g., Skype/Zoom) to meet together due to the COVID-19 pandemic, as they would normally, if allowed, perform the meetings on-site.

With such a large group and heterogeneous background, two of the three users stated that their ideal meeting tools would allow for "large group meetings for overall discussion and presentations, then breakout groups (size up to 10[or 5]). [The software would allow] afterwards [for] a synthesis in a large group again." These meetings serve principally today to share insights one has found to others, and how those insights can support a given person in his/her current work. Ideally, however, those meetings would also serve to advance in common projects by brainstorming theories to better understand phenomena for which no answer has been provided yet. However, based on the participants' comments, it seems that those meetings are not efficient for both requirements, as they "have no clue" on what is going on, and that these meetings, at the end, "are often more to update each other than advancing on the work", with sometimes "experts with a certain expertise [who] are not always good in explaining results.” Mostly, this is due to data being hard to grasp and understand at first glance, and time constraints.

I thus derive the following design goals:

G1 The CSCW system should target co-located collaborators.
G2 The CSCW system should allow for more than five simultaneous users.
G3 The CSCW system should not hinder users to take notes and to use other tools such as whiteboards.
G4 The CSCW system should allow for data exploration in order to “advance on the work.”

Data Types and Visualizations

The data those domain experts contains many physical properties. The participants listed the followings:

- Direct values: temperature, velocity, pressure, density, turbidity, wave heights, and storm frequency.
- Derived values: turbulence and buoyancy frequency.

The three participants of the survey stated that they usually work on a subset of their datasets (usually due to time constraints) that they have pre-processed. It seems, from the answers, that the data is usually shared in the form of graphs. For meetings involving only scientists, data can also be shared as spreadsheets that scientists can plot as part of their own routines.

In their current workflow, these scientists look at them side-by-side or use *transect* visualizations. Their datasets are usually in 2D (e. g., images) or 3D (e. g., simulations) with an optional time component. These users avoid merging those properties as it can smooth or hide important properties. Their main issues, however, is to relate 3D datasets with 2D ones, and to relate multiple timed datasets with each other as they may not have been calibrated correctly (difference in time resolutions and starting time).

All the participants agreed that visualizations aid greatly in their meetings, as they allow “to present high-resolution results” and aid users’ “own understanding and capabilities to interpret the data, but also help to present it effectively and clearly to others.” Yet, those visualizations are “mostly 2D graphs”, usually created “during data processing, data analysis, and for presentations”, i. e., when the scientists prepare their meetings and not during them. The number of visualizations users look at may vary depending on the meeting, with difficulties at finding an “ideal presentation of a 3D environment including hundreds of thousands of data points.”

I thus derive the following design goals:

G5 The CSCW system should support users at understanding 3D datasets.
G6 The CSCW system should allow for side-by-side comparisons of 2D and 3D datasets.
G7 The CSCW system should facilitate spatial relationship between 2D and 3D datasets.
G8 Datasets are expected to be pre-processed prior of the meeting.

Ideal Software

One participant stated that “the ideal system should be flexible, useable with different programming languages (different members of [their] team code in different languages), able to work with data input in different formats, allow easy data formatting within the system, and allow output of data and visualizations in different formats. Most importantly: it should be easy to learn as a system that requires a lengthy training process will be unlikely to convince people to move away from what they know. Especially seeing as nobody has the time for lengthy training, a way to start the new program and be able to work with it right away would be vital.” I do not know, however, if these statements concern a collaborative tool for meetings or if they concern a remote collaborative environment where users work alone to, e. g., prepare a meeting.

Another participant stated that the environment should allow for joint and subgroups data analysis and discussion. This aspect is important as the meetings can regroup 5 to 10 persons (see Section 3.2) where the work can sometimes be parallelized.

I thus derive those final design goals:

G9 The CSCW system should allow collaborators to split into multiple subgroups.
G10 The CSCW system should be easy to learn.

To answer these design goals, I describe and justify next the AR-CSCW system I developed, which mainly relies on one AR-HMD and one multi-touch tablet per user.

3.3 AR-HMDs plus Multi-Touch Tablets CSCW Environment

The main need from the participants who answered the survey concerns the collaborative 3D data scientific visualization (**G5**) in co-located settings (**G1**).

Multiple pieces of related work highlight the use of immersive technologies to visualize such data; see Section 2.7.3. As users are working together, they should be aware of their co-workers to collaborate efficiently. The different immersive systems found in the literature (see the output dimension of the survey; Section 2.3) do not give the same level of immersion (e. g., FoV, resolution, head-tracking, isolation) and the same level of *passive* awareness cues that do not involve the technology, e. g., seeing the collaborators' bodies through AR-HMDs. Those technologies do not also share the same specifications (e. g., computing power and the restricted physical environments they can run into).

While collaborating, the users I am targeting need their usual tools (**G3**, external software, e. g., spreadsheets), so a complete VR environment such as VR-HMDs would not be suitable. In CAVEs, users can bring laptops and see each other. However, only a limited number of users can use a CAVE due to the limited refresh rate of its projectors, which does not comply with **G2**. This limitation also applies to Spatial Augmented Reality (SAR) environments which also rely on projectors. However, SAR covers bigger areas (e. g., a whole room) compared to CAVEs which are limited to the size of their screens, usually forming a cube of $\approx 9m^3$ [73]. In comparison, SAR environments allow the use of usual tools put anywhere inside the room, e. g., whiteboards. Finally, AR hand-held displays are too limited in size and computing power, and do not propose any 3D rendering. They do not answer **G5** and barely support **G4**. Their only benefits lie in their availability in the consumer market, their touch input interface, and their mobility, i. e., users can move around while manipulating a multi-touch tablet. To comply with most of the design guidelines, I then propose to rely mainly on AR-HMDs as I show next.

3.3.1 AR-HMDs as the Main Devices

AR-HMDs do not inherently restrict the number of simultaneous users, as users possess their own dedicated device (**G2**). Users can also see each other and their real environment without artifacts, allowing the use of traditional tools (**G3**). AR-HMDs are also not restricted to where they are used into, while CAVEs and SAR systems need time to be reconfigured when moved. Moreover, by relying on AR devices as opposed to VR ones, users can use their room as the whole workspace which allows to place multiple visualizations to visualize all (or a subset of) those stated properties (**G6**). Moreover, this large workspace allows users to separate themselves into multiple subgroups and join together in a seamless manner (**G9**).

They, however, have issues with color fidelity for the moment [287], limiting **G5**. Moreover, AR-HMDs are usually embedded systems with a limited computing power, even if it tends to increase over the time at, now, the same rate as smartphones and tablets. For instance, at the time I am writing this PhD (August 2021), the latest Microsoft's HMD, the HoloLens 2^{nd} gen. released in 2019, includes a SoC Snapdragon 850 comparable to the high-end SoC Snapdragon 845 that was released one year earlier, which, e. g., the smartphone Samsung Galaxy S9 embeds.

I then propose to rely mainly on AR-HMDs for their benefits for collaborative immersive analytics. However, AR-HMDs do not fit every immersive analytics scenario because of their limited computing power, as the actual 3D datasets are big in data size and require a lot of pre-processing and a sufficient graphical power to efficiently visualize them in real time. Similarly, in a project I co-authored with Wang et al. [284], we showed that, while AR-HMDs have benefits, they cannot replace desktop machines which contain (1) all the exploratory tools scientists require in their daily tasks (**G4**), including scripting tools such as Python, Matlab, and R, and (2) the computing power to do so in a relatively small-time frame. Other multiple past projects then relied on hybrid systems using both an immersive device and a desktop machine or laptop [28, 50, 180] to have the computing power to run simulations, and to have a more effective user interface which relies on mouse+keyboard paradigms.

However, my targeted participants pre-process their data before meeting together (**G8**). So, while they need to explore their datasets during meetings, they do not need as much computing power and simulation tools as they would while pre-processing those data. Moreover, in Wang et al.'s [284] study and Bach et al.'s [16] study, users generally preferred to stand up when engaged with their immersive environment. Those users did not interact much with the visualizations and explore them really slightly. They were, however, benefiting from the immersive technologies to understand their data using the 3D coordinate system such technologies propose, as compared to 2D perspective or even orthographic views that scientists currently rely on [38, 284]. Finally, desktop machines might not fit collaborative scenarios. Indeed, in co-located settings (**G1**), only one user at a time can interact with the system. As such, the survey shows that the oceanographers I am targeting rely mainly on slideshow presentations and static visualizations, using projectors, instead of using computers during their meetings. Finally, desktop machines cannot merge the workspace and the collaborative spaces, as one cannot see and interact simultaneously with the visualizations and the collaborators without changing their focus, impacting the discussion.

Based on these observations, I propose as a general guideline to rely on AR-HMDs for users that already know their datasets and have already performed some pre-processing computations (**G8**) before meeting with their co-located colleagues (**G1**). The immersive application, however, still needs basic exploratory tools (**G4**) for users to, e. g., show particular features to colleagues, to benefit from the immersive technology to understand better the spatial relationship of their data (**G5** and **G7**), and to try new ideas and compositions as the discussion moves on. The meetings then serve as events where collaborators discuss and share insights with each other, where 3D visualizations might help, and to debate on the problematics those users are interested in. Moreover, as discussed in Section 1.1.1, I argued that the collaborative, output ($\mathcal{V}$ and $\mathcal{O}$), and input ($\mathcal{I}$ and $\mathcal{M}$) spaces are merged together, based on Bruckner's et al.'s model [57] (see Figure 1.4). With such merging, users are all engaged and they can simultaneously interact with and see the visualizations, while still having their collaborators in sight.

In this section, I mostly discuss about the output capabilities of AR-HMDs. To provide interaction (**G4**), I am coupling them with multi-touch tablets as I explain next.

3.3.2 Interactive Tools: Multi-Touch Tablets

After further discussion with Burkard, and by looking at what current exploratory software proposes, it seems that scientists in their work need several basic tools and functionalities, including, but not limited to:

- Open files.
- Create multiple visualizations based on those files to perform side-by-side comparisons.
- Rotate, scale, and translate those visualizations in the 3D space.
- Change the visualization parameters:
  - Parameterize the transfer functions of volumetric renderings.
  - Change the color scheme (e. g., rainbow mapping, divergence mapping).
  - Change clipping planes.
  - Select a new time frame and play an animation.
  - Select a region-of-interest to further explore.
- Mark specific positions and add annotations, e. g., sketches and texts.
- Add and merge 2D information such as 2D images, coastlines, map, and path lines.
- Perform simulations such as computing pathlines for, e. g., physics particles simulations.
- Save the current visualizations for, e. g., scientific publications or presentations.
- Apply data filters.
- . . . And many more.

Usual AR-HMDs today are usually limited to mid-air gestures, voice, and eye-tracking interfaces. Mid-air gestures, the main interface of AR-HMD, is tiresome [128]. As meetings can last multiple hours, I do not foresee to rely on such an interface for most interactions, as the needs are numerous and require, for some of them (e. g., manipulating the transfer functions, moving visualizations around, drawing annotations), continuous interaction that the other two interfaces cannot provide efficiently.

A multi-touch tablet can solve most of these problems. Its Graphical User Interface (GUI) can give general functionalities to handle the visualizations (e. g., opening, deleting, and cloning visualizations, moving the clipping planes, parameterizing the transfer functions). Moreover, it can be lowered and set at a comfortable position to reduce the torque on the users' arms. Finally, it can enable new kinds of interaction, such as 3D and tangible interactions (see also Chapter 5).

However, users cannot enter texts on tablets as efficiently as on computers which use keyboards. While keyboards can be used on multi-touch tablets, they are smaller than those of workstations and they would require users to pose the tablet on a solid surface (e. g., a table), as tapping on a keyboard in an empty 3D space is difficult. All the scripting should then be done on a dedicated workstation before a meeting, which **G8** suggests.

The tablets then give exploratory data analysis tools, but on a smaller scale compared to workstations. This minimalistic exploratory environment, however, might be easier to learn compared to other more complete tools which would possess way more filters and other exploratory tools (**G10**), and may suit meetings where not everyone is familiar to use the same tools.

However, the tablet can act as a personal device. The merging between the interactive (tablet) and the output (AR-HMDs) spaces is then broken when both devices are used together. Because of this break, collaborators might lack passive awareness cues about one's doings, such as what objects is a user currently manipulating on his or her tablet. Active awareness cues are then necessary and should be considered.

3.3.3 Their Joint Use

But while the tablets give interactivity, in regular meetings, not everyone needs to actively explore or create new visualizations. Some *passive* users, e. g., the communication department, only need

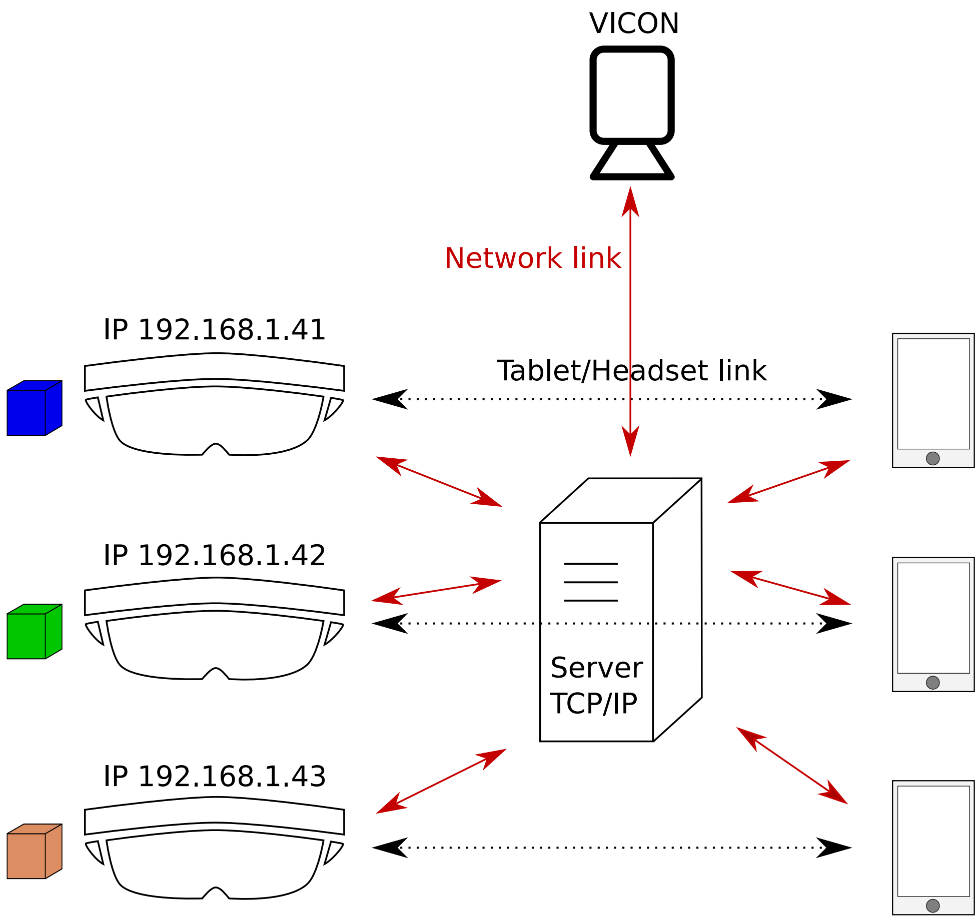


Figure 3.3: Sketch summarizing the architecture of my overall system. All the network packets transit from the server.

to understand the dataset and not interact with them. As it is for slideshow presentations, the AR-HMDs would allow such a behavior, as most of the interactions are delegated to the multi-touch tablets. The AR-HMDs are then mainly used for visualizations, similarly to projectors used for slideshow presentations.

While the AR-HMDs have benefits (e. g., 3D rendering, the use of traditional tools, and the merging of collaborative, output, and input spaces), the solution I am proposing can be similar to other desktop environments for what concern the merging of 2D and 3D datasets, and the merging of multiple timed datasets (**G7**), as the technology itself does not propose an inherent solution as it does for 3D visualizations (with, e. g., motion parallax and stereoscopy depth cues). Moreover, I expect users to pre-process their data beforehand, as AR-HMDs + multi-touch tablets would unfit exploration based on programming tools. Such an environment would then not fit complete remote asynchronous collaboration where users try to understand their datasets with heavy filtering tools. To illustrate this new hybrid environment, I describe next the functionalities of the system I created throughout this PhD to attempt to solve the needs of those oceanographers.

3.4 Main Components

I base my collaborative environment on four main components (Figure 3.3). The two first ones concern the AR-HMDs and the multi-touch tablets. Then, I use a remote server to handle the communication between all the devices. This design allows users to connect and disconnect themselves from the workspace as the server stores the current status of the session. The final component is the VICON tracking system which tracks, in the users' 3D space, the multi-touch tablets relatively to their AR-HMDs. While this might hinder the characteristics of AR-HMDs to run in every room space without calibration, I emphasize that technologies to enable the tracking of tablets with standalone AR-HMDs already exist. Indeed, some standalone VR headsets such as the Oculus Quest can track its remote controllers by tracking their IR LEDs using the constellation tracking algorithm [104]. Currently, the VICON tracks only one user as I used it only for the Tangible Brush project (Chapter 5).

The server, the HMDs, and the tablets possess a version of the data needed during the session. While the server might stream the data or the resulting visualizations to all the other devices, it is to note that the data can be large (e. g., 1.6 gigabytes for seven timesteps for the data I gathered from HZG), with possibly millions of data points (five millions in my case). A dedicated smart network structure and protocol is required to enhance that part, which is outside the scope of this PhD. For the moment, I only handle binary VTK files representing VTK_STRUCTURED_POINTS 3D objects, and a custom format to handle cloud point datasets I used for a dedicated project (see Chapter 5). I used the documentation VTK provides to create my own parser in C/C++ for all the devices. To handle time, a dataset comprises multiple additional files whose names are followed by .X, X being the time entry. For example, the data "data.vtk" might comprise multiple files called "data.vtk", "data.vtk.1", "data.vtk.2", etc.

I emphasize that my application is a research prototype. Some basic functionalities may then be missing. This application, however, serves as a proof-of-concept and gives insights to future designers that want to create their own applications whose targets are closed to what I am targeting here. It is also the only software I know that allows the collaborative visualization of volumetric datasets using AR-HMDs.

I describe next each part of the system in more details. For what concerns the specifications of 3D points (Chapter 4) and volumes (Chapter 5), and the abilities to render different visuals, one per user, for a given 3D visualization anchored at a given 3D position (Chapter 6), I encourage the reader to read the corresponding chapters.

3.4.1 The Server

The server handles the communication between all the devices and saves the status of the current session. This architecture allows devices to disconnect and reconnect whenever they want, as the current status of the environment can be resent upon request. The server runs on a linux distribution. I developed it using C++ and the standard library of UNIX systems (unistd.h). Except for the communication with the VICON which relies on the VRPN [257] protocol, the server does not rely on third-party libraries that I did not write myself.

Every component discusses with the server and awaits for its approval. Each device, mainly the multi-touch tablets, requests for an action or change. Depending on the state of the application, the server may accept the changes which are broadcast to all devices.

The communication relies on a custom protocol on top of TCP/IP, which works well on a private wireless network. All sockets are configured to send the packets as soon as possible, i. e., no optimization is performed to reduce the bandwidth (i. e., to optimize the ratio between the overall data size, and the useful data size, as multiple packets lead to more metadata). This allows for a smoother experience for real-time applications. As some commands are necessary to be conveyed without loss, e. g., the path of a dataset file, I did not want to rely on UDP protocols. Moreover, some binary data, e. g., the volumetric mask defining volumetric selections, and the data allowing the synchronization of the 3D space (i. e., that objects are perceived by all at the same 3D place) using the Microsoft's HoloLens' API, should be transmitted without errors. For the purpose of my PhD, most of the messages sent and received are logged inside a JSON file for later analysis. Using the Microsoft's HoloLens' API, the first connected AR-HMD sends the data of the room to the server, which broadcasts it to the other AR-HMDs to create a shared coordinate system between all AR-HMDs.

Four threads are used to handle the incoming data of the sockets. Then, for each client, the server creates a new thread to handle the writing part of this client's socket in order to not block the system with unresponsive devices (e. g., an AR-HMD taking time to parse the data synchronizing the 3D space). A last thread is created to send, at a frequency of 20Hz, the AR-HMDs positional data and their current actions. I decided on this value based on a tests I made, and also because I wanted to minimize the network overhead.

The 3D Datasets and their Visual Representations

As VTK files contain multiple variables, not all being necessary for every work session, the user can select which variables he or she is interested in. This optimizes the use of the memory use of the devices, especially the embedded devices.

Once opened, the user can add as many visual representations as necessary. Each representation can either be public, i. e., seen by all, or private, i. e., only seen by its owner. Private visualizations can be made public, but public ones cannot be made private. A user can, however, duplicate a public representation and make it private. A public visualization is locked by a collaborator for 1 second once he or she performs a modification over it, to ensure synchronization. I reset the timer each time a user manipulates the visualization.

Each visual representation possesses a transfer function. To define the transparency of a data point, I am using the Kniss et al.'s [154] multi-dimensional transfer functions based on Gaussian functions, named Gaussian Transfer Function and Triangular Gaussian Function. I map the color based on the data point magnitude, rescaled from (0, $\sqrt{\sum_i^{dimensions} maxVal_i^2}$) to (0, 1). $maxVal_i$ is the maximum value of the i^{th} dimension of the dataset. The new amplitude is then mapped following three color spaces: the rainbow color space, the divergence color space, and the Moreland's [192] MSH color space, which is an adjusted divergence color space. The color space can also be clamped in order to highlight certain part of the dataset. In addition, each visual representation has two depth clipping planes (one from top–bottom, and one from bottom–top) to explore, along the depth axis, the dataset.

I handle an unrestricted number of visual representations because the AR space is as wide as the room the users are in is, allowing multiple non-overlapping visualizations. This allows for side-by-side comparisons in the users' 3D space (**G6**). Users can open as many visualizations as they want as long as the computation power and the memory of the devices can handle them.

2D Log Annotations

Additionally to the 3D datasets, the application handles 2D log annotations (see Figure 3.4), i. e., external data sharing a spatial relationship with the linked 3D datasets.

My collaborators at HZG are interested both in forecast predictions and in actual measurements they made in-situ during expeditions. Their data contains the paths of the vehicles during an expedition and the measurements. In the application, users can open and configure those log annotation data. Then, the users can spatially link a visual representation of a 3D dataset with an annotation which are then spatially merged, solving the issue of relating 2D with 3D datasets (**G7**). The user can finally configure the color mapping. I allow two color mappings: a uniform color mapping, or a mapping reusing the transfer function of the 3D visuals applied to the dimensions the user selected for the visual instance of the log annotation data.

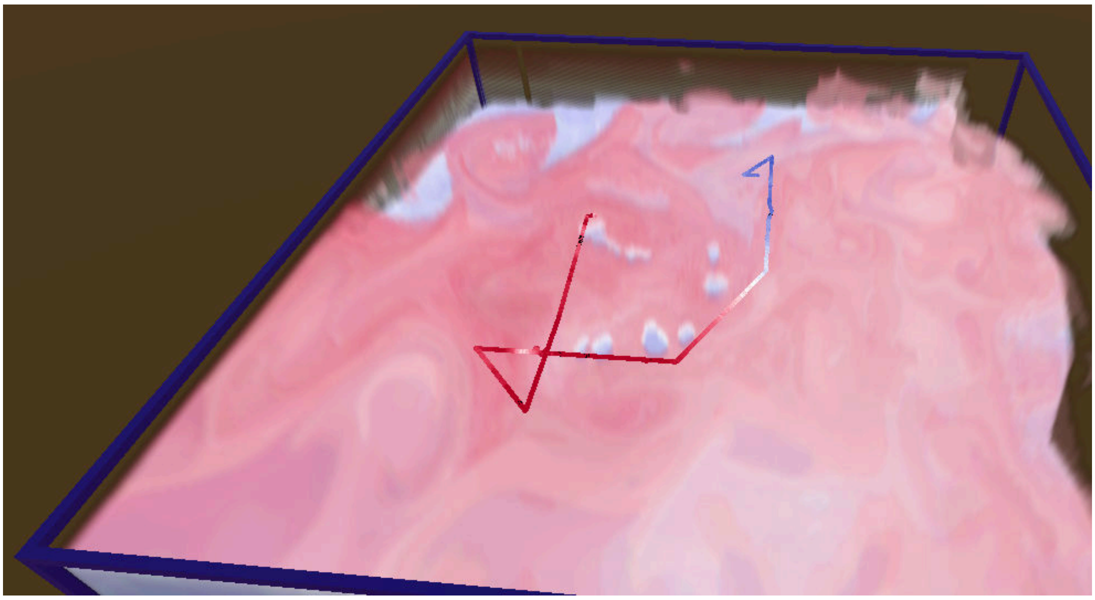


Figure 3.4: Screenshot I took from Unity (for better readability). It shows a 2D log annotation anchored on top of a visual representation of the Cabo Verde dataset. Both visuals are spatially merged. This visual 3D representation shows the temperature scalar field of the dataset. Its transfer function focuses on the mid–high scalar value range. The track represents the path of the boat. The color of the track is mapped to the temperature value as the captured. I emphasize also that the transfer function extracted an eddy visible on the top-left part of the screenshot.

3.4.2 The AR-HMDs

The AR-HMDs are mainly output devices for the visualizations and some additional information. The only 3D input AR-HMDs provide in my software concerns the specification of points; see Chapter 4. I therefore discuss here the visualization aspects of the device.

Their software relies on Unity 3D, C#, and C++. I developed C++ libraries to (1) reuse the parsing of VTK files done for the tablet's and the server's software, and to (2) access low-level information such as the depth cameras of the AR-HMDs; see Section 4.4. While KitWare provides the VTK library to handle VTK files, this library is not yet available on devices relying solely on DirectX, which is the case of the HoloLens, hence the custom VTK parser.

At each frame update (i. e., at a maximum of 60Hz), the AR-HMDs send their positions to the server, which are broadcast and then used by the other devices to locate those AR-HMDs.

Volumetric Rendering

I am focusing in this PhD on volumetric datasets, i. e., 3D scalar fields such as meteorological datasets. To visualize such data, I perform a volumetric rendering using a ray-marching algorithm [209]. I first compute the 3D texture corresponding to the visualization on the CPU based on the transfer function parameters. The computation of the transfer function is done in parallel, as every voxel is independent from each other. Each update considers the current time animation, i. e., at which time the user is currently looking at in the data space. As such, one update needs a linear interpolation between timestep t and timestep $t + 1$.

The transfer functions rely either on the Gaussian Transfer Function or on the Triangular Gaussian Transfer Function as defined by Kniss et al. [154]. For the datasets coming from HZG I am working with, which have a dimension of $370 \times 280 \times 50$, this computation takes 2 seconds on the HoloLens 2^{nd} gen.. To avoid making the devices unresponsive, I update the visualizations on demand on a separate thread. Once the computation is finished, I push the resulting 3D data to the Graphical Process Unit (GPU) for the rendering.

Table 3.1: The list of the current implemented glyphs floating above the users' heads which depend on their actions. In an ideal software, each action should trigger one awareness cue, e. g., manipulating the transfer function should yield an awareness cue.

Action	Input Device	Glyph
Nothing	None	
Rotation	Tablet	
Scaling	Tablet	
Translation	Tablet	

Once the image is computed, for each frame, I determine what parts of the screen the visualization should take considering its Model-View-Projection 4×4 matrix. Then, I render a 2D quad corresponding to this screen area and perform the ray-marching algorithm on the GPU. As the graphics pipeline of embedded systems, today, follows the SIMD architecture using GPU, e. g., via OpenGL and DirectX, the HoloLens computes multiple pixels simultaneously. I then convert the ray cast inside the screen-space to the model-space, and make it traverses the 3D texture until either the ray has a cumulative transparency of 95%, or that the ray crosses an opaque object, which I determine using the depth texture provided by Unity. The ray step is proportionally inverse to the higher dimension of the resolution of the texture.

The scalar field I am exploring with HZG contains $370 \times 280 \times 50 = 5,180,000$ data points. Considering the specifications of the HoloLens 2^{nd} gen., the device requires to read a maximum of:

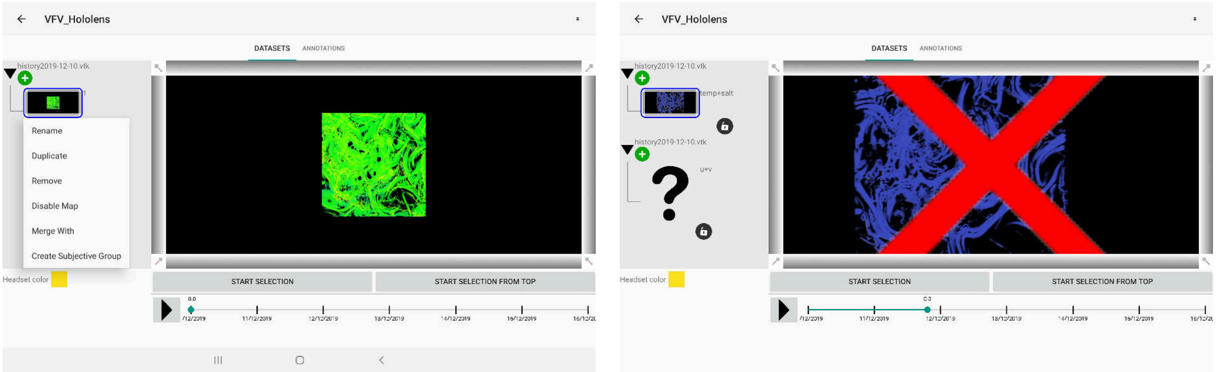
$$\begin{aligned} maxDim \times diagonal \times stereoscopy \times pixelX \times pixelY &= 370 \times \sqrt{3} \times 2 \times 2048 \times 1080 \\ &\approx 2.8 \cdot 10^9 \end{aligned}$$

data values, which I store as 16-bits color values (4 bits per channel) as an optimization, for the whole ray-marching algorithm. This corresponds to approximately $5.7 \cdot 10^9$ to $1.7 \cdot 10^{10}$ bytes being read, which depends on whether there are bilinear, trilinear, or no filtering parameters to apply. The equation considers that we look at the 3D visualization through the diagonal of the 3D window, i. e., the longest segment, and that the visualization fills the whole screen.

The required computation power and memory bandwidth out-scale the current hardware of the HoloLens for a framerate high enough to ensure real time applications ($\geq$ 30Hz), as required by Azuma et al.'s definition [14] of AR systems. The framerate, in this case, depends mostly on the number of visualizations being rendered simultaneously and the size they take in the camera space. As the number of pixels of the headset's screen are far greater than the dimensions of the dataset, I propose to render the visualizations on smaller render textures, which are then drawn on the screen, speeding the whole process and allowing the system to remain reactive.

Awareness Cues

As multiple collaborators can join the session and manipulate the 3D space, it may be hard to understand which user is performing what actions. To solve this issue, I am reusing the concept of 3D windows introduced in Section 2.2.5. Basically, each visualization is bound in space by a 3D colored wireframed cube. As I am relying on the multi-touch tablets for most interactions, and because multi-touch tablets are private screens typically hidden from the other collaborators, I am using the color of the wireframes of the 3D windows to represent which user is currently



(a) (b)

Figure 3.5: Screenshot of the interface of the tablet for what concerns the Datasets Panel. The tree on the left represents the opened datasets and their respective spawn visualizations. A long press on one of them opens a menu for further functionalities. If another user manipulates a public visualization, a red cross shows that this visualization is locked for the moment (b)

manipulating this specific 3D visualization. A default color of blue signifies that no user is currently manipulating this visualization.

To know which user corresponds to which color, I render a 3D glyph above each user. The color of the glyph corresponds to the collaborator's ID and the shape of the glyph corresponds to the user's current action, which the server sends at a frequency of 20Hz. I list all the awareness cues I am currently supporting in Table 3.1. This list could further be extended in the future as I did not implement awareness cues for every action. This list serves, however, as an example and a proof-of-concept for future designers. Moreover, Rekimoto et al. [228] stated that awareness cues, if any, should reach the workspace and not stay on the user. Here, for what concerns the workspace, only the colors of the wireframes change which might not give enough information to users and need further thoughts.

3.4.3 The Multi-Touch Tablets

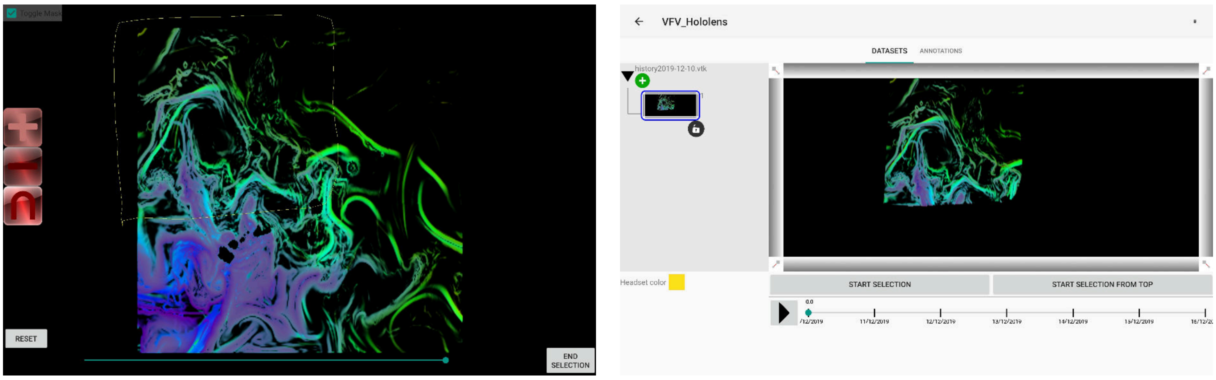
The multi-touch tablets are at the center of the interaction. Their software relies on the Android SDK and NDK, using Java and C/C++. I use *libpng* to open images, *OpenGL* to render the visualizations, and formerly *lib3ds* at earlier version of the software to render 3D arrows where I focused on vector field visualizations before shifting to volumetric rendering. I do not use any other third-party libraries, except those that I wrote myself to not duplicate similar codes between the server and the tablets.

To pair the tablets with their corresponding AR-HMDs, a user enters on their respective tablet the IP address their respective AR-HMD displays. During pairing, each user selects his/her handedness that is necessary for the project concerning the specifications of 3D positions; see Chapter 4. Once the pairing is finished, the message displayed on the AR-HMD disappears and the tablet loads the status of the current work session (e. g., datasets opened, visualization instances created, transfer functions applied, etc.).

The graphical user interface of the tablet comprises three main views which I describe next.

Datasets Panel

The first main view handles all the visualization this user has access to (Figure 3.5). The user can duplicate, add and remove visualizations representing opened datasets, and move those visualizations from the user's private space to the public one.



(a) (b)

Figure 3.6: The degraded volumetric selection interaction (Figure 3.6a) of the tablet interface after having pressed “Start selection from top” (see Figure 3.6b and Figure 3.5). The camera of the view is placed on top of the visualization and faces it. The user can rescale, with the slider at the bottom of the view, the orthographic projection to have a better overview. To select parts of the data (and hide the rest), the user can applies boolean operator (OR, NOT, AND). Then, the user draws a lasso surrounding what he/she desires and presses “End Selection”. Figure 3.6b shows the corresponding result of this example.

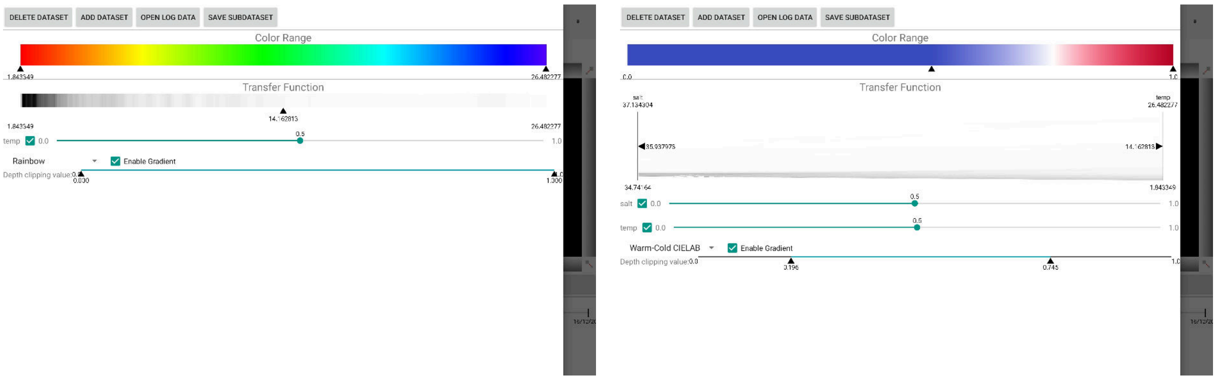
Using the FI3D widgets [298], the users can also scale, rotate, and translate the visualizations in the 3D space. For that, the tablet considers the AR-HMD’s 3D position in the AR space. The translation directions (e. g., forward) are defined with respect to the user’s current orientation. I then handle 2 Degree-of-freedoms (DoFs) for rotations, 3 DoFs for translations, and a uniform scaling, which have been proven to be sufficient for this application. I use an orthographic rendering on the tablets by default as this is a key rendering mode for scientists in their respective software to measure distances, and because the AR-HMDs already provide perspective views.

The bottom part of this panel handles the time component. The user can play the animation or set the current timer at a desired value. Because volumetric rendering takes time to compute, the timer awaits for 3 seconds before moving to the next step. Finally, this view presents a specific functionality to specify region of interests using the multi-touch tablet, with two possibilities. The two of them rely on a 2D shape the user draws on the tablet. In the first case, the user can later move the tablet around to extrude that 2D shape, creating a 3D shape. All the points inside that shape are considered for the selection. Chapter 5 gives further details about this functionality. However, moving the tablet around requires an external tracking system for the moment. We rely on the VICON system which is not as mobile as the HoloLens and the multi-touch tablets are. Then, the second method allows the user to select along the depth axis everything that is inside the drawn 2D shape; see Figure 3.6. This has also for benefits that users can select everything along the horizontal plane, while refining their sections along the depth axis using the Visualization Parameters Panel as I show next.

Visualization Parameters Panel

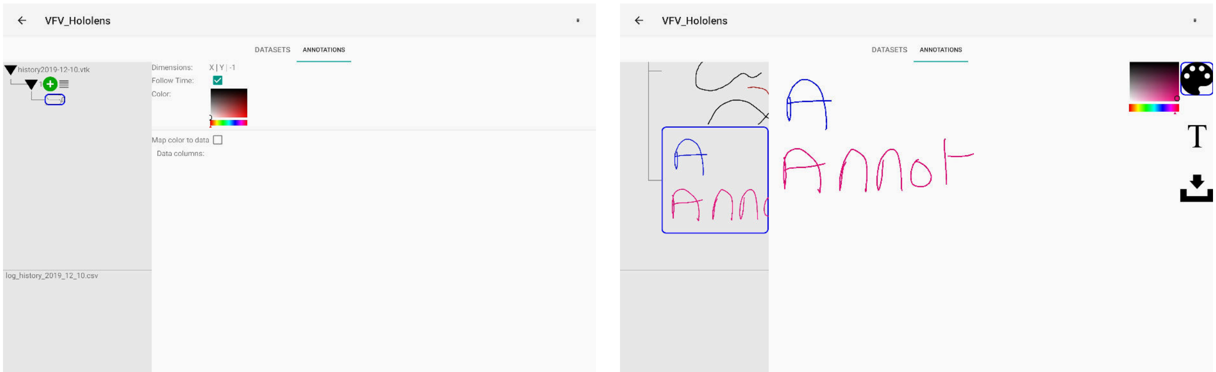
Numerous visualization parameters need to be considered for such an application. I then separate the User Interface (UI) widgets that manipulate those parameters in a dedicated panel (Figure 3.7).

This second main view allows, first, to parameterize the transfer function of the current visualization. Users can select the type of the transfer function by considering or not the gradient of the field as Kniss et al. [154] described. The user can change the Gaussian parameters, change the colormap, and clamp it to have higher magnitude of changes in the data value space (see “Color Ranges” in Figure 3.7).



(a) (b)

Figure 3.7: Screenshots of the interface of the tablet for what concerns the Visualization Parameters Panel. Users can change the transfer function, clamp the data range, and set the depth clipping planes. Figure a shows the interface for one-dimensional transfer functions, and b shows the interface for multi-dimensional transfer functions.



(a) (b)

Figure 3.8: Screenshots of the Annotations Panel interface of the tablet. Figure a shows a "Positional" annotation. A file "log_history_2019_12_10.csv" has been opened. A "Positional" annotation has then been linked to the instance "1" of the dataset "history2019-12-10.vtk." Figure b shows a "Canvas" annotation. The three buttons on the right are respectively to draw strokes, write texts, and import images. In the current version, users can only draw strokes.

The user can change the size and the center of the Gaussian function along all the dimensions of the opened dataset. To give insights about the data, I provide a histogram to represent the distribution of the data in the data space. When multiple variables are considered, I display this distribution using a continuous parallel coordinate plot [123]; see Figure 3.7b. For the moment, the parallel coordinate plot is not interactive (i. e., it is not possible to move the panels). However, this is a strong requirement for a more complete software as it allows to efficiently explore the data space.

Finally, the user can set two depth clipping planes: one from the top and one from the bottom. To select along the other two local axis of the dataset, users can spatially select what they are interested in by drawing a 2D lasso as explained in Section 3.4.3.

Annotations Panel

The last panel facilitates the handling of annotations (Figure 3.8). I aim to handle, in this software, two kinds of annotations.

The first one, annotation canvas, allows the user to draw sketches, enter texts, and add images. As the discussion goes on, collaborators need to share and save thoughts for later uses. Sketches are essential in collaborative tasks which the survey (see Section 3.2) confirmed. The transmission of the sketch data to the server and the managements of these canvas on the AR-HMDs, however, unfinished. While I did not finish to implement this functionality, I would like to emphasize that this functionality would not be comfortable if relying only on the AR-HMD input devices. Let us first consider sketching. Arora et al. [12] showed that 2D sketches have better accuracy and fairness when using a physical 2D surface compared to drawing mid-air, even if the system can project mid-air strokes into a 2D plane. Moreover, drawing in mid-air tires the user's arms quickly; see the Gorilla-arm syndrome [128]. Finally, entering texts on the AR-HMD is slow and tiresome. All those functionalities can, however, easily be implemented on a multi-touch tablet and other hand-held displays; see the numerous consumer applications for smartphones and tablets allowing to type texts and perform 2D sketching.

Moreover, while I describe the anchoring process of 2D and 3D objects in Chapter 4, I did not implement a way to convert the number of pixels saved by the canvas to the metrics system used by the HMD, i. e., how many pixels fit a meter. This might depend on what users intend to do with these annotations. I then would suggest future designers to have such a behavior parameterizable on the multi-touch tablets. An alternative way would be to add simple 3D mid-air gestures on the AR-HMDs to reorientate, reposition, and rescale those canvases. As we expect users to hold their tablets on their hands, however, a study might be necessary to find which method (tablet against AR-HMD) works best. Finally, multiple research questions remain for those annotations. For example, once anchored, how should a canvas be orientated per user? As 2D canvases contain information only visible from its front, users might either need to move in front of those canvases to have access to their information, or the system should render canvases in such a way that they always face the user. However, such layout would lower the workspace coherency and might introduce communication difficulties concerning complex annotations. This research question is also stated in our survey (Section 2.6.2).

The second type of annotations concerns the reading of 2D data as discussed in Section 3.1. I am currently handling 2D timestamped positional data, as the scientists at HZG want to understand the movements of the plane and the ship across time during their expeditions. First the user opens the file containing the data. Currently, I only handle CSV file format. Second, the user selects what is the column corresponding to the time data entries. Once this CSV file is opened, the user creates as many LogAnnotationPosition objects as necessary. Each object represents a view on the CSV file format, allowing the user to define what are the X, Y, and Z column entries corresponding to the positional data. Then, the user can link as many visual representations as they want to these LogAnnotationPosition, creating an instance of a LogAnnotationPosition. The user can set the default color to this instance, or map the column entries to the visual channel. The color computation reuses the transfer function of the linked 3D visualization (which can be multi-dimensional).

In a more complete software, the user interfaces to visualize such 2D abstract data embedded in, e. g., CSV files, can follow current recommendations of InfoVis researchers. For example, they can take into consideration the grammar of visualizations [289]. The multi-touch tablets, with their general-purpose graphical user interface, are more appropriate than the AR-HMDs to implement and use such a complex user interface.

3.5 Is the Current Hardware Limiting?

The specific hardware I use in my implementation and their inherent limitations have impacts on the design of my application, which I discuss next.

Some issues are likely to be answered by future iterations of the hardware. For instance, the HoloLens can have a better tracking accuracy and be used in brighter environments. The color fidelity can be improved, and future headsets can surely track additional devices (e. g., tablets), similarly to what the Oculus Quest by Facebook is doing with its remote controllers, for instance.

However, other issues may unlikely be answered. While computing power tends to increase, users' demand and data size increase as well. Former issues about limited computing power answered by new hardware will rapidly be relevant again after a few time. AR-HMDs, by being embedded systems, will likely always be less powerful and have lower rendering resolutions compared to workstations and dedicated machines. Rendering dark colors is also an issue for Optical See-Through displays. One solution would be to rely on Video See-Through devices, but the resolutions and latencies of their screens and cameras should increase dramatically for users to not feel disturbed by the headsets compared to Optical See-Through devices which have, for what concerns the real world, an "unlimited" resolution and framerate.

Finally, both the AR-HMDs and multi-touch tablets are not suitable for scripting applications, which limits all the heavy data analyses and computations steps (e. g., the scripting part of ParaView would hardly be possible in my hybrid environment). While this is compatible to my design goals (**G8**), designers should consider how powerful and complete their applications need to be.

3.6 Evaluation

With the COVID-19 pandemic, I was not able to evaluate this system with end users in a well-designed user study. However, as in August 2021, I was able to test this system with two oceanographers. One of them is Burkard Baschek (user U_a), my collaborator. He knew already the system and how to use it, while the second user (user U_b) had not used it in the past. U_a is the director of the Institute of Coastal Ocean Dynamics, and U_b heavily works on the forecast predictions of the Cabo Verde area at different resolution scales. This session lasted approximately thirty minutes.

While the study was informal, the way those two persons discussed and collaborated yield hypotheses that I want to highlight in this PhD which might be the starting point of future work.

3.6.1 Events

The dataset those users worked on corresponds to the Cabo Verde area, which contains multiple eddies. The dataset contains water temperature, salinity, longitude and latitude velocity, and magnitude velocity. Users were sit next to each other. Instead of moving themselves around the datasets, U_a moved the visualizations around him using the FI3D widgets.

U_a opened the dataset and loaded the temperature. He put this instance near him. Meanwhile, I was talking and explaining the system to U_b. Then, U_b opened an instance which was at position $(0, 0, 0)$. This location was near U_a and entered into conflict with U_a as, at this point, he was working alone around this physical area. Then, he specifically asked to move that visualization near U_b to have his space free.

After some minutes, U_a wanted to discuss with U_b, who was not trained with the system yet. Especially, U_b did not yet understand the FI3D widgets. From this small session, it seems that FI3D widgets require some training and thus cannot be launched to newcomers, entering into conflicts with **G10**.

U_a visually linked the temperature and the velocity properties together, but he was not able to match correctly the insights both visualizations gathered. U_b, based on his knowledge of the graphical area, suggested to open the salinity property, as this area is particularly prone on salinity changes. U_a opened the salinity and used the Triangular Gaussian Transfer Function to visually remove regions that have small gradient values (i. e., areas where the scalar field is

Figure 3.9: A Screenshot taken from the HoloLens showing all the opened properties of the datasets (velocity, salinity, temperature) and their respective instances for the informal evaluation. All the visualizations rely on a Rainbow color scale and use Triangular Gaussian Transfer Functions which weight opacity with (1) the area of interest in the dataset domain (using a Gaussian: Center c and Size σ) and (2) the gradient of the field. From left to right: Velocity magnitude, Temperature (in front), salinity (in back), and salinity again. The labels are orientated similarly in the 3D space for all the users.

almost constant). This allows to highlight eddies where gradient should be strong. Figure 3.9 shows a screenshot of all the opened visualizations.

Surprisingly, U_b did not need to be trained to follow the discussion. U_b, to follow the discussion, laid down the tablet and focused on the AR view. Both of them were able, using hand gestures, to communicate seamlessly.

After U_a gave few explanations about how he parameterized the visualizations, he started a time animation process. At one point in the discussion, U_b stated the following key sentence: "this eddy was the one I had trouble with two weeks ago." Compared to what they usually do, here they looked from the bottom of the visualization to check for its 3D structures (see Figure 3.10). I asked whether they are used to rely on 3D visualizations, and they answered that they currently check, on their software, the depth values slice per slice (i. e., using 2D visualizations).

3.6.2 Discussion

While this session was short, it gives few indications to later investigate in future work. First, the system seems complete for small data exploration (i. e., it contains most of the key functionalities). Especially, as users already knew the datasets, which corresponds to my initial focus scenario (see Section 3.1), users were able to retrieve some properties they already found in their 2D software. However, compared to the 2D software, here they were able to understand and visualize the 3D structure of their datasets which they never have looked at before. Instead of passing more of their time working from the top, most of the discussion considered the bottom part of the "salinity" visualization. The immersion of the devices surely influences that behavior, and solves the issue of finding suitable 3D data representations (**G5**).

Second, and most importantly, users were able to communicate seamlessly. Even without training, U_b was able to participate in the discussion and gave some thoughts about what he was seeing. In this system, users initially passive can use AR-HMDs and understand what they see, ask questions, and then be part of the discussion which make them active users. The decoupling between the interactive devices (the multi-touch tablets) and the visualization devices

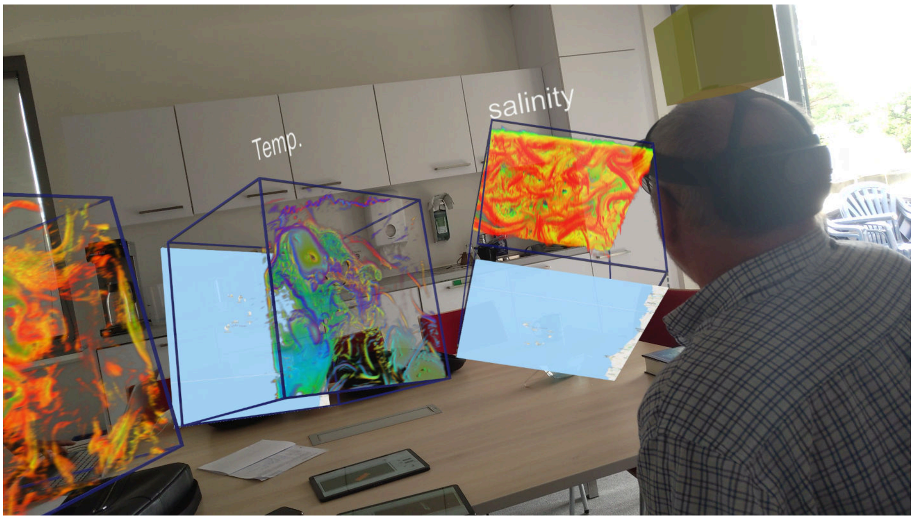


Figure 3.10: A Screenshot taken from the HoloLens of user U_a showing the salinity property (right visual instance) of the dataset viewed from its bottom, where the two users were discussing and visualizing the 3D structures of the eddies.

(the AR-HMDs) might straighten that point, as here users only need to visualize and not interact with the system. The whole view contains then only things users can visually grasp, and where all the mechanics are hidden. Moreover, AR-HMDs have been proven to immerse the users more than traditional tools such as slideshows (see the survey Chapter 2), and might suit non-domain experts such as people at a communication department (see Section 3.2).

Third, private and public spaces are important, and where new visualization instances spawn should be considered. Research should investigate whether new instances should spawn in the "public area" for every user or whether they should spawn near, e. g., the active user. If they should spawn at the "public area", correct algorithms might be required to avoid visualizations to overlap with areas other users are working in.

Fourth, one initial assumption is that users would prefer to walk around the visualizations. However, by providing a remote way of placing the 3D visualizations using the multi-touch tablets, the two users preferred to stay sit. U_a preferred to move the visualization instances around him and U_b, who understood the rotations part of the FI3D widgets, rotated the visualizations instead of moving himself around them. The assumption of moving around should then be reconsidered when multiple visualization instances are spawn and where the software targets multiple collaborators. Indeed, the working space is more cluttered with multiple visualization and users may be influenced by social norms, which may influence the users' will to move in their 3D space.

Finally, by being sit side-by-side, those users could not see the glyphs above the head of their collaborator. They noticed it when I pointed it out to them and were surprised that those glyphs existed. As Rekimoto et al. [228] stated, workspace awareness should reach the workspace and not stay on users. Here, only the colors of the wireframed cubes change, wireframe that is inside the workspace. More thoughts should be put for what concerns the awareness cues.

Again, I based those points on a not planned user study with only two users, even if they are domain experts. I can then only ask questions and give primary results to further investigate.

CHAPTER 4

THE SPECIFICATION OF POINTS

This chapter is based on an article that is accepted for publication at the Springer's journal *Virtual Reality* [263]. Any use of "we" refers to me and my co-authors which are Lonni Besançon and Tobias Isenberg. This is the first project focusing on the "exploration" part of this PhD.

In this research project, we compared three techniques to specify 3D positions for Augmented Reality (AR) collaborative visualization. As interactions are key components of exploratory visualization tasks, but because the AR literature lacks from fundamentals for what concerns data exploratory analysis, we adapted from the Virtual Reality literature three distinct techniques to specify points in 3D space. While selections rely on existing objects, we use the term "specify" for empty locations. 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 however differ in the way users behave, their accuracy, and speed.

This chapter then targets **RQ2** that I stated as:

> **RQ2** To specify points in a volumetric dataset, what are the implications of the different available metaphors on the users' understanding, co-presence, and performance, in a collaborative AR environment?

4.1 Introduction

Well-designed interaction is essential for exploratory 3D visualization tasks [144, 147, 229, 271] within immersive analytics [59, 181], including the selection of Region-Of-Interests (ROIs) [38, 270, 296, 297]; see also Chapter 5, specific features (e. g., vector orientation) [141, 241], and 3D positions [36, 106, 129, 137, 153, 298]. To select objects, systems generally propose pointing actions, e. g., via raycasting [11, 103, 114, 174], or via direct manipulation [249]. In 3D datasets

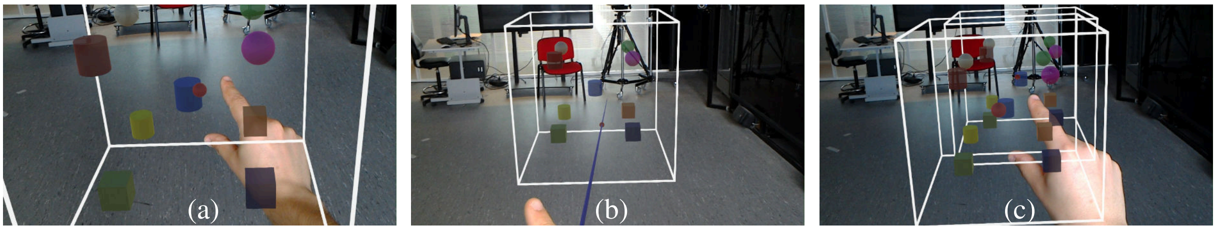


Figure 4.1: The three techniques we studied: (a) *Manual*, (b) *AR-Go-Go*, (c) *World-In-Miniature (WIM)*. We display in all cases a red spherical 3D cursor (radius = 2 cm) to every user. Our study uses an abstract dataset which supports communication with its colored geometries. The outline of the dataset turns white when the 3D cursor is inside the dataset's bounding box.

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 [38, 132, 297] but focuses mostly on selecting ROIs and does not use AR Head-Mounted Displays (AR-HMD), to the best of our knowledge.

Compared to ROIs, pointing at a 3D position allows users to show this point to others, to probe its data (e. g., velocity and temperature), and to place objects at this location. These objects can be annotations, which are important in visualization systems [78, 119, 246] and are core features of collaborative systems [134, 152, 255]. To study the efficiency of 3D point specification techniques[1], we mostly focus on whether users are (1) efficient with them in solo tasks, and (2) if their partners understand what a given user is doing, as both are important in collaborative environments [42]. For example, students in medicine should be able to specify the location of a tumor and their teachers should be able to guide them to this specific location.

Our contributions are threefold. First, based on Poupyrev et al.'s taxonomy [223] of selection techniques in virtual environments, we adapted and implemented three Virtual Reality (VR) 3D selection techniques to AR-HMD environments (Figure 4.1). Indeed, the VR literature is richer than the AR one for exploratory data analysis, hence this choice. The first technique allows users to specify points by physically reaching the point of interest. The second one relies on a ray crossing the user's chest and dominant hand. The distance between the chest and the hand controls the position along the ray. The third one creates a proxy of the dataset, which acts for the original. Second, we studied these techniques in a collaborative setting for guidance tasks based on 3D point specification, to measure the subjective co-presence and respective understanding of the participants. Third, in another controlled experiment, we evaluated the techniques' performance in non-collaborative AR work to give readers insights about how well do these VR techniques work in AR. We report on participants' behavior, performance, fatigue, and the efficiency of the awareness cues they perceived.

4.2 Related Work

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

4.2.1 Collaborative Systems

While co-located AR setups tend to provide better awareness cues among collaborators compared to VR setups, as I stated in Chapter 1, yet there may not be enough of it. Spaces relying on both virtual and real objects must be coherent regarding these two types of objects. Since more literature addresses interactive data visualization in VR than in AR, we reused existing VR interaction techniques to AR spaces, and adapted them if they do not directly apply to the changed environment. For example, we adjusted the *Go-Go* technique [222] by changing the arm extension (which would not be aligned with reality) to a displayed 3D cursor.

The survey presented in Chapter 2 shows that most studies focusing on awareness cues and collaborations focused on remote collaboration, as both the embodiment and conversational cues are easily available in synchronous co-located setups. In this work, we focus on giving insights about awareness cues that are available to specify a 3D position in co-located settings. The only work we found focusing on co-located collaborative 3D selections using AR-HMDs as supports is Oda and Feiner's *GARDEN* system [199] that my co-authors and I reviewed in our survey. To summarize, the system facilitates 3D referencing in shared AR based on the scene's depth data. They studied both the roles of *indicator* (the one pointing at objects) and *recipient* (the one

[1] We use the term *specification* and not *selection* as we are interested in arbitrary points that have no intrinsic meanings inside the dataset, which usually *selection* refers to.

understanding the intended location). The recipient was more accurate with their technique than with laser-pointers, virtual arrows, or video sharing techniques.

4.2.2 3D Visualization

I justified in Chapter 3 the general interface this work relies on based on previous work. I remind the reader that my interface relies on multiple AR-HMDs combined with multi-touch tablets. The primary idea is to delegate most of the interaction, mostly for what concerns WIMP interfaces, on the multi-touch tablets. In this research project, however, we are interested in 3D point specification which is 3-dimensional by nature. We thus investigate the use of immersive 3D devices for such a task by first looking at the VR literature.

VR-Based Selection Techniques

Because of the 3D and direct interaction nature of VR [57], researchers intensively studied point and region selections in this environment [11, 38]. Most of these techniques rely on geometry or ray casting metaphors to allow users to reach far targets [11, 38]. 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 [38, 296].

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 [21, 114, 230, 235]. 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. 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* [113]) or by adapting the radius of the cursor [235] depending on its location, as there is no well-defined target in our case.

Multi-Touch Device-Based Selection Techniques

The tablets I am relying on can also allow users to select (and specify) 3D points via touch input, e. g., the *Balloon* technique [27, 79].

For collaboration, however, non-tangible tablets may not allow other co-workers to be aware of actions because they cannot see this personal screen. Additionally, the multi-touch tablet could be used as a tangible device which, by its 3D manipulations, give awareness cues to all collaborators. However, interactions based on motions may be physically too demanding compared to using hand gestures because users are forced to hold the device, which has a certain weight, 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 (e. g., *SpaceCast* does not apply for datasets without a scalar importance measure).

4.3 Motivation

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

- Egocentric ⟶ **World in Miniature (WIM)**
- Exocentric
  - Virtual Hand Metaphor
    - Remote ⟶ **AR Go-Go**
    - Non-remote ⟶ **Manual**
  - Virtual Ray Metaphor ⟶ ~~Inapplicable~~

Figure 4.2: Summary of our techniques' choice to cover an extended version of Poupyrev et al.'s taxonomy [223].

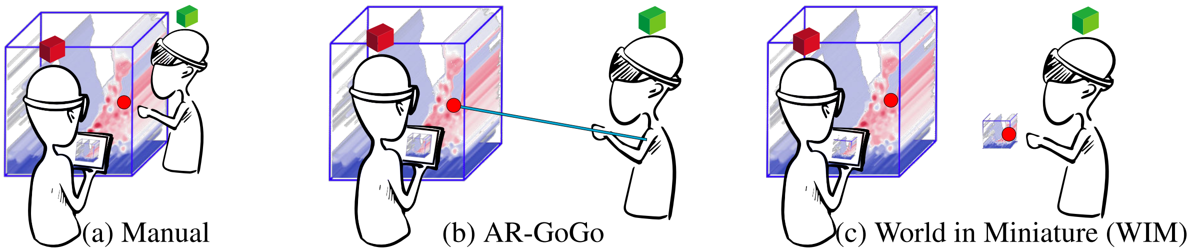


Figure 4.3: Sketches showing the three representative techniques we selected. We give screenshots of their implementations in Figure 4.1.

4.3.1 Three Interaction Techniques for Point Specification

We are studying 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 one 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 (Figure 4.2). We aim, with this approach, to understand the effects of most techniques in collaborative contexts by studying three techniques representative of the whole taxonomy. As 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 [222] and exocentric techniques [212, 249]. Poupyrev et al.'s taxonomy [223], however, does not consider the techniques' reach. Users can indeed virtually reach any object in VR with their hands [222], 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 think 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, so we also included remote interaction techniques in our experiments. We summarize our choices in Figure 4.2 and show our techniques schematically in Figure 4.3.

4.3.2 Scenarios

We then wanted to determine the techniques that work best in collaborative scenarios. We identified particularly the following three scenarios: (1) a solo task where the user is alone in the environment, (2) a collaborative task where collaborators work on the same object of interest, and (3) a collaborative task where collaborators are working in parallel with two distinct regions of the environment. As the previous chapters discussed (see Chapter 1 and Chapter 3), our system comprises multi-touch tablets to allow users to do more than only specify 3D positions, while allowing users to move around their data. This research project targets then a real-case scenario instead of focusing solely on one particular task. Studying these three scenarios in one study would, however, be cumbersome due to the limited time participants can focus their attention (≈ 1.5 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 reuse the collaborative system described in Chapter 3.

4.4 Implementation

We describe in this section only the specific aspects of this project for the system we used, and we redirect the reader to Chapter 3 for more details on the overall system. As we performed this project before the HoloLens 2^{nd} gen. were available, we relied on multiple Microsoft's HoloLens 1^{st} gen. which we found provided the best overall AR experience among available hardware at that time (as of end of 2019). The software and its environment are similar to what I described in Chapter 3. The tablets are again the Samsung's Galaxy Tab S4 with Android 9. The git repository of the server specific to this project, which links to the other parts of the system, can be found at `https://github.com/MickaelSERENO/SciVis_Server/tree/CHI2020`.

Next, we describe the three techniques we chose and implemented, as they are shown in Figure 4.1 and 4.3. 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 1^{st} gen.'s 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 Microsoft's Media Foundation API in "research mode." To validate a position, the user performs a "Tap" gesture, which is natively recognized by the HoloLens' API. This gesture is compatible with our hand detection but generates a small shift (called the Heisenberg effect [51]) 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 HoloLens' API than in our custom algorithm without removing it. To reduce the effect as much as possible, we applied an exponential smoothing ($\alpha = 0.3$) as a low-pass filter to the position input. Participants in our study did not report any issue (such as latency) regarding the filtering and its parameters.

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. [283] and Issartel et al. [140]. Current standalone HMDs, including VR (e. g., Oculus Quest) and AR (e. g., Microsoft's HoloLens 2^{nd} gen.) HMDs, show a tendency of using visual tracking based on RGB and depth 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.

4.4.1 Manual technique

In the *Manual* technique (Figure 4.1(a), 4.3(a)), a user must physically reach a position of interest—a local virtual hand metaphor. We render to all co-workers a small red sphere near the user's thumb that acts as a 3D cursor. We use a heuristics thumb position because we think it is more comfortable than the captured 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. Four of these persons strongly preferred and three preferred using the estimated thumb position compared to the captured 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. Bruder et al.'s [58] showed evidence that users seem more accurate when a small offset is applied, at the cost of speed. We prefer, however, to emphasize accuracy over speed due the tracking system relying on image processing which is subject to noise and because 3D data exploration usually requires high accuracy.

We compute the heuristics 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 it to determine the forward and left (or right) axes. This noise is comparable to the noise of Microsoft's head-gaze interaction technique (as they are both based on the head orientation) and it has a low impact because 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 the HMD is highly inconvenient. Other solutions such as a separate Raspberry Pi would decrease the portability and thus would create new limitations [140]. 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.

4.4.2 AR Go-Go interaction technique

The *AR Go-Go* technique (Figure 4.1(b), 4.3(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 $d = 13.0 \times |(hand - (head + (0, -0.15, 0))| - 0.20)$ *(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 [222], 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 [54]), because an 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. [174] 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 [103, 223]. 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.

4.4.3 World In Miniature (WIM)

The *WIM* technique (Figure 4.1(c), 4.3(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.

4.5 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, we created in a first experiment a pair guidance task done using a simple scene to measure the collaborative aspects of our techniques. These tasks rely both on the "actor" performing the task and the "guider" who has more information than the "actor." In our study, a guider is someone who knows where the target position is and guides the actor to that point. 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 points, 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 [166]. 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, benefit from the awareness cues. Moreover, our motivation is interactive 3D visualization such as voxel datasets rendered volumetrically. One of the core tasks 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 medical datasets, hurricane features in meteorology datasets). Then, scenarios where there are issues regarding occlusions should be rare; see also the screenshots presented in Chapter 3 (Figure 3.9 and Figure 3.10). Our sample dataset thus seems suitable to represent such cases, without participants actually needing to understand these sciences. Finally, and to the best of our knowledge, no studies compared these three techniques together in AR and VR contexts. We thus also discuss performance results for all techniques based on a second experiment.

4.5.1 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,

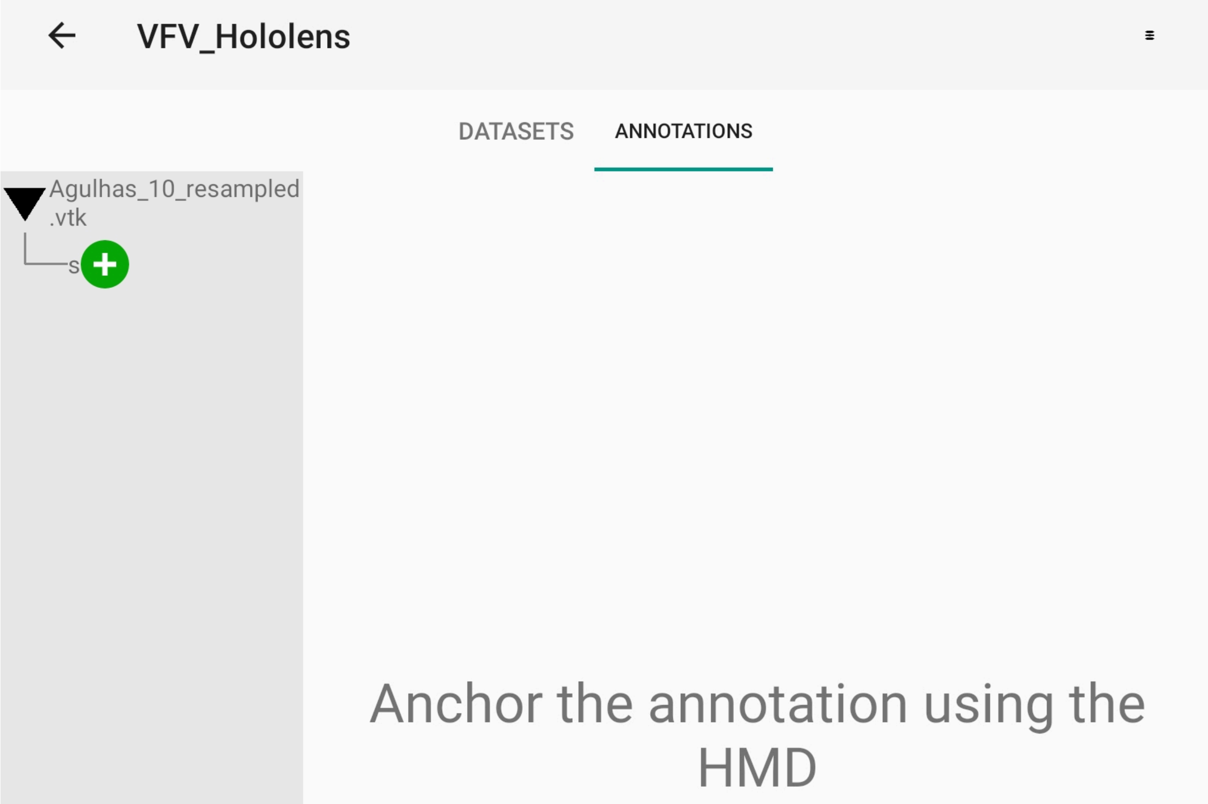


Figure 4.4: View asking the user to anchor the annotation using the AR device after pressing the green “+” button. Users can select which interaction technique to use during the training session using additional menus (top-right menu icon).

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,

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.”

4.5.2 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 $50 \times 50 \times 50cm^3$ dataset that comprised simple colored geometries (cubes, spheres, and cylinders). While the tablet was not mandatory for such experiment, this thesis provides cues (Chapter 3) about its usability for 3D data exploration using AR-HMDs. As this work concerns 3D interaction to explore 3D datasets, we enforced its use in this experiment.

To start the annotation process, participants tapped on the “+” button on the tablet shown near the dataset label (Figure 4.4), where we encouraged them to “Anchor the annotation using the HMD.” We thus clearly separated the annotation placement from normal tracking. Once the annotation was anchored using the chosen interaction technique, we rendered a small yellow 3D star at that location in the AR space for all collaborators and stopped the hand tracking. As a constraint in every condition, we asked participants to be at minimum 1.5 m away from the center of the dataset (they can then move if needed), before beginning the anchoring process,

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

Table 4.1: Number of $(accuracy; speed)$ data points discarded per technique per study. Each condition comprises 192 trials.

forcing them to move in the *Manual* condition. We represented this radius by placing 4 markers on the floor, one for each direction. We did this because we believe that, during a normal session, 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. We applied no further positional constraints.

Participants went through all interaction techniques in each experiment. We counter-balanced the technique order with a 3×3 cyclic Latin square. Each participant used, for every trial, a position from his/her assigned pre-generated pool to reduce learning effects. We reused these two pools for every pair of participants in this study. Each pool contained eight 3D positions, determined by a uniform random generator. Values ranged from –0.5 to +0.5 along the x-, y- and z-axes, defined in the dataset's local coordinate system. We ensured that these positions do not overlap with the dataset geometries. 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.

4.5.3 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 experimental data analysis [9, 10, 86, 111, 187, 274] and follow current APA recommendations [276] by reporting our results using estimation techniques with effect sizes and confidence intervals (CIs), instead of p-values. A p-value reading of our results can still be inferred [156] and our effect sizes refer to the means we measured and do not use standardized effect sizes [71], whose reporting is not always recommended [18]. We interpret our effect sizes and CIs as providing different degrees of evidence about the population mean [34, 35, 75, 85].

We pre-registered our experiments at `osf.io/43j9g`, following current best practices [69, 70]. The current scripts and our gathered data are available at `osf.io/7a6yr`. We modified the pre-registered scripts to correct for logical errors, as recorded on OSF. We also removed miss-clicks performed by users, i. e., detection of false-positive gesture which were reported by users 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 [98]. 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 removed outliers for the speed and accuracy metrics. Outliers are defined as values being distant of the mean by more than three times the standard deviation. We summarize the numbers of discarded values in Table 4.1. Except for speed and ranking analysis, we represent all statistical results using a 95% Bootstrap CI (BCA method). In addition to this analysis, we also video-recorded every session for exploratory analysis.

At the end of the study, we asked participants to rate their familiarity with immersive HMDs on a 5-point Likert scale (separately for AR and VR, ranging from "never used" (1) to "use daily" (5)) and the subjective accuracy of the hand tracking on a 7-point Likert scale. We then asked participants if they had comments and to rank each technique (1 is best) regarding *co-presence*, *message understanding*, *accuracy*, *speed*, *parallel* tasks, and *overall* ranking.

4.5.4 Participants

A total of 12 participant pairs (6 females, 18 males, mean age=28.4, median=25.0, SD=7.37, range=23–56) completed our controlled experiments. 23 participants were right-handed and one was left-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 Pi, $i \in [0, 11]$; for each of these we denote participants as either ID 0 or ID 1; so overall we refer to $Pi_j, j \in [0, 1]$.

4.5.5 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.

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^{st} 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 $50 \times 50 \times 50cm^3$ dataset.

Training. We first launched 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. To select the technique to train with, we provided a menu on the touch device. At the end of the training session, we launched the first part of the study.

Tasks. For 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. We then asked this participant to guide his/her partner to the marked location, to be as accurate as possible. The star served as a proxy position for structures that users would mentally target in a real dataset, e. g., the eye of a hurricane. Because in a real dataset such explicit object may not exist and certainly is not known *a priori*, we did not use visual highlighting when the 3D cursor was near the location. 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, and we measured each technique's performance in our second experiment.

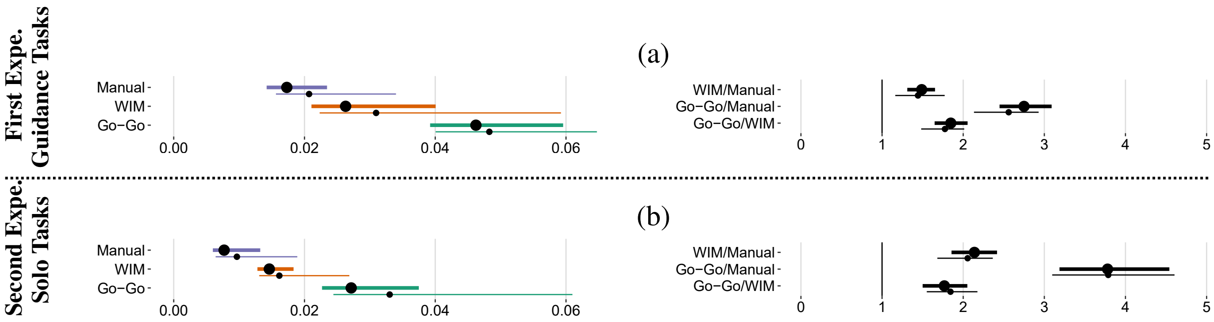


Figure 4.5: Accuracy results for the two experiments of the study. For each, the left graph represents using a 95% Bootstrap CI the distance error of each technique in meters. The right graph represents the pair-wise comparisons (ratio) 95% CI for the three techniques. For all metrics, the top CI is computed with outliers removed and the bottom CI represents all data.

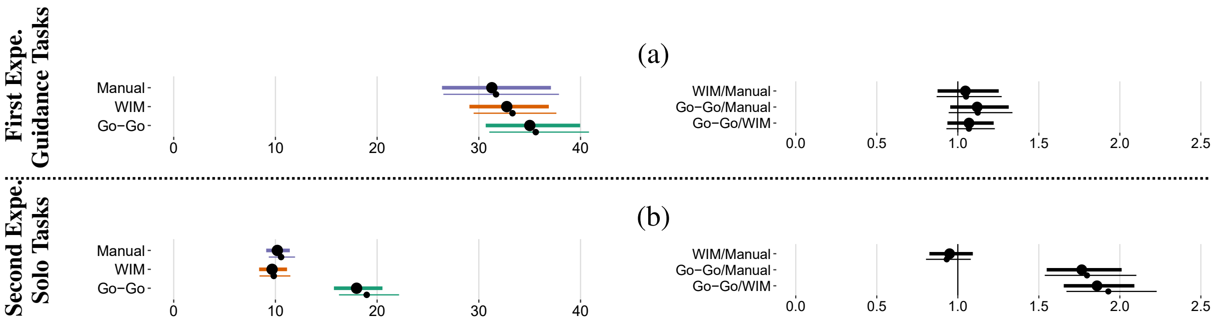


Figure 4.6: Time completion task (TCT) for the two experiments of the study. For each, the left graph represents using a 95% CI the TCT of each technique in seconds. The right graph represents the pair-wise comparisons (ratio) 95% CI for the three techniques. For all metrics, the top CI is computed with outliers removed and the bottom CI represents all data.

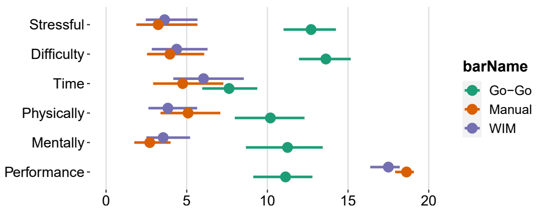


Figure 4.7: The results of the NasaTLX questionnaire represented using a 95% Bootstrap Confidence Interval (BCA method).

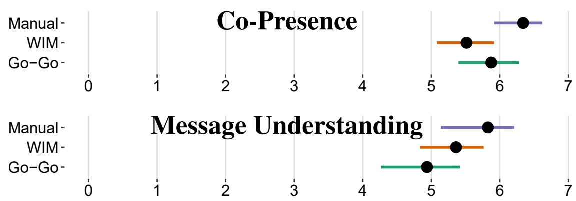


Figure 4.8: 95% Bootstrap Confidence Interval (BCA method) of the Co-Presence and Message Understanding metrics. We averaged per metric and per user the six related 7-points Likert scale questions.

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 [120]'s *Networked Minds Measure of Social Presence* questionnaire. We were inspired by related studies, e. g., by Piumsomboon et al. [218] and Lee et al. [166], who also used a subset of the six metrics proposed by the questionnaire. We averaged the results of all questions per factor item to get one metric per factor, considering positive and negative questions.

Pre-registered Results

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

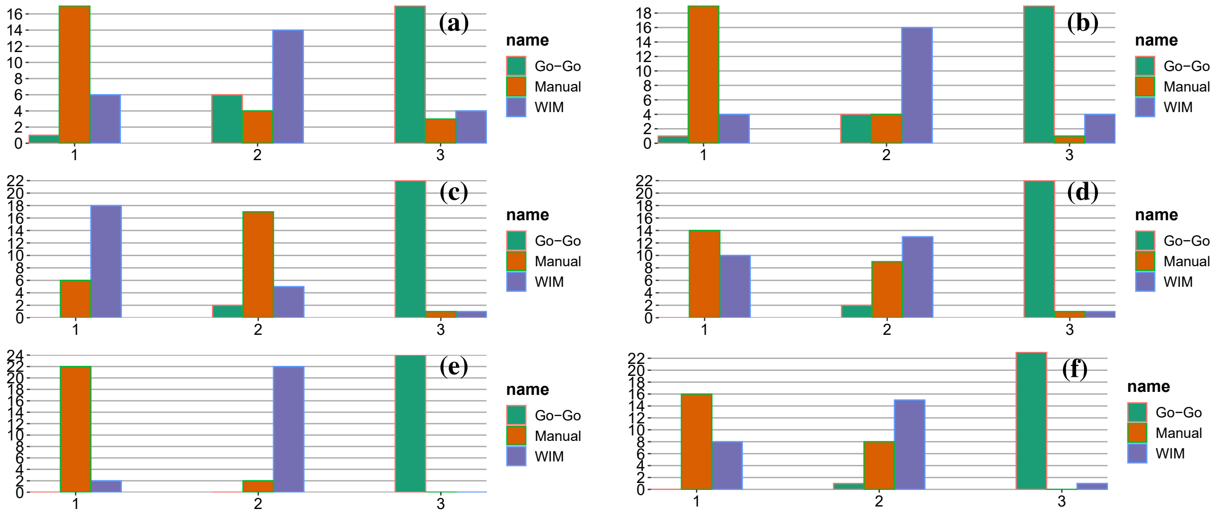


Figure 4.9: Ranking counting results, 1 being the best result: *(a)* co-presence, *(b)* message understanding, *(c)* parallelism, *(d)* speed, *(e)* accuracy, and *(f)* general.

Accuracy. Figure 4.5a 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 seem to reveal a difference between these techniques. When inspecting the pair-wise ratios, however, the CIs reveal strong evidence of a difference in accuracy between the techniques. We see that *Manual* is about $1.5\times$ as accurate as *WIM* and about $2.5\times$ as accurate as *Go-Go*. *WIM* is a bit less than $2\times$ 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 [238] which is commonly used in the HCI literature (e. g., [33, 34, 143, 163, 297]). We thus arrived at geometric means (Figure 4.6).[2] We started the timer when the participant pressed the "+" button and ended it when the annotation was anchored. We enabled the experimenter to restart the timer to resolve issues just-in-time, without altering the metric. In such cases, the experimenter asked the participant to re-press the "+" button. Note that a participant needed both hands to restart the timer; one hand to hold the tablet, and one hand to press the button. To whole phase of refinement, the longest phase, thus needed to be redone from the beginning and the validity of our speed results is preserved. Using pair-wise ratios, we found no evidence that one technique would be faster than another one, which contradicts **H1** and **H3**.

Co-Presence and Message Understanding. Figure 4.8 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

[2] While an arithmetic mean uses the sum of a set of values to obtain the mean, a geometric mean uses the product of the set's values. They dampen the effect of potential extreme trial completion times which could otherwise have biased an arithmetic mean. We plotted results and pair-wise ratios in Figure 4.6a.

questionnaire shows different results regarding these metrics, with *Manual* providing the most co-presence and best supporting mutual understanding.

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 (Figure 4.1) 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* [239] because concepts such as ceiling and wall exist for virtually every visualization. We also observed $P0_{0,1}$, $P1_{0,1}$, $P2_{0,1}$, $P3_{0,1}$, $P4_{0,1}$, $P5_{0,1}$, $P6_{0,1}$, $P7_{0,1}$, $P8_{0,1}$, and $P9_{0,1}$ often to 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 *Manual* is the most direct method, i. e., there is no artifact between the object of interest and the interaction. Indeed, based on Bruckner et al.'s model (see Figure 1.4), *Manual* merges the manipulation, interaction, visualization, and output spaces together. 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 interferences, $P6_1$ made a "v" sign using two fingers to show the target location. To be as accurate as possible, $P0_1$ 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 gave only voice instructions. We also saw some participants ($P4_1$, $P8_1$, $P9_0$) staying behind their partner. $P4_1$ 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 *AR Go-Go* allows users to communicate about the same object without standing near each other. This setup not only facilitates hand tracking, it also provides better deictic and gestural communication in the guidance task as $P7_0$ 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 was confirmed by $P1$ who asked his partner to physically move to a location with better visibility.

Except for $P4$ and $P8$, participants in the *AR Go-Go* condition used the ray 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. This technique allowed them to adjust first the orientation (2 DoFs) and then one translation (1 DoF), the latter being adjusted using the ray direction. 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 $P1_0$ (actor) asking multiple times to his partner if his ray is aligned with the target, and $P1_1$ (guider) asked his partner to align the ray with his finger, before adjusting the depth and notifying his partner when the trajectory was correct. $P6_0$ suggested to his partner to first guide her with 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 [11].

WIM 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 Section 4.5.5. $P0_0$, $P1_{0,1}$, $P4_1$, and $P6_0$, when helping their partner, first looked at and guided the partner through the original dataset, before focusing on the copy version. This may be explained by 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. $P9_0$ and $P11_1$ asked if they can interact with the original in the solo tasks for better accuracy, instead of with its copy. $P8_{0,1}$, 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. $P1_1$, $P10_{0,1}$, $P11_0$ did the same in the solo tasks. This specific behavior made *WIM* and *Manual* similar, which might explain why results reported in Section 4.5.5 may contradict **H4**.

Others just used only the original dataset ($P0_{0,1}$, $P6_{0,1}$, $P9_1$, $P10_1$) 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 actions. When users looked at the original and not the *WIM* version, one may assume that they lose the directions axis (e. g., forward and left axis). However, guiders managed to guide their partners using these directions, saying sometimes "closer to me" in an efficient way without ambiguity. For unknown reasons, $P11$ performed half of the trials using the original (sometimes starting with the copy and then heading towards the original) and the remaining trials using the copy only. Creating two versions of the dataset, however, made participants sometimes look at both simultaneously which can confuse them as pointed out by $P1_0$, and may disable deictic communication. $P0_0$ and $P10_1$ said that they cannot see their partner's pointing action when they were manipulating the copy. In comparison, $P9_0$ commented that he looked at both versions of the dataset to both target the point in the WIM and see his partner's pointing actions in the original dataset, helping himself with the replicated 3D cursor. One idea would be to render the guider's 3D cursor to support pointing actions such as "just here."

4.5.6 Second Experiment—Solo Tasks

To the best of our knowledge, the *Manual*, *Go-Go*, and *WIM* techniques together have not yet been compared 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. Again, 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 [121], see Figure 4.7.

While our main experiment studied collaboration, this second experiment thus compared the techniques' pure performance and made our results comparable with experiments that compare exocentric metaphors with direct manipulations or with other ray+cursor techniques.

Pre-registered Results

We now analyze our questionnaire responses and logged data of this second experiment the same way as in Section 4.5.5. The computed pairwise comparisons are again stable and our main conclusions remain unchanged regardless of outlier treatment, for both speed and accuracy.

Accuracy. We again see (Figure 4.5b) strong evidence for *Manual* and *WIM* being more accurate than *Go-Go*, and *Manual* being more accurate than *WIM*. *Manual* is about $4\times$ as accurate as *Go-Go*, *WIM* about $2\times$ as accurate as *Go-Go*, and *Manual* about $2\times$ as accurate as *WIM*. *AR Go-Go* is then the least accurate technique, supporting **H1**, and *Manual* the most accurate one, supporting **H5**. The ranking questionnaire results (Figure 4.9) yield similar conclusions.

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, similar devices could be a prerequisite for collaborative 3D data exploration tasks as explained in Chapter 3. $P6_0$, 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). When looking at pair-wise comparison ratios (Figure 4.6b), there is 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. Figure 4.7 shows the subjective workload results based on the NASA-TLX questionnaire. The figure shows strong evidence that, to our participants, *Go-Go* feels the least performant, the most difficult and stressful, and the most physically and mentally demanding technique. It 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 the *Manual* technique. We found no evidence regarding the stress, difficulty, time pressure, or physical and mental demand metrics between *WIM* and *Manual*.

4.5.7 Discussion

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

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.

Message Understanding

The low self-estimated performance with *Go-Go* (Figure 4.7) 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 substantially reduced the mutual understanding provided by the technique in the same way the "co-presence" did. Future research is needed to investigate that phenomenon.

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 ($\approx 1cm \pm 1cm$), it 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 to the multi-touch tablets, either to scale up the dataset, or to use other 2D interaction techniques [27, 29].

AR Go-Go. The Heisenberg effect [51] amplified by the ray metaphor may explain why *Go-Go* is the least accurate technique. Even if the displacement is still acceptable ($\approx 3cm \pm 1cm$), it creates frustration as $P5_1$ pointed out. This may have strongly impacted the perception of the tracking accuracy reported in Section 4.5.4. 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 [174]. This technique may then be good only for pointing, e. g., to show remote locations. One may say that better tracking system may solve this issue. However, by modifying a spatial jitter into an orientational one, the accuracy of *Go-Go* will always be lower than that of *Manual*. Moreover, it is known that orientational jitter affects people's behavior and speed, even with high-quality tracking system such as the Lighthouse system used by the HTC Vive [24]. Moreover, 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 [77] 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 an effect on accuracy. We expected *Manual* to be $2\times$ as accurate as *WIM* in both conditions as the dataset's copy was half 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*.

Speed

The lack of strong evidence for *AR Go-Go* being slowest in the collaborative condition may indicate that (1) the "guiding" step of the task takes more time than the "clicking," (2) that the ray metaphor better supports direction instructions (see Section 4.5.5), 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 was difficult to recognize by the system. They tended to reduce the click amplitude due to accuracy issue, it becoming 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 Figure 4.9), not supporting **H3**.

We highlight that our tracking remained stable, with a mean accuracy of ≈ 0.5 cm for a TCT of ≈ 10 s for *Manual*, the direct interaction technique.

Parallel Work (Hypothesis)

We asked participants to state their intuition about the parallel work potential of each technique. Because we had not designed the tasks to study such a scenario, we only report our hypothesis about this metric based on these responses and our observations. We hypothesized, based on the rankings and the comments, that *Manual* may not be suitable for parallelism because of workspace obstruction and tracking conflicts when both users would manipulate the environment, as reported by $P6_1$. 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. In comparison, most participants stated that *WIM* may allow them to work in parallel as it does not obstruct the view, while facilitating accurate-enough input. Nonetheless, questions remain about whether the appearance and disappearance of the *WIM* are disturbing for participants and whether the cloned 3D cursor should be placed inside the original dataset.

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 $P8_0$. 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 Section 4.4.3), smaller and generally placed at a comfortable 3D position, which reduce arm fatigue. $P6_{0,1}$ and $P9_0$, however, wanted to modify the copy's position by dragging its edges, and $P9_0$ 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 Figure 4.7) with it, which supports **H2**.

Some participants in *Manual* put down their tablets on a desk for the solo tasks. Indeed, the tablets 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.

4.6 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. This work is part of my PhD for which I already explained the combinations of multi-touch tablets with AR-HMDs (see Chapter 3). 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 exploratory 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 as we discussed. 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 [53, 103, 223]. 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 telling that a miss-click occurred, or by outliers removals. 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 results, modifying slightly the time required (which was particularly true for *AR Go-Go* where participants tried to stay still as much as possible), 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 our participants had not received dedicated training for how to collaborate with their partners. We purposefully did not train users for this aspect because, in real life, users often collaborate with changing partners. Our time completion task results then reflect such scenarios, and may have to be re-evaluated for people who frequently collaborate with the same partners and 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.

Some of our parameter choices may also have affected the results. For instance, a smaller/ bigger *WIM* may substantially affect the collaboration, fatigue, and accuracy, while an *AR Go-Go* with a lower gain-factor may affect people' preferences. Scenarios with strong changes may thus lead to different results. We chose these parameters on purpose, however, to satisfy meaningful constraints, such as being able to see the *WIM* entirely (its basic purpose) and to have a suitable range for *AR Go-Go*.

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.

4.7 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 an AR technology which relies on markerless tracking system, which seems to be the method that standalone HMDs will be based on 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 non-remote 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. The *AR Go-Go* technique 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 apparition of proxy objects may disturb collaborators who do not expect these. We then saw weak evidence that the *Manual* technique provided a better co-presence and supported users' comprehension better than the *WIM* technique, even if the co-presence results are satisfying in all cases.

This research project was possible because the VR literature, or in a more general way, the immersive analytics research area, extensively studied pointing techniques. Still, interaction techniques from VR are not always directly transposable to AR, as the real world in AR is not occluded compared to VR. In the next chapter, I am focusing on 3D volumetric (RoI) selections. As this exploratory tool was not extensively studied in VR, I am focusing on one-user environments based on my system. I am then giving a discussion (see Chapter 7) and primary hypotheses about the capabilities of volumetric selection interaction techniques in collaborative environments based on this research project tackling 3D point specifications. Drink a coffee and take a break. Next teleportation: again a new AR world!

5

CHAPTER

The Specification of Regions

This chapter is based on a work I submitted at Proc. EuroVis 2022 [264] with my co-authors Stéphane Gosset (former intern under Tobias Isenberg and my supervision), Lonni Besançon, and Tobias Isenberg. While Chapter 4 explores the specification of 3D positions, this chapter takes a step further and discusses about the specification of regions. Any use of "we" refers to the aforementioned authors.

In this chapter, we study the tangible aspect of the multi-touch tablets we have combined with AR-HMDs, as described in Chapter 3, to perform spatial 3D selections. We are again primarily interested in the exploration of 3D unstructured datasets such as cloud points or volumetric scientific datasets. Because AR-HMDs merge the visualization, interaction, and the users' physical spaces, users can also use the tablets as tangible objects in their 3D space. Nonetheless, the tablets' touch displays provide their own visualization and interaction spaces separated from those of the AR-HMDs. This raises several research questions compared to traditional setups. In this chapter, we theorize, discuss, and study different available mappings for manual spatial selections using a tangible touch tablet within an AR-HMD space. We then study the use of this tablet in a 3D AR environment, compared to its use with a 2D external screen.

This chapter then targets **RQ3** that I stated as:

> **RQ3** To specify regions of interest, what are the implications of a tangible multi-touch tablet where its 3D position has meanings in the AR space compared to the original Tangible Brush which decouples the input and output spaces?

5.1 Introduction

In this project, we are interested in spatial selection[1] of 3D scientific datasets using the combination of an AR-HMD with a multi-touch tablet. We reused the system I described in Chapter 3 where I already stated the benefits of such a combination.

Such spatial selection—one fundamental exploratory tool [147,272]—is essential for users to specify regions of interest (ROIs) in unstructured datasets to explore them further, in particular because they often cannot be isolated with filtering by some property set. While the interactive spatial selection has been studied in the past, most research has focused on 2D rendering systems as opposed to immersive ones [39].

As I stated in Chapter 1, immersive AR-HMDs merge the interaction, visualization, and the users' physical 3D spaces together, and their immersive displays are suitable for 3D data visualization [102,132,157]. However, before the introduction of the recent Microsoft's HoloLens in 2016 (1^{st} gen.) and 2019 (2^{nd} gen.), computing power of standalone AR-HMDs could not allow

[1] We use the term "spatial selection" to refer on the user interaction of selecting data points based on their spatial properties

the exploration nor the visualization of such complex datasets. As this interaction environment is then new, many research questions remain regarding how to explore large scientific datasets and how AR-based exploration tools differ from past setups.

As previously stated, this PhD focuses on a hybrid environment which relies on multi-touch tablets combined with AR-HMDs to benefit from both input modalities (see Chapter 3). In this work, we further study this hybrid environment by investigating those forms of spatial selection in which users fully control what they can select, which requires a high Degree of Freedom (DoF) [38]. While mid-air 3D input suffers from instability [11, 12] and induces fatigue [128], the use of touch tablets for 3D tasks [190, 252], such as spatial selection [190], has shown benefits within immersive 3D systems.

We based our spatial selection on Besançon et al.'s [38] Tangible Brush, as it is flexible and also uses a tangible touch tablet. The Tangible Brush project is a project I co-authored while I was under- and, later, post-graduating. This original implementation, while being performed before I started this PhD, is the basics of this research project my co-authors and I describe in this chapter. Originally, my co-authors and I used traditional projected views shown on an external 2D screen to allow users to understand the 3D scene. The user inputs a 2D lasso using the tablet's touch screen which he or she can extrude in 3D by moving the tablet around. While this manual interaction showed benefits compared to a state-of-the art semi-automatic approach which combines the user input with computer-based decisions, Tangible Brush decouples the user's output and input spaces, as the output is rendered on an external 2D display and the input is entered in the user's physical space. This decoupling might have implications, such as the high reported mental effort from participants. Tangible Brush can be re-implemented in our new AR-HMD plus multi-touch tablet setup by replacing the projections on the external 2D screen with stereoscopic views via the AR-HMD. This new configuration yields numerous new research questions, as now the position of the multi-touch tablet has a meaning in the users' physical and visualization spaces. Compared to the original technique, this new setup might lower the mental demand from users as both the input and the output spaces—for mid-air interactions—are merged in the user's physical environment.

We thus studied this new setup and compared it with the original design. As this interaction technique is less studied in immersive environments compared to the specification of 3D positions (see Chapter 4), where the VR literature is extensive for the latest, we focus in this work in one-user environments and do not study collaborative scenarios to provide fundamental knowledge on which further research can build on. Our contribution in this work is threefold. First, we discuss and implement multiple position mappings that can be used for manual spatial selections using the tablet as a tangible device inside the AR 3D space. We based the discussion of the possible pros and cons of each mapping mostly on the directness of the interaction using Bruckner et al.'s model [57] (Figure 1.4) of spatial directness. Second, we studied these mappings to understand them better and to find the most suitable one in the AR environment. Finally, we compare our AR-based spatial selection technique using the selected mapping to the original Tangible Brush technique, to better understand the implications of merging the user's input space within the user's physical and visualization 3D spaces.

Our results show that the most direct technique, with respect to the AR 3D space, is not the most accurate nor the most preferred one amongst participants. Comparing then one remote but more direct interaction with the original implementation, our results show that the use of stereoscopic AR views substantially decreased our participants' workload compared to a setup with only projected 2D views, while both conditions had similar accuracy and speed.

5.2 Related Work

In our work we rely on past research on hybrid interfaces that includes immersive devices, AR-related input techniques, and manual spatial selection including the traditional Tangible Brush as we summarize below. I already provided motivations in Chapter 3 about the use of multi-touch tablets combined with AR-HMDs. I am thus focusing here on the tangible aspects of multi-touch tablets for such an environment whose interaction space is closely related to the user's physical space and the current display space [72].

5.2.1 Mid-Air Gestures and Tangible Interactions

AR-HMDs tightly connects the physical world with the display and interaction spaces for 3D physical interactions [72], e. g., some mid-air input and tangible interactions. This close connection may reduce the users' cognitive load [72] and be suitable for tasks that require input with many degrees of freedom (DoF) [16, 127] and tangible interactions. Bach et al. [16] showed that this tight connection might explain why AR-HMDs performed the fastest, while having similar error results than other forms of input for the placement of cutting planes, which needs detailed manipulations (4 DoF).

Surale et al. [252] explored the use of a multi-touch tablet in a Virtual Reality (VR) environment for CAD software. Depending on the usual trade-off between speed and accuracy, users prefer to manipulate their views and data either on the tablet or through the 3D interfaces of the VR environment. In our system, a user performs a 3D spatial selection by extruding a 2D shape in its physical space, similarly to the extrusions of CAD software. We use the touch tablet to input the 2D shape accurately and rely on the physical user space for the coarser extrusion which is a 6 DoF interaction. In addition to its function as a 2D interface, Surale et al. [252] also used the touch tablet for 3D tangible input, e. g., as a cutting plane. The choice of using the touch or the tangible property of the tablet depends on the task and the needed precision. Reipschläger and Dachselt [226] augmented a multi-touch tabletop with an AR-HMD for CAD sketching. Most interactions happen on the tabletop, where users draw and extrude 2D sketches, while the AR-HMD was used for 3D visualization. In our approach, we also use the AR-HMD mainly for 3D visualization and delegate most of the interaction on the tablet.

Compared to mid-air gestures using hands, passive haptic feedback and planar constraints of tangible tablet can improve the comfort and accuracy of 2D and 3D input [12, 80, 127, 190]. Montano-Murillo et al. [190] showed that using a real tablet compared to a virtual one in a VR setup improves the user's performance for complex pointing tasks, at the cost of speed for simpler tasks. Arora et al. [12] showed that 2D sketches placed inside the 3D space have better accuracy and fairness when using a 2D surface than drawing in mid-air, even when the system projects the mid-air input strokes into the plane of interest.

5.2.2 Volumetric/Spatial Selection

Many techniques to select parts of data in 3D space exist. We separate the selection of explicit objects from the spatial selection of subvolumes. The first group relies on pre-defined shapes or objects in the 3D space that users can explicitly select. The second one allows users to select some part of the 3D space [11, 38, 39]. For object selections, most of the interactions in immersive environments rely on the raycasting paradigm, sometimes by adding another degree of freedom [21, 114, 230, 235], e. g., our AR Go-Go technique presented in Chapter 4.

We focus here on unstructured datasets which do not include explicit objects but instead sample physical properties of, e. g., particles or volume cells, which one can filter to make selections. Jackson et al. [141] filter vector fields based on a tangible cylinder's orientation. Hurter et al. [132] brush paths using two VR controllers and select paths that connect both targeted points. Akers et al. [6] filter brain white matter pathways using 3D boxes associated to Boolean operations. Yu et al. [297] select regions based on the density of the field and 2D lasso input. Montano-Murillo et al.'s [190] hybrid VR-HMD plus tablet setup allows one to select points in cloud point datasets. The user first frames the ROI using the touch tablet and then uses the tablet as a magical lens to target at with a VR controller and raycasting. Transfer functions can also extract ROIs [116, 154]. Wiebel et al. [288] selects a ROI based on the 3D color and opacity field as defined by the transfer function, the virtual camera transformation, and the user's 2D input. While transfer function manipulations and Wiebel et al.'s algorithm [288] can be implemented on tablets, they often require time to find the suitable function, and it may not fully capture the user's intended ROI. Moreover, not all datasets possess suitable physical properties. We are thus interested in general techniques that work for every type of datasets (i. e., point clouds [67, 248], volumes [288], and vector fields [141]). We also cannot only rely on pre-defined geometries, e. g., boxes [273], as we want a flexible, data-independent way to specify ROIs. Moreover, we are interested in better understanding the users' mental models in AR environments, compared to non-immersive ones, for 6 DoF interaction techniques that make use of a tangible tablet. We thus rely on Tangible Brush [38] as we show next.

5.2.3 Tangible Brush in AR

The Tangible Brush [38] spatial selection technique relies on a tangible touch tablet. The user draws a lasso on the tablet and extrudes it in the virtual 3D space by moving the tablet around. The lasso must be entered on an orthographic projection to be accurate as perspective views do not allow users to understand the size of a drawn shape with respect to the objects in the scene, as their sizes on the screen depend on the position of the virtual camera. Moreover, an orthographic 2D projection likely leads to less users' cognitive load and higher accuracy as both its interactive and its visualization spaces are in 2D [16, 57].

Because the tablet does not provide any depth cues, my co-authors and I in the original implementation et al. [38] provided a second stationary 2D screen which gives the position of the virtual camera in the virtual space. We originally compared Tangible Brush with SpaceCast [297], a semi-automatic technique that infers the selection based on minimal user's input. Tangible Brush is slower but it is more accurate and versatile. However, we originally recorded a high mental load which may be explained by the decoupling between the input (3D tangible interactions) and the output (2D external screen) spaces on which Tangible Brush relies. In this new work, we thus modify Tangible Brush by replacing the stationary screen with an AR-HMD. While VR-HMDs work as well, users can more comfortably move in their physical space with AR-HMDs because they can see the real world. Drawing on a touch tablet that users can see, moreover, yields better accuracy than drawing on a physical surface visually replicated inside the VR-HMD [12]. CAVEs are also less practical for end-users due to their high maintenance costs [284]. These motivate the use of AR-HMDs over other forms of immersive systems, making this interaction suitable and implementable in the overall system I described in Chapter 3.

5.3 AR Concept and Differences to Tangible Brush

In this work, we adapted the traditional Tangible Brush interaction as follows. In the traditional Tangible Brush interaction [38], the user draws a 2D lasso on a multi-touch tablet that he or she then extrudes into 3D space. The coordinate system was each time reset based on the orientation of the tablet as there was no absolute 3D coordinate system. In contrast, Augmented Reality possesses an absolute coordinate system which corresponds to the user's 3D space. We can then map user's movements in many different ways as we discuss next. Figure 5.3 shows the actual representations of the three AR-based approaches we implemented. To describe and analyze

Technique	Focus	*vi*	*om*	Class	Axis aligned
Original	Tablet	quasi	no	5	**Yes**
Original	External Screen	yes	no	4	No
NA	Tablet	quasi	no	5	**Yes**
NA	HMD	yes	yes	1	**Yes**
RF	Tablet	quasi	no	5	**Yes**
RF	HMD	yes	no	4	No
RA	Tablet	quasi	no	5	No
RA	HMD	yes	yes	1	**Yes**

Table 5.1: Classification summarization of the different techniques and mappings using Bruckner et al.'s [57] model (Figure 1.4).

both the original Tangible Brush technique and our own AR-based approach, we use Bruckner et al.'s [57] model of spatial interaction directness (Figure 1.4). We describe both parts of the interaction (sketching + extrusion) and discuss both sets of input + output devices (2D tablet + 2D display vs. 2D tablet + 3D AR) to describe the differences of both set of output modalities.

5.3.1 Lasso Input

The lasso step of both techniques is identical. A user draws the lasso for later extrusion on the provided tablet, which uses an orthogonal projection. Here we consider i to be an inverse of v as a 2D position on the tablet corresponds to the near-clipping plane of the camera, and vice-versa. The orthogonal mapping naturally allows users to measure distances, in contrast to the use of perspective projection [38]. Moreover, $\mathcal{O}$ and $\mathcal{M}$ collapse due to 2D input being captured virtually on the tablet's display. We thus classify this interaction as belonging to Class 1.[2]

5.3.2 Original Extrusion Interaction

In the original implementation, the user moves the tablet and its lasso in the physical space, yet the user can only observe the selection effects on a 2D screen that shows the (partially selected) dataset and the tablet's motions in two perspectively projected views. The local coordinate system of the tablet (which is needed to map from $\mathcal{M}$ to $\mathcal{I}$) was reset each time the user pressed the extrusion button, with the forward axis being defined by the normal of the physical tablet during the interaction, and the x- and y-axes being defined by its screen physical orientation. We call this mapping ***Relative-Full*** (RF). For both screens we thus have a three-dimensional interactive space (both $\mathcal{I}$ and $\mathcal{M}$ are in 3D), yet two distinct 2D output spaces $\mathcal{O}$. We analyze the spatial directness for both cases next.

Tablet as the focus: The tablet renders an orthographically projected visualization $\mathcal{V}$. When focusing on the tablet, we consider vi to be quasi-inverse as motions in the interactive space $\mathcal{I}$ along the tablet's normal may lead, given the near- and far-clipping planes, to the same visualization $\mathcal{V}$. Moreover, om is not the identity as the 2D output $\mathcal{O}$ has to be related mentally to the 3D tangible device $\mathcal{M}$. Because the physical tablet ($\mathcal{O}$) shows a positionally correct

[2] Bruckner et al. [57] classified interactions as belonging to Classes 1–6, where Class 1 interactions are the most and Class 6 ones the least "direct."

projection of the 3D $\mathcal{V}$ as positioned in the similarly 3D $\mathcal{M}$, however, helps users in their interaction which we classify as Class 5.

External screen as the focus: When focusing on the 2D screen instead, we consider $\mathcal{V}$ and $\mathcal{I}$ to be identical: forward movements of the physical and, thus, also the virtual tablet can directly mapped to translations in the visualization space. Like before, om is not the identity due to $\mathcal{O}$ being the 2D screen and $\mathcal{M}$ being the physical 3D space. Moreover, the lack of axis alignments between $\mathcal{O}$ and $\mathcal{M}$ during most extrusion processes can result in a high mental load. Indeed, based on previous work [175, 285], we hypothesize that an existing or missing alignment can impact a user's mental model and, thus, the ease of creating space $\mathcal{U}$: people manipulate input devices ($\mathcal{M}$) based on what they see ($\mathcal{O}$) and their mental frame of reference ($\mathcal{U}$). This is especially true for novel interactions that users need to learn, as opposed to trained interactions such as mouse input where, even though $\mathcal{M}$ and $\mathcal{O}$ do not align (with a vertical screen and a horizontal mouse), users have an established mental model via training. We classify the Tangible Brush with an external screen as Class 4.

In summary, the original Tangible Brush uses indirect interactions, in particular due to the dimension differences between spaces $\mathcal{O}$ and $\mathcal{M}$ for both screens. Moreover, $\mathcal{O}$ and $\mathcal{M}$ lack alignments in the x- and y-axes when users focus on the external screen. This analysis may explain why Besançon et al.'s [38] participants experienced a high mental load, as directness is usually linked to the users' mental model, especially for novel interaction paradigms [57].

5.3.3 Extrusion with AR Tangible Brush

The main difference of our AR Tangible Brush to the traditional setup is that we replace the 2D external screen with an AR-HMD. Users can then see the data stereoscopically in front of them, with the input (i. e., the tablet with the lasso) being possibly at the physically and virtually correct position. We can now define three different mappings for the actual realization of our AR-based Tangible Brush selection (see below). All three cases have in common that users need to set the virtual camera for the tablet (which still relies on an orthographic projection), before actually drawing the lasso. This allows them to relate their extrusion shape to the dataset. We use López et al.'s [175] approach of freezing the 2D orthographic view on the tablet for a given position, to then allow the users to draw their lasso on a static view for precise input. Next we introduce and discuss the three mappings for the actual selection input using Brucker et al.'s model.

Naïve Approach

As a straightforward or ***Naïve Approach*** (NA), we map the physical position, orientation, and size of the tablet in the user's physical 3D space directly to their virtual counterparts.

Tablet as the focus: If the tablet is the focus, the situation is almost similar as for the traditional Tangible Brush and we fall back at classifying it as Class 5 because om is not the identity. Moreover, $\mathcal{M}$ and $\mathcal{O}$ do not share the same axis orientation anymore. vi is quasi-inverse due to the orthographic projection used on the tablet.

AR-HMD as the focus: If the AR-HMD is the focus, $\mathcal{V}$, $\mathcal{I}$, $\mathcal{O}$, and $\mathcal{M}$ collapse, corresponding to Bruckner et al.'s Class 1. While it is the spatially most direct approach, NA has limitations. First, the size of the tablet (10.5" diagonal for the Samsung Galaxy Tab S4 that we use) is a strong limitation, as most 3D visualizations would have to be scaled down to "fit inside the tablet" and the global size of the data representation constantly has to be adapted to the intended selection. Second, users sometimes need small targeted selections. As visualizations can be large in size, a user may lose the overall spatial context as they need to be physically close to the part area to be selected. This may lead to parts of the data not fitting in the user's field-of-view anymore. Finally, interacting inside an intangible object like a 3D visual representation can feel uncomfortable due to the vergence-accommodation conflict [64, 226]. Moreover, the tablet possesses an active screen which is then hidden by the AR-HMD, leading to conflicts between the two. These led us to create a virtual tablet selection different in size, position, and orientation, compared of its real counterpart that uses relative motions. We propose two types of relative motions: ***Relative-Full*** (RF) and ***Relative-Aligned*** (RA), as we discuss next.

Relative Mappings: Relative-Full and Relative-Aligned

For RF and RA, the user first initializes the virtual tablet within the 3D visualization ($\mathcal{V}$) space using NA. They then move the physical tablet in the 3D manipulation ($\mathcal{M}$) space and observe the motions of the virtual tablet in the visualization ($\mathcal{V}$) and output ($\mathcal{O}$) spaces, to either start an extrusion or to replace the virtual tablet from afar. Those mappings are then suitable for remote interactions on datasets that require large overviews. For RF, similar to the original implementation, we use a clutched interaction so that the tablet's actual orientation may differ from its virtual alignment. In contrast, the physical and the virtual tablets have the same orientation in RA.

Tablet as the focus: As for the original Tangible Brush technique, we classify RF and RA as Class 5 because the dimensionality of $\mathcal{O}$ and $\mathcal{M}$ differ from each other. In the RA mapping, however, similar to NA, $\mathcal{M}$ and $\mathcal{O}$ do not share the x- and y- axis orientation, making this mapping even more spatially indirect than the RF mapping.

AR-HMD as the focus: When focusing on the 3D AR view, $\mathcal{O}$ and $\mathcal{M}$ collapse for RA but do not for RF, the latter due to the difference in axis orientation. Indeed, the user redefines a new reference coordinate system for the manipulations in $\mathcal{M}$ each time they start the extrusion process using the RF mapping, compared to their (world) coordinate reference in physical space. RA, in contrast, ensures that the reference of the 3D motions in $\mathcal{M}$ match the coordinate system of $\mathcal{O}$. In both cases vi is inverse. We thus classify RA as Class 1 and RF as Class 4.

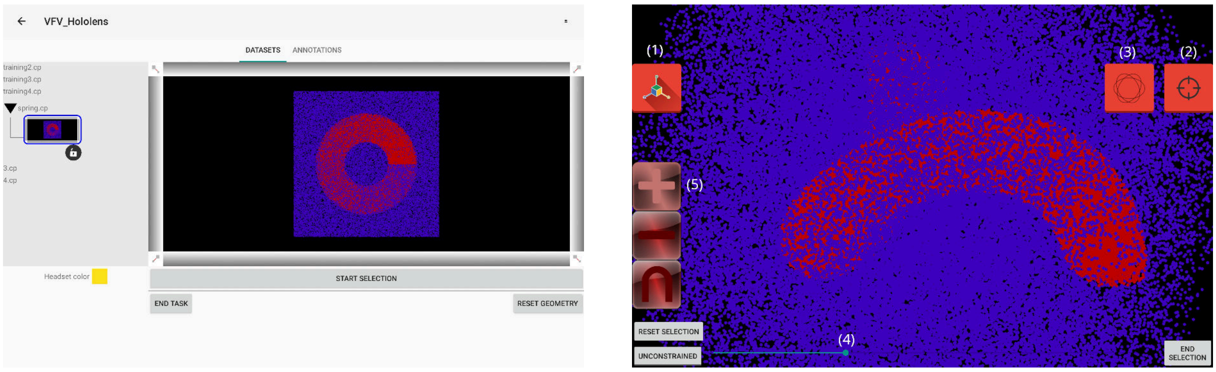
In summary (see Table 5.1), NA is the spatially most direct method: it merges, for what concern the AR-HMD, the interaction/manipulation and the visualization/output spaces within the user's physical space. NA possesses however multiple limitations as we described. RF should feel more comfortable than RA for a user if he/she concentrates on the tablet view due to the axis alignments between $\mathcal{O}$ and $\mathcal{M}$, and RA should feel more comfortable than RF if he/she concentrates on the AR space for the same reason. We also consider NA and RA to be "direct" forms of interaction (Class 1) when compared to the original implementation, even if one relies on absolute positions (NA) and the other on relative motions (RA). This difference makes however NA more direct than RA.

To better understand the implications of the difference of spatial directness that the AR-HMD induces compared to the original implementation, we need to study the different interaction mappings empirically. Such an experiment also promises to provide us with a better understanding of the user space $\mathcal{U}$ (i. e., the user's mental model) during our interaction, which may depend on the screen the user focuses on, due to the different level of directness.

5.4 Implementation

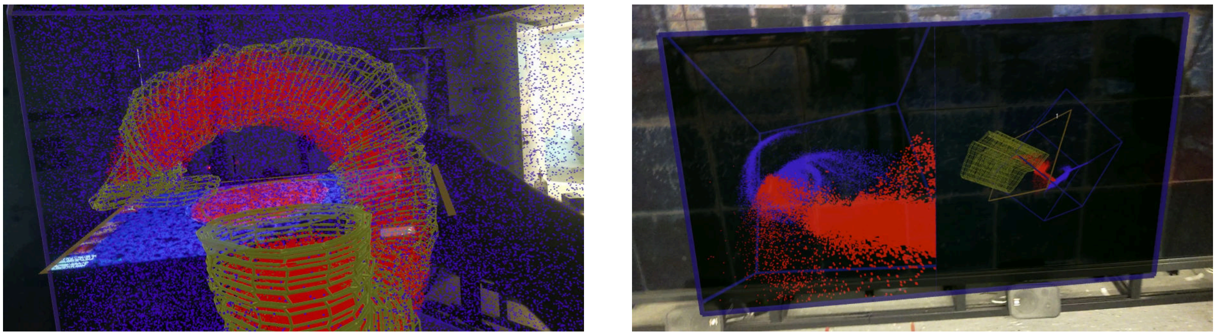
The overall system is motivated and described in Chapter 3 which this project relies on. We describe here only what is relevant to this project. The source code of this specific project can be found at `https://github.com/MickaelSereno/SciVis_Server/tree/volumeSelect` and `https://github.com/MickaelSereno/SciVis_Server/tree/volumeSelect2D`. The first branch corresponds to the source code we used in our first experiment, and the second branch to the second experiment.

We launch the server with the user's ID as a parameter. The server loads the needed dataset depending on how far the participant is in the study. To later analyze users' data, the server logs



(a) (b)

Figure 5.1: Screenshots of the tablet interface at the time of the user experiments, which slightly differs from what I presented in Figure 3.5. a is the main interface for general operations. When users start a selection, we show the Tangible Brush interface b on the tablet; (1) enables the Tangible mode, (2) Position mode, (3) Tangible Rotation, (4) is the slider to rescale the virtual tablet and its associated view, and (5) facilitate Boolean operations (OR, AND, NOT). The position mode is only available in the 3D condition, and the tangible rotation is only available in the 2D condition.



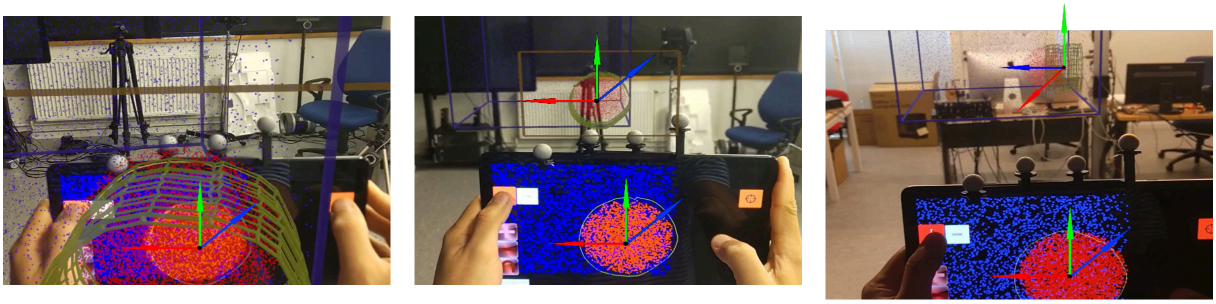
(a) (b)

Figure 5.2: Screenshots of the HoloLens' interface during extrusions for a the AR–3D condition and b the 2D condition.

all sent and received network events in JSON format. We can thus recover from crashes of the tablet or the HMD, should these occur, without affecting the results of the study. We extract useful metrics (e. g., speed, precision) from the logs using Python and R.

All datasets measure $50 \times 50 \times 50\,\text{cm}^3$ in the AR space. Users can move, rotate, and scale the visualization using FI3D widgets [298] on the tablet. In the AR conditions, the camera associated with the FI3D widgets faces the current visualization to always have it in sight. As the main view is the AR space and because the user can physically move around in the AR space, parameterizing the camera based on the user' position and orientation would be unusable. However, in our replication of the original Tangible Brush technique, the camera associated with the FI3D widgets follows the virtual tablet position and orientation as this condition always relates to this virtual tablet which is static most of the time (i. e., outside any selection processes).

The visualizations are placed at the origin of the HMD when the application is launched. Participants can relocate the visualization vertically (y-axis) for comfortable interaction, but we locked motions along the x- and z-axes to maintain similar conditions across participants in AR conditions. In our condition that replicates the original Tangible Brush, however, we allowed participants to move the dataset along all axes because, in the AR condition, users can move horizontally on the ground, which acts as x- and z-translations that were impossible in the



(a) (b) (c)

Figure 5.3: Representations of the three AR mappings we implemented. a depicts the Naïve Approach (NA). The user extrudes, in a direct approach, the drawn 2D shape. The position of the virtual tablet overlaps the position of the physical tablet. b depicts the Relative-Aligned (RA) approach. The orientations of the virtual and physical tablet with respect to the physical space is the same during an extrusion. However, the user controls the virtual plane with relative motions. In this example, the user has stepped back before starting the extrusion process remotely. c depicts the Relative-Full (RF) approach. This mapping creates two coordinate systems. Movements from the one defined by the screen of the virtual tablet are mapped to the coordinate system defined by the user's physical space. In this example, the user, after having placed the tablet in front of the dataset, has stepped back and rotated around the dataset which the user now faces sideway. As the user moves along the normal of the physical tablet, its virtual counterpart also moves along its own normal axis (the blue arrows).

original implementation. Rotations can only happen on the x- and y-axes, with the x-axis being defined by the user's current position in the physical space in the AR condition. While one can argue that a z-axis rotation can be useful, we note that the drawing of the lasso can be adjusted on the tablet, replacing a z-axis rotation. We made this choice to keep all the possible interactions (rotations, moving, scaling) simple in our overall system. Finally, we provided a button that resets the default positions and orientations for both the 3D dataset and the virtual tablet.

5.4.1 3D Tracking and Mapping of Position and Orientation for the Selection

Users can start a 3D selection by pressing a button on the tablet. We then show an outline that corresponds to the virtual tablet in the AR-HMD in yellow, and the tablet shows a full-screen orthographic projection of the data corresponding to its virtual position. Normally the physical motions of the tablet do not affect the virtual one, but users can update the size and the position of the virtual tablet on demand. Using spring-loaded "position" mode, they can reposition the virtual tablet in the 3D space using a 1:1 mapping. Alternatively, another spring-loaded "tangible" mode allows them to apply position offsets on top of the current mapping to either re-adjust the position of the virtual tablet or to extrude a drawn 2D lasso. Users can adjust the size of the virtual tablet using a slider, which modifies the outline displayed on the HMD and the orthographic parameters on the tablet. As in the original work, users can also use Boolean operations to adjust the selection (AND, OR, and NOT). Finally, again as in the original implementation, we provide a constrained extrusion mode that moves the virtual tablet only along its normal. This mode may improve the accuracy of the selection when it should be performed along the depth axis because it removes hand shaking and tracking noise along the other two axes.

We send the position and orientation of the HMD in its own coordinate system to the server at each frame, at a maximum frequency of 60 Hz. The server also receives the positions and the orientations of both the tablet and the HMD captured by the VICON at 60 Hz. The server then converts the tablet's coordinates into those of the HMD, the latter having been created by the HMD's private API. We then send these positions and orientations to the tablet which applies the current mapping. The tablet then sends back its updated position and orientation to the server, which passes them on to the HMD. While this design relies on several messages (VICON–Server + HMD–Server, Server–Tablet, Tablet–Server, Server–HMD), we did not perceive any latencies that would be caused by network traffic on our dedicated local network. With this design, all the decisions concerning the mappings and their implications happen on the tablet. This design allows us to change the interaction by only modifying the tablet's source code.

We derive the virtual tablet's orientation depending on the used mapping. Only for RF it actually differs from the orientation of the physical tablet, which we compute as $o_{beg} \cdot o_{pos}^{-1} \cdot o_{cur}$ using quaternions, where o_{beg} is controlled for tangible input, o_{pos} is controlled for position input, and o_{cur} is the current tablet orientation in the user's 3D space.

5.4.2 Volume Selection

After placing the tablet into the 3D space, the user can draw the lasso (to be extruded) on its static orthographic view. A rapid tap removes the lasso. To optimize the computation of the volumetric shape (computational complexity of *O(n)*, n being the number of points of the lasso), we enforce a minimal distance between points on the lasso of 0.05 units. We automatically close the lasso if its first and the last points are closer than 0.20 units. These distance units arise from OpenGL's camera coordinate system ($-1.0 \ldots +1.0$ along the screen's x- and y-axes).

For each movement during the extrusion, we compute the 3D positions of all lasso points p_i. For all time steps $t; t > 1$, we display a 3D wireframe corresponding to the selection on the HMD, which we update each time by connecting $p_{i,t}$ to $p_{i,t-1}$. We also ensure that two consecutive steps, on the wireframe, are at least separated by 5 millimeters to reduce clutter. We close the 3D selection volume when the user presses the "Done" button, selects another Boolean operator, ends the selection, or uses position or tangible rotation modes to reposition the tablet. This way, we ensure that we do not connect endpoints that the user did not intend to connect. Updating the 3D volumes does not consume much computing power, allowing us to run it in real-time, which is a strong requirement for AR applications [14, 15].

Compared to the original Tangible Brush implementation which updates in real-time the selection state of the visualization, users can check the intended selection volumes before they validate or invalidate them. Once validated, we compute the ROI selection on the server, which sends back the Boolean mask (one bit per data point) to each device. For a dataset of $160K$ points (i. e., the galaxy collision dataset in our study; Figure 5.4), the resulting mask represents $(160 \cdot 10^3 + 7) / 8 = 20$ kilobytes of data, which does not prohibit real-time interactions. The same applies for our Cabo Verde dataset I described in Chapter 3 which has $5.13 \cdot 10^6$ data points, resulting in 626 kilobytes of data for the boolean mask.

In our study, we use point cloud datasets to measure the accuracy of the techniques we compare. We render the points as cubes using a geometry shader on both output displays. When a participant validates a selection, we consider every volume and its associated Boolean operator. We determine, for each data point, if it is inside or outside a given volume using a raycasting algorithm. We cast a ray from the point into a random direction and, depending on how many triangles of the enclosing selection shape the ray crosses, the point is inside or outside the volume. For optimization purposes, however, we first segment our 3D visualization in a $16 \times 16 \times 16$ grid. Then we store, for each cell, the triangles that are part of a given cell. We then cast the ray from the point (which is at a position (i, j, k) in this 3D grid, and (x, y, z) in the 3D world), parallel to the x-axis. As a further optimization, depending on the maximum number of triangles the ray may collide with, we either cast the ray along the $-x$ or the $+x$ direction. We apply then a ray–triangle intersection algorithm to determine the triangles the ray intersects with, filtering the triangle lists using the segmented 3D grid. We discard triangles we already considered. If the number of triangles crossed is even, the point is outside the volume; otherwise, it is inside. We further parallelized the algorithm using OpenMP to process multiple data point simultaneously.

5.4.3 Replicating the Original Tangible Brush Implementation

For a manageable yet fair comparison, we simulated the original Tangible Brush technique inside the AR-HMD. We thus put a floating 1.22 m × 0.68 m virtual screen at a distance of 1 m from the participant. Because optical see-through displays cannot render black colors, we render a dark green background color to simulate an opaque virtual screen. We set the virtual tablet's position by default at the front-top-left corner of the 3D cube encapsulating the data, facing the center of the cube. We made this choice to avoid having to perfectly align the mathematically generated datasets that have axis-aligned features with the default tablet's position and orientation (along x-, y-, and z-axis).

Like in the original implementation, on the left side of the virtual screen we render a perspective view from the point of view of the current virtual tablet. Its field-of-view (fov) depends on the current size of the virtual tablet: $verticalFOV = atan(3 \cdot sizeX)$, with $verticalFOV$ in radian and $sizeX$ (the virtual tablet's width) in meter. On the right side of the virtual screen, again as in the original version, we render the scene using a birds-eye-view camera which we placed at (0, 1 m, -1.75 m) and which faces the center of the scene, where we place the dataset by default.

As in the original Tangible Brush, we apply translations with respect to the tablet's coordinate system, with the z-axis defined by the normal of the virtual tablet. This control is particularly comfortable when looking at either the tablet or the virtual screen's left side, as the axes of these respective cameras are aligned with the translation motion.

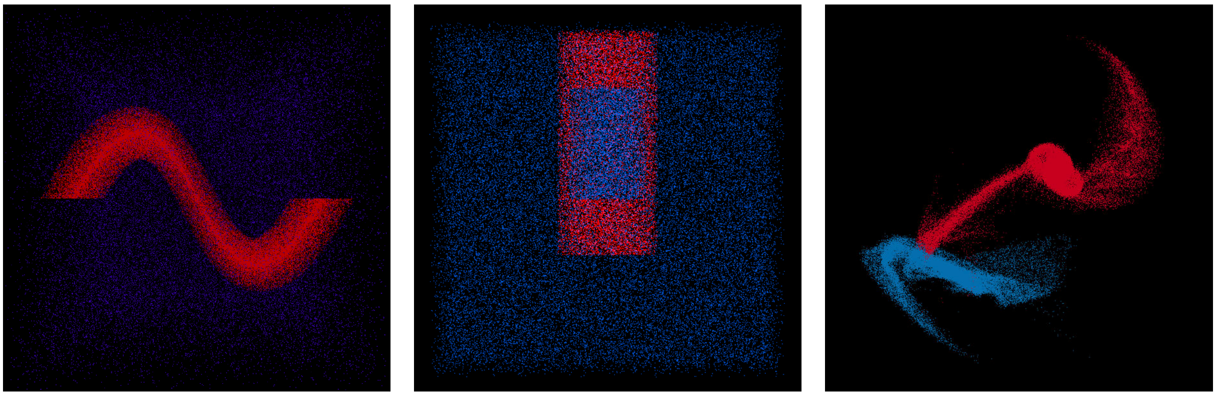
Based on a pilot study, we allow participants to rotate the 3D scene around the center of the dataset using the tangible properties of the tablet. Such rotations are strong depth cues which allow users to align both the tablet along an axis they can understand and the 3D datasets, and to review the 3D extruded-wireframed volumes within the virtual 3D space.

In summary, while the interaction of our replication is similar to the one we originally created [38], we modified the interface with respect to the following aspects, to accommodate the AR-HMD setup and to make the AR-based interactions more comparable to the traditional Tangible Brush:

- the interactions to move and rotate the datasets,
- the bird-of-view position to ease users with translations which do not rely on tangible movements anymore,
- the addition of 3D wireframe volumes for the participants to validate or invalidate,
- tangible rotation to adjust the 3D scene at any step of the extrusion (placing the tablet, drawing the lasso, extruding the 3D shapes, and reviewing the 3D volumes), and
- the possibility to scale the tablet's view.

5.5 Experimental Verification

As the AR setting possesses three different conditions and the original technique possesses only one, we conducted two distinct controlled experiments. We designed the first one to allow us to select the "best" AR mapping, while using the second one to compare this mapping with the original Tangible Brush technique. All the gathered data are available on our repositories, which



(a) (b) (c)

Figure 5.4: The three datasets our experiments relied on. Participants needed to select red dots while avoiding the blue ones. a depicts a helical spring shape, which represents non-linear objects like blood vessels, which we intended to force participants to perform unconstrained operations. This dataset also appears in Figure 5.1 and Figure 5.2a. b represents the outer shell of a cylinder and c the collision of two galaxies as a real use-case scenario. We reused datasets b and c from Besançon et al.'s [38] study.

also contain the pre-registrations for both experiments. We only did not pre-register the criteria for selecting the AR mapping for the second experiment because they depended on the results from our first experiment.

We recruited 18 participants per experiment. Each participant performed only one of the two experiments to avoid learning effects. We reused the Galaxies and the Cylinder datasets from Besançon et al.'s experiment [38]; see Figure 5.4b–c. As all Besançon et al.'s datasets rely on extruded shapes along constrained axis, we added a helical Spring as a third dataset for its non-linearity properties; see Figure 5.4a. Because of the COVID-19 pandemic and its associated difficulty to recruit participants, we decided prior to running the study to limit these experiments to three datasets instead of four. The first experiment showed that we were right at constraining ourselves with three datasets, as it could last up to 2.5 hours.

5.5.1 Tasks

For both experiment we used the same tasks. For each mapping, participants had three training trials with simple shapes. Then, after a break, they performed the selections on the three datasets (see Figure 5.4), while we measured the quantitative metrics. For each dataset, participants performed two consecutive trials, resulting in six trials per mapping per participant. We asked participants to select the red dots while avoiding the blue dots, and to be as accurate as possible without substantially influencing their speed. To avoid overly-long trials we explained that it was nearly impossible to have a perfect accuracy. We rendered selected dots with a lower saturation (i. e., tending to white) than non-selected dots, allowing participants to understand which dots are selected and their classification (to select vs. not to select).

5.5.2 Experiment 1: Best AR Mapping

We recruited 18 participants (age: MIN=20, MAX=45, AVG=26.33, SD=5.80) to study the three AR mappings in our first experiment, pre-registered at `https://osf.io/pwauq`. We documented the differences with the pre-registration on OSF (see the file Errata.docx). Participants performed all the trials for a mapping, before continuing with the next one. For a given participant, we presented the datasets in the same order for each mapping. We counter-balanced the order of the datasets using a cyclic Latin square on $(P_{ID}/3) \mod 3$ (integer division), and the order of the techniques on $P_{ID} \mod 3$, where ID ranges from 0 to 17 and is unique per participant. At the end of each mapping, participants answered the corresponding part of the questionnaire (see our additional materials). At the end of the experiment, participants answered the remaining general questions.

Participants were generally unfamiliar with AR (RANGE=1–5, AVG=2.17, SD=1.46 on a 5-point Likert Scale) and VR technologies (RANGE=1–5, AVG=2.11, SD=1.08 on a 5-point Likert Scale), except for the 5 AR or VR researcher participants, but rated the tracking of the tablet as accurate (RANGE=3–7, AVG=5.78, SD=1.00 on a 7-point Likert Scale).

Hypotheses

Our hypotheses and our reasons for them are as follows.

H1 NA will be the fastest approach, as it is the most direct one.

H2 NA will be the most physically tiring mapping where users need to hold the tablet high. Indeed, visualizations are expected to be aligned with the user's eyes or chest to be comfortable, similar to Bach et al.'s [16] study.

H3 The user' accuracy and cognitive load using RF and RA will depend on the output device on which a user focuses, as both mappings have different degrees of directness but rely on the same input modalities. If the tablet is the focus, RF will be more precise and requires less cognition than RA, and vice-versa otherwise.

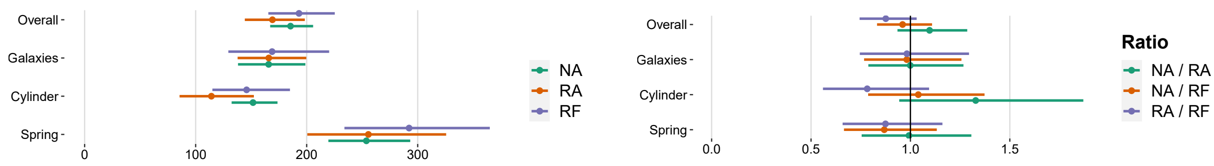
H4 However, users will need some depth perception to correctly move the tablet in the 3D space, where the AR view outperforms the view of the tablet. We then expect users to mainly focus on the AR view.

H5 Users will prefer RA. The selection often requires to scale up the virtual tablet as users cannot, otherwise, select what they are willing to because the frame will be too small. NA is then inappropriate to get the overview of the whole scene, as discussed in our designs. Moreover, the users will focus on the AR space (**H4**), making RA more suitable than RF (**H3**) due to the axis alignments between $\mathcal{O}$ and $\mathcal{M}$.

Results

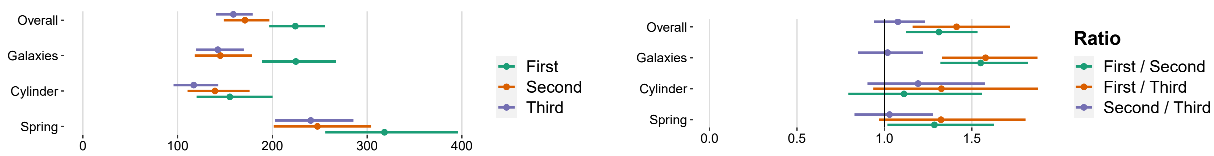
We gathered in total $nbParticipants \times nbConditions \times nbDatasets \times nbTrials = 18 \times 3 \times 3 \times 2 = 324$ data points. We report our results using estimation techniques based on 95% confidence intervals (95% CIs) instead of p-values, due to the recent APA recommendations [276] and the recent criticism of NHST to analyze experimental data [9, 35, 85, 274]. We additionally refrain from dichotomously interpreting our results as significant or not to follow current best practices and limit the risks of non-replication [35, 69, 124]. Krzywinski et al. [156] details a p-value reading of our approach. We removed one trial $(P_{ID}; Technique; Dataset; SubTrial) = (P_1; RA; Cylinder; 1)$ as the participant selected only a very small portion of the dataset, surely due to miss-clicks with the "reset selection" or the "end-task" buttons.

Due to issues with the HoloLens (e. g., opening the startup menu or losing itself in the physical space), we adjusted two further trials to remove the time that was lost to fix the system, removing 90 and 30 seconds for $(P_{ID}; Technique; Dataset; SubTrial) = (5; NA; Spring; 0)$ and $(7; NA; Cylinder; 0)$, respectively, whose durations we recovered from our video recordings. Finally, the HoloLens sometimes went into its saving power mode during breaks. Due to some potential errors, e. g., creating ghost TCP/IP connections, we relaunched all the components



(a) (b)

Figure 5.5: Raw data a and pair-wise within-subjects comparisons b of the Task-Completion Time in the first experiment. Error bars: 95% CIs.



(a) (b)

Figure 5.6: Raw data a and pair-wise within-subjects comparisons b of the Task-Completion Time following each participant's technique order during the first experiment. This analysis was not pre-registered. Error bars: 95% CIs.

(server, tablet, HoloLens) to start again from where the participant was, resulting in two sub-log files instead of one. This concerns P_1, P_3, and P_{11}.

Except otherwise stated, we computed all within-subjects pair-wise comparisons by bootstrapping geometric means on ratios. This allows us to visually compare the size-effect of the comparisons in percentages.

Completion Time We started the timer at the beginning of each trial. We removed the computation times of volumetric selections from the total amount of time spent in a trial. To correct for positive skewness, we analyzed log-transformed speed metrics [238] and present the confidence intervals of their anti-logged results in Figure 5.5, as it is the standard for such data [34, 37, 143]. From the vast overlap of CIs in both subfigures, we failed to find evidence that one mapping is faster than another, disproving **H1**. We experienced a learning effect (see Figure 5.6) as participants tried to find a strategy to solve the problems during their first respective mapping, before sticking to it for the other mappings. This is especially true for the Galaxies and the Spring as the solutions for those datasets were not obvious to find or to apply, compared to the Cylinder where participants understood directly that they can (1) select the outer part by extruding a cylinder, and (2) remove the inner part by extruding a smaller cylinder. At a lesser extent, users got used to the devices as all techniques use the same input $\mathcal{M}$ and output $\mathcal{O}$ modalities, strengthening again this learning effects. This may explain the lack of evidence.

While we did not pre-register the comparisons between datasets, the data shows strong evidence that selecting the Spring took longer to complete, compared to the other datasets. Participants found two strategies for this dataset. Some used multiple Boolean operators to mathematically define the spring. Others followed the spring using unconstrained unions. For the first strategy, the number of operations is higher than in the other datasets, resulting in a slower completion. For the second strategy, unconstrained operations need more control for accuracy, resulting in slow movements.

Workload We saw strong evidence that NA feels the least performant (Figure 5.7) and some evidence that RA feels the most performant to participants. We also found some evidence that RA requires less effort than RF. Participants P_{0-1}, P_{4-5}, P_{10-11} stated that they cannot correctly see both the 3D view and the tablet using NA when interacting inside the data. P_1 stated, however, that she felt more in control with NA after some trials, possibly be due to its higher directness.

Figure 5.7: Raw data a and pair-wise within-subjects comparisons b for TLX workload, 1st experiment. Error bars: 95% BCIs.

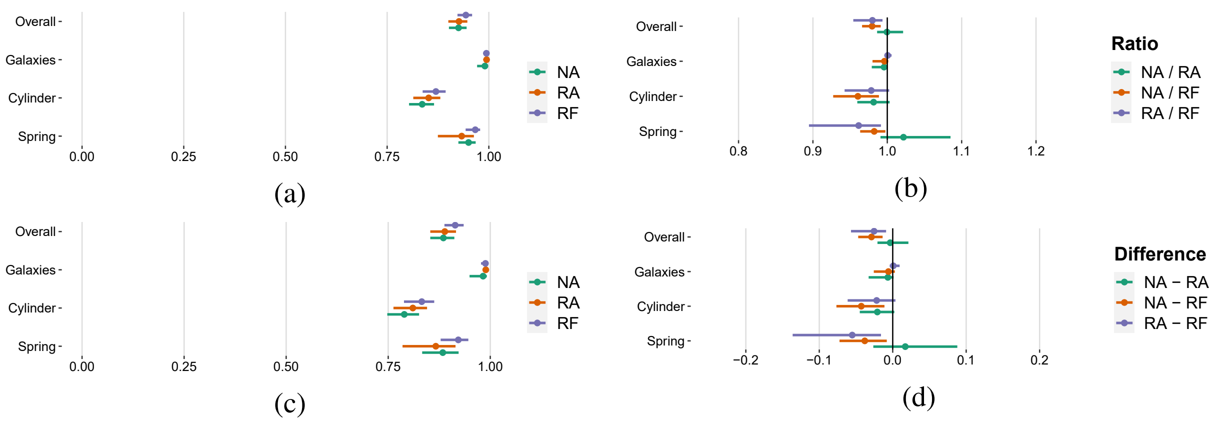


Figure 5.8: Accuracy for the first experiment. a and b correspond to the F1 score, and c and d to the MCC. a and c represent the raw values, and b and d within-participant pair-wise comparisons. Error bars: 95% Bootstrapped CIs (BCIs).

P_0 did not feel RA and RF to be different: she mostly relied on constrained operations, making both mappings similar. Finally, we found no evidence of a difference between NA, RA, and RF with respect to mental or physical workloads, not supporting **H2**.

Accuracy We computed the F1 and MCC metrics to represent the accuracy similarly to Besançon et al.'s [38] study. Both scores rely on three values: True Position (the number of correctly selected points; TP), False Negative (the number of points that should have been selected but were not; FN), and False Positive (the number of incorrect points that were selected; FP). $F1 = P \cdot R/(P + R)$ with $P = TP/(TP + FP)$ and $R = TP/(TP + FN)$. MCC considers also the number of particles that were correctly classified as such (True Negative; TN).

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

All the mappings have a similar accuracy (Figure 5.8a and c). Figure 5.8b and d show some evidence that RF can be more accurate than both NA and RA. There is also weak evidence that RA is more accurate than NA for the Cylinder and the Galaxies. The effect in both cases, however, is relatively small (i. e., less than 5% on average).

In our video recordings we saw that participants P_{0-1}, P_{3-5}, P_{7-10}, and P_{12-16}, for some trials with RA and RF, operated the 2D virtual plane in a sideways fashion, giving them a better perspective compared to the most direct mapping (NA). This functionality is useful, e. g., for the Cylinder whose inner part relies on two parallel planes easily visible from sideways. Using RF, P_{3-6}, P_{10}, and P_{12} placed the virtual plane in front of the dataset, before rotating around it by 90° to perform the extrusion. With the Cylinder, participants aligned the virtual tablet with the cylinder's circle side. They then physically moved around it to face the virtual plane at an angle

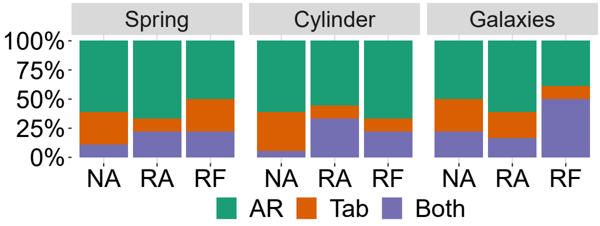
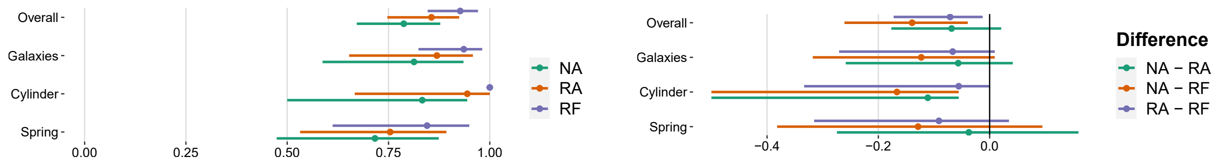


Figure 5.9: Proportion of which screen participants focused on during the extrusion steps for the first experiment. If participants have looked about equally at both screens, they crossed both answers on the questionnaire, resulting in a "both" value.



(a) (b)

Figure 5.10: Raw data a and pair-wise within-subjects comparisons b of the proportion of constraint operations that contributed to the final results over unconstrained ones for each condition of the first experiment. Error bars: 95% BCIs.

of 90°. They then moved the tablet during the extrusion toward the cylinder's rectangular side, which corresponds to the tablet's normal. The virtual tablet and thus the extrusion by extension then move along the depth axis of the cylinder, which corresponds to the virtual tablet's normal.

In the Galaxies data, users can select the red galaxy by extruding along the normal of the virtual plane defined by the blue galaxy, where users stop to maximize their accuracy. P_3, P_{9-10}, and P_{17} rotated the dataset before starting the selection. P_{17}, in addition, lowered the dataset to put the red galaxy at a comfortable position. He then selected it from top to bottom. P_0, P_3, and P_4 placed the virtual plane near the blue galaxy, before selecting the red one which was on top. They kneeled or lowered themselves to place the virtual plane and check its position. Finally, P_{11} selected the red galaxy using the unconstrained mode. In RA, he focused on the AR view sideways while moving the tablet around.

Focus Figure 5.9 shows the proportion on which of the two screens participants focused during their extrusions. We told every participant to mark both options if they have focused about equally on both screens. The figure shows that both screens were useful during extrusion. Still, except for RF with the Galaxies, participants mainly relied on the AR view, partially validating **H4**. We cannot, however, validate **H3** as we found evidence that RF is the most accurate mapping for participants who primarily relied on the AR view. This is especially true for NA (Spring and Cylinder), which we explain by its level of directness with respect to the AR view. As the Galaxies data contains much empty spaces compared to the other datasets, the screen of the tablet was also visible when users interacted within the dataset. This may explain why participants focused more on the tablet in the Galaxies, compared to the other datasets using NA.

The level of directness may also explain why NA and RA might invite more unconstrained operations than RF (Figure 5.10). We are also surprised to see that participants solved the Spring primarily using constrained operations (Figure 5.10), while we designed it to encourage unconstrained ones. P_0, P_9, and P_{17} stated that it was really difficult to perform unconstrained operations for this dataset, for which they spent a lot of time (Figure 5.5). P_{17} stated that he used only constrained operations using the relative mappings, just as he does with CAD software.

Constrained movements correspond to a pointing task from point A to B. Reflecting on studies about Fitt's law [19, 188] and P_9's comments concerning her screen focus, we hypothesize that users look at the AR view during their initial coarse but fast movements (ballistic phase [107]), and later at the view of the tablet for their refined movements. Indeed, the AR view facilitates a good understanding of the geometry of the dataset and its spatial relationship with the virtual

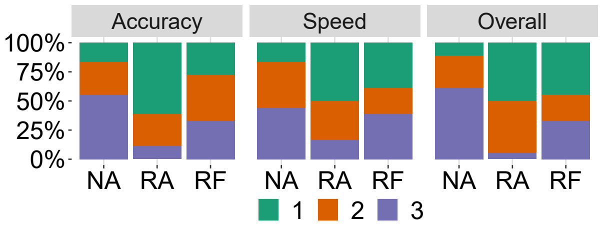


Figure 5.11: Proportion of ranking results for each condition of the first experiment; 1 is best.

plane. These facilitate fast movements towards the target, while the view of the tablet shows a precise position of where the virtual tablet is within its neighborhood (i. e., around its position) by acting as a clipping plane via its orthographic projection. We take this hypothesis into consideration for our second experiment.

Preferences We found evidence that the relative mappings were preferred compared to NA (Figure 5.11). Participants rated RA as being more accurate than RF, while the data suggests the opposite. We found strong evidence that participants preferred RA over RF overall. As participants mainly focused on the AR view during extrusion, we can confirm **H5**.

Opinions differ between participants with respect to RF and RA. P_1 did not fully understand how to move the tablet in RF. P_0 stated that, in RF, she could place herself in a suitable orientation. For example, after having placed the virtual tablet, she physically moved around the dataset to face the virtual tablet at the data's back side, and then moved toward the virtual plane such that the virtual tablet moved toward her. P_5 stated that RA was hard to use when she did not rely on constrained operations.

Discussion

The NA condition was not the preferred one (**H5**) mostly because of a lack of scene overview and because the AR hologram is rendered on top of the tablet, making both hardly readable. P_{10} used the relative mappings to place the tablet before the extrusion instead of using the "Position" button (which uses NA), to avoid this conflict. While we expected RA to be the most accurate (**H3**) with participants mainly focusing on AR (**H4**), the data shows that RF was the most accurate. However, our participants perceived RA to be more accurate than RF. Yet, the directness difference of RA and RF, when AR is the focus, may have biased the participants. Directness is then not always an indication for accuracy and may even bias users. Still, all mappings had a good accuracy. The main difference between RA and RF, user-wise, is due the subjective effort and mental model. Results show that participants preferred RA over RF, possibly due to the virtual and the real tablet sharing the same orientation in AR space, on which most participants focused during extrusion.

Some participants benefited from the two coordinate systems of RF ($\mathcal{O}$, the AR-HMD, and $\mathcal{M}$, the tablet). After placing the virtual tablet, they walked around the datasets by 90° or 180° before extruding their 2D shapes to have a better depth perception. This may explain its better accuracy compared to the other techniques and why participants using RF focused mainly on the AR-HMD. The users' performance may depend on the possible rotation values (e. g., 45°, 90°, or 180°), and these may be interesting to investigate in future work. The implications of such interaction may be different for different numbers of DoF a person uses simultaneously: whether the user is only performing a constrained translation (1 DoF), reorient an object (3 DoF) or plane (2 DoF), or combine all of them (5–6 DoF). Still, we saw some participants hesitating about how they should physically place/orient their body in the 3D space before extruding the lasso in RF, which may be due to the disconnect between $\mathcal{O}$ and $\mathcal{M}$, explaining our workload results.

Based on this experiment, we suggest to use relative mappings instead of NA. Users can still perform absolute operations by applying the relative mapping at a relative distance of 0. Using the "Position" mode, users can replace the virtual tablet once they have drawn the lasso before starting an extrusion. In the next experiment, we compared the original setup with this AR one. As the AR mapping, we selected RA as it was the most preferred one. One may have chosen to compare RF for both the original and the AR setups due to better accuracy. Should we find strong evidence that RA appears to be more accurate than the original setup in this second study, we can also be confident that RF as well will be more accurate than the original setup, hence our choice. Moreover, as all mappings reached similar accuracy, we preferred to focus this experiment on studying the users' mental model rather than their level of performance.

5.5.3 Experiment 2: AR vs. Original

We recruited 18 participants, as pre-registered at `https://osf.io/rvpuc`. We had to discard one participant who rushed the experiment and did not respect our experiment protocol (his results are still available on OSF). We replaced him with another person, for a total of 18 participants (age: RANGE=21–36, AVG=25.2, SD=4.1). The second experiment used the same protocol, tasks, and datasets as the first. We counter-balanced techniques (our *AR* technique and the original *2D* one) using a cyclic Latin square on $P_{ID} \mod 2$ and datasets on $(P_{ID}/2) \mod 3$. In the 2D condition, participants could use the complete set of translations provided by the FI3D widgets. Similar to the first experiment, errors in manipulating the HoloLens happened and some participants asked for additional breaks, resulting in multiple log files for participants P_{1--2}, P_6, P_8, P_{14}, and P_{17}. For user P_{14}, especially, the HoloLens' startup menu showed up too frequently when he manipulated the tablet. Another gesture to open that menu may be needed.

Participants are not familiar with AR (RANGE=1–2, AVG=1.39, SD=0.50 on a 5-point Likert Scale) and VR technologies (RANGE=1–3, AVG=1.67, SD=0.69 on a 5-point Likert Scale), but rated the tracking of the tablet as accurate (RANGE=3–7, AVG=5.44, SD=1.10 on a 7-point Likert Scale).

Based on our theoretical analysis (Section 5.3) and on our first experiment (Section 5.5.2), we hypothesized the following:

H6 *AR* outperforms *2D* both in cognitive load and users' preferences as the users now have better depth perception through the 3D stereoscopic view, and that *AR* is more direct than *2D* when focusing on the AR view.

H7 For the same reasons, and because the input modalities are similar, *AR* is more accurate than *2D*.

H8 Moreover, as *AR* allows users to place the virtual tablet in a 1:1 mapping prior extrusions, it is faster than *2D*, which needs to either place the tablet using relative translations, or to move the dataset using the FI3D widgets.

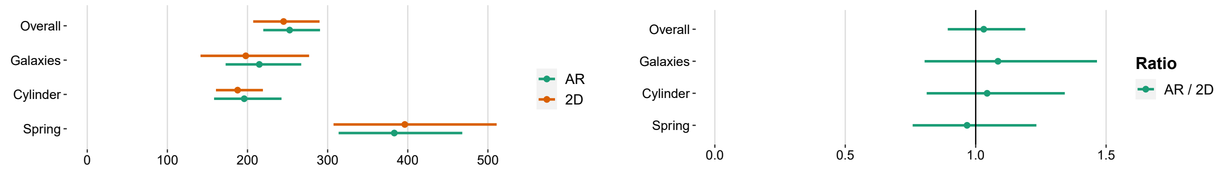
H9 During extrusions, users use the AR view for coarse movements and use the tablet's view for refinements.

Results

We use the same reporting and analyzing strategies as for our first experiment 1 (see Section 5.5.2).

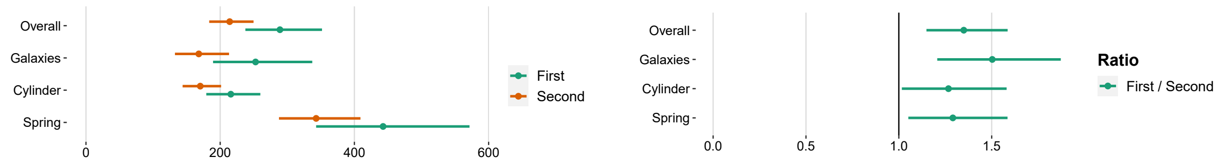
Completion-Time Similar as in the first experiment, we did not find evidence that one technique would be faster (see Figure 5.12), disproving **H8**. We again observed a learning effect across all datasets between the first and the second techniques that participants encountered as they followed their technique order (see Figure 5.13).

In the *2D* condition, participants sometimes placed the visualization (using the FI3D widgets) with respect to the fixed virtual tablet, instead of placing the virtual tablet with respect to the fixed



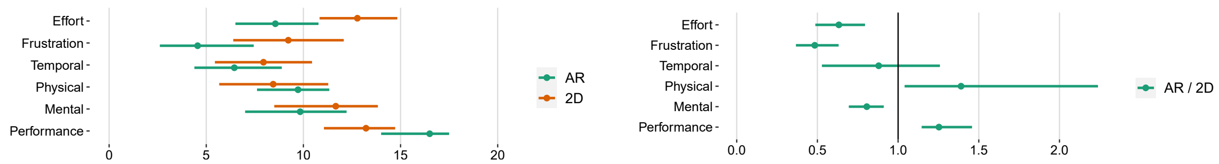
(a) (b)

Figure 5.12: Raw data a and pair-wise within-subjects comparisons b for Task-Completion Time, 2nd experiment. Error bars: 95% CIs.



(a) (b)

Figure 5.13: Raw data a and pair-wise within-subjects comparisons b of the Task-Completion Time following each participant's technique order during the second experiment. This analysis was not pre-registered. Error bars: 95% CIs.



(a) (b)

Figure 5.14: Raw data a and pair-wise within-subjects comparisons b, of the TLX workload results for the second experiment. Error bars: 95% BCIs.

visualization. Participants placed the dataset quickly by rotating it correctly, before translating it on the depth axis of the virtual plane, resulting in few but accurate operations for the placement. Finally, we again see strong evidence that the Spring dataset was the slowest to solve.

Workload We found strong evidence that users felt more performant, were less frustrated, needed to work less hard, and needed less mental effort at solving tasks using *AR* compared to *2D* (Figure 5.14). We also note that the difference is quite large in most cases. This may be due to the AR view offering strong spatial cues about the spatial relationship between virtual tablet and 3D visualization, due to the difference of the directness of the techniques, and because the AR view helps users to understand the 3D visualizations. We can thus confirm **H6**. The lack of evidence for differences in physical demand may be due to both techniques relying on the same input modalities.

To reduce his workload, P_0 rotated the whole scene in the 2D condition (using "Tangible Rotation") to arrange the plane of the virtual tablet to be parallel to the view shown on the external screen for the Cylinder and the Spring (where he used only constrained movements). This allowed him to reduce the problem along one dimension to remove, e. g., the inner part of the Cylinder. P_7 used the same strategies all over her trials in the 2D condition.

Accuracy Both techniques are similarly accurate (Figure 5.15). *AR* seems to be more accurate than *2D* for the Galaxies and the Cylinder, but with a small size-effect ($< 1\%$ for the Galaxies, and $\approx 3.6\%$ for the Cylinder). As users can observe more closely in the AR space compared to the 2D screen which does not propose any zooming; we suppose that they were able to

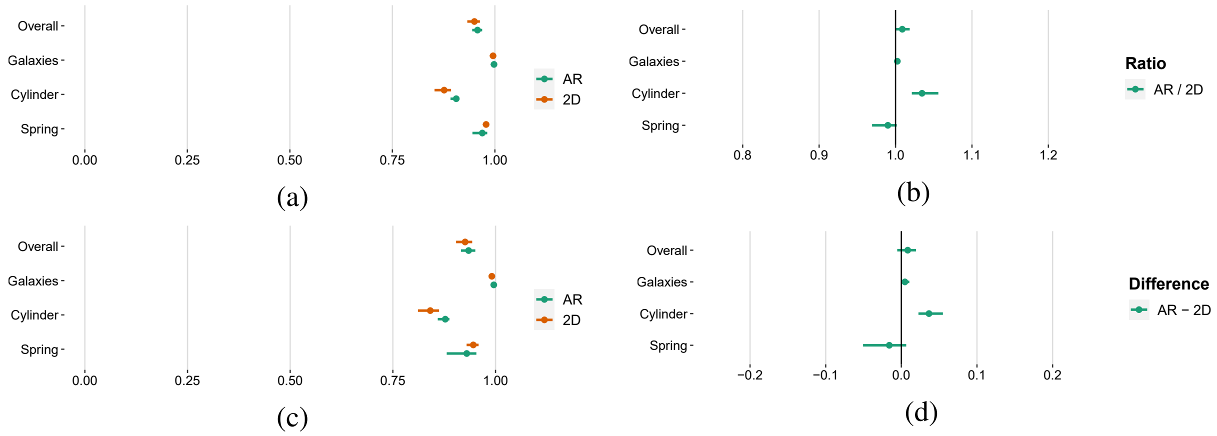


Figure 5.15: Accuracy for the second experiment. a and b correspond to the F1 score, and c and d to the MCC. a and c represent the raw values, and b and d within-participant pair-wise comparisons. Error bars: 95% BCIs.

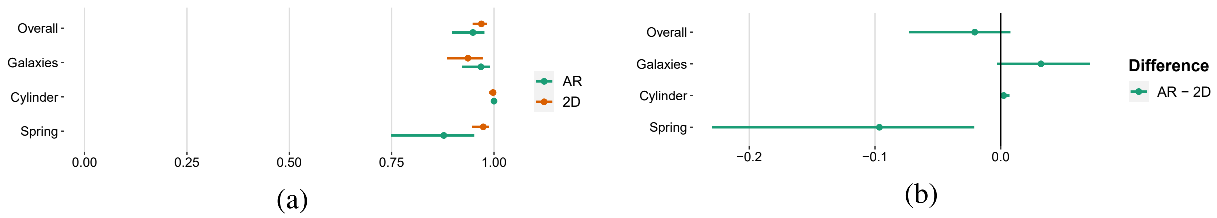


Figure 5.16: Raw data a and pair-wise within-subjects comparisons b of the proportion of constraint operations that contributed to the final results over unconstrained ones for each condition of the second experiment. Error bars: 95% BCIs.

better fine-tune their selections by removing and adding small additional number of dots in *AR* compared to *2D*. There is also weak evidence that *2D* was more accurate than *AR* for the Spring ($\approx +1.5\%$). When examining the ratio of useful constrained operations over total number of useful operations (Figure 5.16), it seems that *AR* invites more unconstrained operations compared to *2D*. This may be due to users having a better 3D understanding of the dataset in AR, making them understand how the Spring is created, and thus making them willing to follow the 3D shape. Moreover, the level of directness of RA (*AR*) over *2D* may invite more 6 DoF operations. However, unconstrained operations for the Spring are less accurate than constrained ones. This may explain the difference in accuracy. Participants performed unconstrained operations in a *Naïve Approach*: they placed the tablet near the Spring, drew a circle, and again placed the tablet near the Spring, before starting the extrusion.

P_0 tried twice to solve the Spring in unconstrained mode in *AR*. He then tried and succeed at solving the Spring using constrained operations. P_9 used constrained operations for the Spring in *2D* (his first condition) and used unconstrained operations in *AR*, where the 3D geometry was better understood. P_6 and P_8 used unconstrained operations in *AR* for their first trial of the Spring, but reverted to constrained operations for all the three other trials involving the Spring.

Overall, we thus cannot conclude that one technique is more accurate than the other. We thus cannot confirm **H7**. We also cannot conclude that the RF mapping in AR would be more accurate than the original implementation.

Focus We found strong evidence that users focus more on the External Screen in *AR* (i. e., the 3D AR view) than in *2D* (i. e., the 2D external screen) when placing the virtual tablet before starting an extrusion (Figure 5.17). Our participants also found that the AR view supported them more than the 2D view (Figure 5.18). The 3D AR view gives strong spatial cues about where

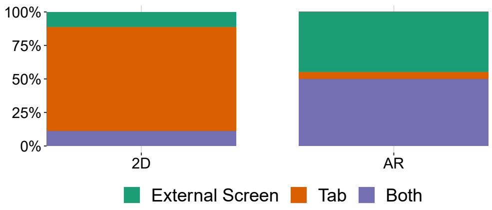
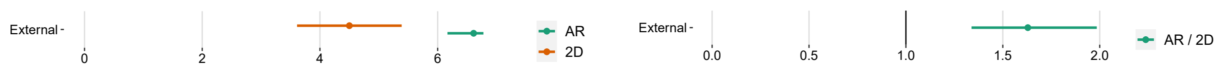


Figure 5.17: Proportion of which screen participants focused on while placing the tablet with respect to the visualization prior extrusions during the second experiment.



(a) (b)

Figure 5.18: Raw data a and pair-wise within-subjects comparisons b, on the support the external screen provided to the participants, 2nd experiment. Error bars: 95% BCIs.

the virtual tablet resides with respect to the dataset, compared to the 2D screen. In *2D*, however, the camera associated with the FI3D widgets on the tablet matches its virtual position. This helps users to orientate the dataset with respect to the virtual tablet. Once oriented correctly, P_{13} commented that he used the 2D external screen to adapt the depth position (with respect to the virtual tablet) of the dataset before starting an extrusion.

Extrusions in *AR*: Most participants (P_{0-6}, P_{8-14}, and P_{16}) primarily focused on the AR space when extruding. P_0 and P_{10} said that they looked at the tablet only to remove the inner part of the Cylinder which helped them with its near-clipping plane. Interestingly, P_8 used the AR view to check when to stop and to check if she was doing as expected, but used the tablet to know where to start using its near-clipping plane. P_1 said that she used the AR view to understand the relationship between virtual tablet and dataset. She commented that, compared to *2D*, she almost never looked at the tablet, except when she needed accuracy. P_5 used the tablet only to check if he had reached the end of the extrusion. P_{13} stated that the selection process was only a sliding movement when using constrained operations and did not look much at both screens. He, however, sometimes looked at the AR view to check the start and the end of the movement when required, e. g., with the inner part of the Cylinder. P_2, P_6, and P_9 focused on the tablet only during unconstrained operations (for the Spring), to check that their lasso was aligned with the orthographic view of the tablet. P_2 also commented on the conflict between the AR and tablet views when both are merged.

In addition, some participants used the AR view during the whole process or to refine their movements. P_2 looked at it to check when the extrusion should end. P_3, P_4, P_{11-12}, P_{14}, and P_{16-17} used the AR view during the whole process.

Finally, P_7 and P_{15} used the tablet almost all the time while using the AR view to check their selections. P_{15} commented that he paid special attention to the tablet at the beginning and the end of his selections. Overall, we found conflicting evidence regarding **H9** depending on participants' preferences and strategies and, therefore, cannot firmly conclude on it.

Extrusions in *2D*: In *2D*, however, most participants (P_{1-2}, P_{4-5}, P_{7-8}, P_{10-14}, P_{11}, and P_{16-17}) looked mostly at the tablet during extrusions. This might be due to the mapping being defined by the screen of the tablet (axis alignment between $\mathcal{O}$ and $\mathcal{M}$ when the tablet is the focus). P_{12} stated that he better understood the mapping in *2D* than in *AR* when he focused on the tablet. P_5, P_8, P_{12} used the external screen only to verify their extrusions, and P_{10} used it only when she was lost in the 3D space. In addition, we noted that she was not facing the virtual screen for some trials. Participants P_6 and P_{10} explicitly said that the perspective view from the viewpoint of the virtual tablet (left part of the external screen) was not useful at all. P_{11} never used the external screen during extrusions. P_{17}, surprisingly, used the external screen when she wanted to be accurate. She did not comment, however, what part of the external screen she was looking at.

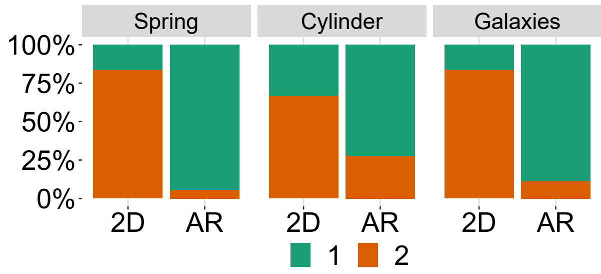


Figure 5.19: Proportion of ranking results for each condition of the second experiment. 1 is best. We use a value of 1 for both conditions when participants ranked them equally.

Some participants, however, used both screens about equally. P_3 used the external 2D screen to understand the 3D geometry and extrusion axis, while he used the tablet's near-clipping plane to check when to stop. He also commented that he had issues with understanding the 3D scene in projected *2D* compared to *AR*. P_{15} looked at the tablet only at the beginning and the end of the extrusions, and relied on the external 2D screen when making coarse movements. Finally, P_6 and P_9 primarily used the right part of the external screen for their extrusions. However, P_9 commented that he focused more on the tablet during the last trials when he understood how the near-clipping plane of the tablet supports him.

Preferences There is strong evidence that users prefer *AR* over *2D*. The size-effect, however, is weaker for the Cylinder dataset. As *2D* allowed participants to manipulate the 3D visualization with respect to the static virtual tablet, some first oriented the Cylinder to face the virtual tablet along its extrusion axis and then moved it along the depth axis. Using the constrained mode, they only needed to move the tablet forward to first add the red dots, and then to remove the inner part of the Cylinder, which is a simple strategy. This may explain the difference in the size-effect.

Discussion

While we did not find evidence that *AR* would improve the users' accuracy and speed, our results answer our main research question about the users' workload (**H6**). Users largely preferred and felt more comfortable using a 3D AR view over a 2D screen. By providing them with an AR-HMD, we shifted the users' attention from the tablet to the AR view. This is a fundamental difference compared to the original Tangible Brush where most users primarily relied on the view of the tablet for almost everything. Our evidence shows that the AR-HMD and the tablet can be used jointly with different available strategies. Some users used the near-clipping plane of the tablet to be accurate, and others relied on the AR-HMD for most interactions. We saw three strategies to place the virtual tablet with respect to the 3D dataset. The first one relies on the FI3D widgets. Designers may consider what parameters to give to the FI3D camera, which may depend on the users' tasks (e. g., placing the dataset in the 3D space or with respect to the virtual tablet). While Besançon et al.'s [38] participants seemed to have troubles with rotating their datasets using tangible rotations, we did not experience such an issue with the FI3D widgets. The second strategy is to place the virtual tablet in a 1:1 mapping using the "Position" mode. The last one is to remotely control the virtual tablet using the "Tangible" mode to apply the current mapping. More research is needed to understand which strategies work best and under what scenarios.

Finally, we call for more research to understand if AR views invite more full 6 DoF operations compared to projected 2D views. However, it seems that users would still mainly rely on constrained movements, even when they face selections that invite for unconstrained interaction (e. g., our Spring dataset).

5.6 Overall Discussion

We now discuss the implications of our findings for future research and the limitations of our experimental design.

5.6.1 Full 6-DoF Manipulations May Not Always Be Preferred

We found some evidence that users in a more immersive context might use full 6-DoF unconstrained manipulation more than in non-immersive contexts. This seems to echo previous research [59, 88] on the benefits of a 3D rendering environment to offer more intuitive 3D object manipulations. Nonetheless, we note that, overall, our participants tended to use constrained manipulations very frequently. Our results are therefore particularly interesting with respect to the discussion of integrating or separating DoF when performing 3D manipulations. Past research found conflicting evidence on what can be preferred: on the one hand, DoF integration would seem to lead to faster manipulations [282], but on the other hand DoF separation can lead to more precise and less frustrating manipulations [37, 183, 195, 250, 277, 282]. Moreover, past research focusing on docking tasks for touch displays show that users generally decompose the task into a translation task and a rotation task, instead of doing everything simultaneously [56, 184]. This fact may explain why some of our participants used the FI3D widgets instead of tangible operations to align the virtual tablet and the dataset correctly in the 2D condition, as the FI3D widgets facilitate such a decomposition.

While most research so far highlighted that DoF integration/separation depended on the input technology [161], our findings tend to suggest that the degree of immersion of the output technology also matters, as we saw that AR technologies seem to invite more DoF integration. These results are also particularly important when considering how interaction paradigms such as tangible input and mid-air gestures are praised for their "naturalness," mimicking real-life 3D interaction techniques [139]. Our results shed light on the fact that such unconstrained manipulations also might not be preferred to be constantly used, even in a fully immersive setup.

Yet, since this observation was not one of our primary research outcomes, further investigation on the impact of immersion on preferences for DoF integration is needed. Moreover, all those points remind us of discussions and our experiences that there may be a limit of our human biomechanical or mental abilities that leads to users avoiding to manipulate more than 4 DoFs simultaneously [56, 184].

5.6.2 Spatial Directness Affects Mental Models

Using Bruckner et al.'s model [57] (see Section 5.3), we showed that not all our AR implementations, even though they rely on the same set of input and output devices, have the same directness. Still, we did not perceive a correlation between directness and performance. Specifically, while all our AR implementations are more direct than the original Tangible Brush, we did not perceive that directness would strongly influence performance. However, the users' mental model, strategies, and perceptions are different. We especially see that relying on a 3D display might shift the user attention from what they manipulate (the tangible tablet) to the explicitly shown data, which gives scene overviews and spatial cues.

5.6.3 Tablet Control for AR-HMDs

There are several research projects that combine AR-HMDs used as output display with workstations or tablets used mostly as input devices but which possess also their own output spaces [50, 226, 284]; see also the survey at Chapter 2. Our work adds to this discussion that rendering AR holograms through the AR-HMD around the secondary input device engenders conflicts and makes users uncomfortable. For instance, in the Spring, P_2 of the second experiment said that he had hard time to follow the Spring using unconstrained operations as he could not check on his tablet if the lasso was aligned with the “circular shape” of the spring. Designers should thus consider the density of the used visual representations in AR. At the same time, however, they should also distinguish visualizations where users need large overviews to visualizations that need small overviews because spatially large representations combined with a relatively small mobile device would require constant focus switches. Except for the Galaxies, all our datasets were dense and require large overviews. A future study may shed light on different interaction designs for sparse datasets that require small overviews.

5.6.4 Limitations

Despite our careful design, our experiments have limitations. Our measurement of speed is a major one. Through our observations of the videos, it seems that participants spent a lot of time in their first trial with a dataset attempting to understand it and trying different strategies, before sticking to a suitable one. This may impact our results of the speed metric and of the participants' strategies. We share this limitation with previous experiments in the literature using similar experimental designs [38, 297]. Moreover, we did not associate a dialog box with the “Reset Selection” action which was sometimes pressed inadvertently, impacting the speed metric.

The accuracy of the tracking is another limitation. As the server merges the VICON tracking and the HoloLens one, fast head rotations may lead to a lag between both devices, which leads to synchronization conflicts on the server. Moreover, the HoloLens has difficulties to relocate itself during fast movements, i. e., the origin of its coordinate system is noisy, which strengthen this synchronization issue. Similar to raycasting paradigms, rotation noise can lead to strong displacements in, e. g., the orthographic projection of the tablet. However, this is true only when users rotate their head really unnaturally fast. While all our participants rated the tracking as being accurate, we could solve this issue with an integrated tracking on the tablet, such as what current VR-HMDs do with their controllers.

5.7 Conclusion

We studied three different mappings that tangible objects have when they are associated with AR-HMDs, in the specific case of volumetric selections based on the Tangible Brush technique [38]. Using Bruckner et al.'s model [57], we showed that our three adaptations of the Tangible Brush technique inside an AR environment using AR-HMD are more direct than the original implementation. Directness in AR, however, is not always synonymous to performance, preference, strategies, and users' workload as we showed in our first experiment (Section 5.5.2). Specifically, we found strong evidence that users generally prefer to rely on remote interactions over interacting inside the visualization where the tangible display conflicts with the AR view, conflicting with traditional thoughts that users prefer direct manipulations in 3D spaces. We thus compared one remote adaptation in AR to the original Tangible Brush setup in a second experiment (Section 5.5.3). We found strong evidence that AR views, for 3D volumetric selections, lower substantially the users' workload compared to the 2D condition while having similar accuracy and speed. Our results show also that users behave differently in AR environments compared to 2D ones, which researchers may want to further investigate in future work.

This project is the last one that concerns the “interaction” part of this thesis. Next, I discuss about the “visualization” part of the thesis with the concept of subjective views, or how to move multiple users to different,but related, AR worlds which are part of the AR multiverse.

6
CHAPTER

Personal Head-Mounted Displays and Subjective Views

This chapter extends an extended abstract I presented at IEEE Vis [267] with my supervisor Tobias Isenberg in 2020. This chapter focuses on giving initial design concepts for the "subjective views" I introduced in Chapter 1, i. e., the capacity of AR-HMDs to alter the visualization per user of a given 3D object anchored at a given 3D position in the users' space. Any use of "we" refers to my supervisor and myself. This project is part of the "visualization" part of this PhD.

As a 3D visualization often relies on multiple data layers, heavy filtering, and several transfer functions, the resulting divergence views (one per user) can lead to conflicts in the users' social space where users collaborate with each other but do not see the same set of features. Here we propose multiple designs to overcome this issue.

This chapter then targets **RQ5** that I stated as:

> **RQ5** What are the advantages and disadvantages of subjective views during the collaboration following the *modifier* and *appearance* dimensions for volumetric scientific datasets? Which interaction techniques and visualizations support them best with regard to the users' understanding, co-presence, and performance?

Given the circumstances of the COVID-19 pandemic, I was unable to perform any evaluation of this concept. Hence, I provide here only initial designs of the *subjective views* concept and how I implemented them in my system that I described in Chapter 3.

6.1 Introduction and Background

My system introduced in Chapter 3 relies on multiple AR-HMDs coupled with multi-touch tablets. As each user wears their own set of active displays, they can each have a personalized rendering of the same 3D scene. For example, past researchers [62,254] have investigated the use of private and public workspaces, where a private visualization is visible only by its owner, and the public visualizations are visible by all. Instead of hiding (i. e., making private) or showing (i. e., making public) 3D objects, I propose to adapt their visuals for each user depending on what those users are exploring.

The datasets on which I am focusing in this PhD, scientific datasets, often rely on spatial 3D data and usually contain multiple properties per data point. In meteorological datasets, for instance, each data point can embed temperature, pressure, and velocity properties, each useful to derive multiple meteorological phenomena (e. g., the strength of a possible hurricane). These multiple properties can be visualized by people with different scientific backgrounds and expertise, who will unlikely be interested in the same set of properties, but will still benefit from analyzing together the data within the same space. In this project we extend the concept

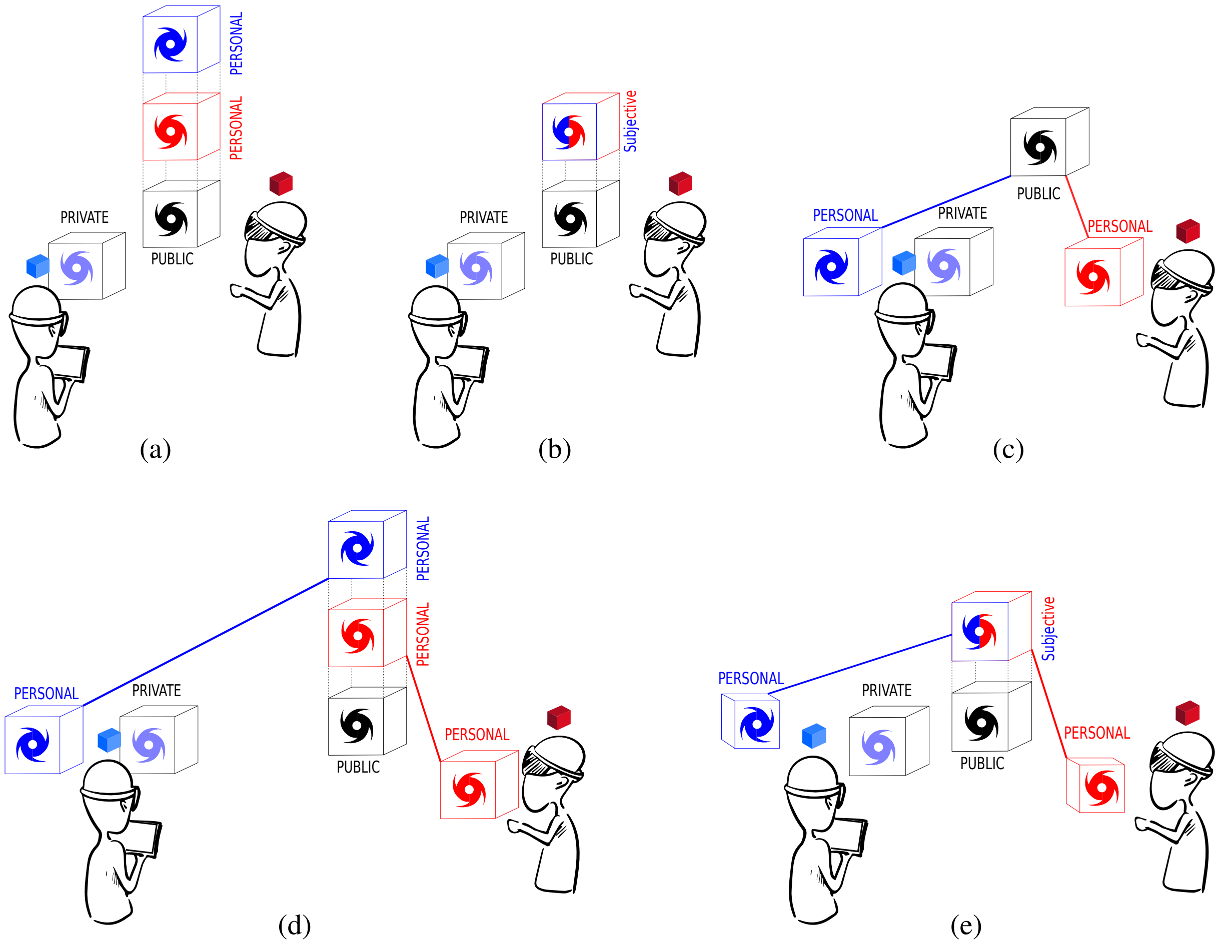


Figure 6.1: Set of the multiple designs to implement *subjective views*. The sketches represent two users and the spawned instances of a hurricane dataset. Figures (a) and (b) are based on a vertical stacking of subjective instances above the paired public 3D window. If the public instance moves, every subjective instance attached to it will move. The subjective instance in Figure (b) is perceived differently per user. Figure (c) proposes to spawn an instance near the user of interest. This instance would be static in the 3D space regardless of the linked public instance. Figures (d) and (e) propose to merge both modes.

of private and public spaces further by searching for the implications of rendering the same visual instance of a scientific dataset to all the users but with different visualization parameters, which would depend on their expertise and roles. We refer to these visualizations as *subjective views* [243], which also have been called *specialized views* [4] in the past.

Smith and Mariani [243] introduced the *appearance* and *modifier* dimensions for the *subjective views*. The first dimension considers the geometry of the 3D object while the second one considers global modifications such as transparency and wireframe render mode. Users can parameterize their datasets by selecting regions of interest (see Chapter 5) and by defining transfer functions. Such functions emphasize some parts of the dataset and not others by modifying the opacity (and color) of the dataset depending on the cell data, related to the *appearance* dimension. This processes in some cells being hidden and others being highly visible (i. e., opaque).

Compared to past work where virtual objects can be adjusted per user without introducing conflicts because these objects are just aesthetic [196] and encode a concept, the volumetric visualizations in this PhD are at the center of the discussion for the experts who work with them. If one user talks about a part of a volumetric visualization that is transparent for another user, a break happens between the social and the visualization spaces. Hence, significant different visualization parameters per user, resulting in significant differences on what information users have access to, can break the workspace coherency and may thus paralyze the discussion. However, one of the main benefits of AR-HMDs for collaborative work lies in merging the social and visualization spaces. Because this problem does not have a trivial solution, we propose some preliminary designs, concepts, and discussion about a possible implementation of *subjective views*, while keeping effective social communication.

6.2 Concepts and Designs

Based on Mahmood et al.'s study [178] and on Reipschläger et al.'s scenario [227], we assume that a typical collaborative session consists of three parts that repeat themselves. First, collaborators talk and explore the dataset together to get a rough idea of what they are exploring. Second, based on their roles, they explore the dataset on their own to parallelize the work. Finally, they gather their insights and discuss together.

We focus on AR-HMDs that allow co-workers to see each other and the visualizations without active focus transitions during all these listed collaborative phases. This work is then implementable in the system I described in Chapter 3. The visualizations of volumetric datasets are often not situated, so they often do not have a pre-defined location in the collaborative space for users to explore and visualize them. Moreover, with AR-HMDs, collaborators have access to a large virtual workspace. We use this aspect to allow users to invoke multiple floating instances of a given dataset, where each instance possesses its own visualization parameters.

However, while we can create multiple instances, the computing power of current commercially available AR-HMDs and general usability considerations limit the number of instances we can invoke, without making the system unresponsive and cluttering the workspace. While future AR-HMDs will likely be more powerful than what we currently have, we also raise the point that those AR-HMDs will likely always be less powerful than workstations, and that there is a general trade-off between visual qualities (i. e., the 3D resolution of the volumetric visualizations) and the number of visual instances users can invoke.

In our conceptual framework, we group multiple related instances into a *visualization context*. Essentially it consists of linking a *public instance* shared among all collaborators to other related instances that individuals may use, each using different user-defined visualization parameters. The *public instance* is used as a common frame of reference to not break the communication while allowing users to parallelize their work. While we propose to have multiple instances gathered together, each spatial instance might possess multiple set of visualization parameters to save space and computing power.

In my system, users use a hybrid system which combines multi-touch tablets with AR-HMDs, as discussed in Chapter 3. We then use an interaction design in which each collaborator uses a multi-touch tablet to control the *visualization context*, in addition to the AR-HMD for the 3D visualizations.

6.3 Visualization Context Instances Layout

The instances of the *visualization context* can be arranged in various ways, but we consider two *layouts* to be most promising, as we summarize in Figure 6.1. The first layout relies on *stacking* the visualization instances. This layout relates to the concept of small multiples, but links the visualizations based on the public reference. Having a wireframe cube framing the visualizations helps to relate spatially those "3D windows" together. When stacking two instances, their frames can easily be related to each other as two of the three dimensions remain the same (Figure 6.1a) and are linked to the public view. This is an additional benefit of the 3D windows concept, in addition to giving spatial cues (e. g., "look at the cellar"; see Section 4.5.5) and giving awareness cues (e. g., modifying the wireframe color to know which user manipulates the view; see Section 3.4.2). All these instances are normally public, i. e., all collaborators can see them the same way. We can also stack private views visible only to a single collaborator, but this makes the space use less efficient. An interesting aspect here is that, due to our use of AR-HMDs, we can show different visuals to the collaborators for the identical position in physical space. We thus can put n (n being the number of collaborators) theoretically private views at the same location in the virtual workspace, which turns them into *subjective views* due to their identical spatial reference frame (Figure 6.1b). This mode allows us first to save space in the collaborative workspace and may facilitate interesting collaborative exploration techniques. Yet it also poses questions on how to reveal the subjective view to collaborators during discussions. By allowing easy comparisons, we think that these modes would be suitable for discussion phases.

Our second layout relies on simple *linking* as we show in Figure 6.1c. This mode allows the collaborators to arrange the *visualization context* instances more freely, potentially with easier access, thus providing a more comfortable work environment. The downside of this mode is that the spatial relationship between different instances is less clear. But this may be less important during independent work periods, as egocentric objects are suitable for private work.

None of the two modes is thus able to support all work phases equally. We thus explored a combination that uses both the stacked and the linked layout as we show in Figures 6.1d and e: the stacked elements that provide collaborators with an easy way of understanding their spatial context are replicated using linking to be more easily accessible. The main drawback of this arrangement is that it requires us to render a lot more instances.

As additional points, these designs can be altered by adding visibility rights to all the instances, e. g., user A's visibility over user B's personal instances might change and depend surely on which phase of the collaboration users are in. Moreover, what differs from the merged layouts (Figures 6.1b and e) and their exploded counterparts (Figures 6.1a and d) is the virtual merging of personal instances. A system might then be able to switch between a subjective and an exploded design, which might allow other behaviors during the collaborative process.

6.4 How Many Visualizations? The Limits of Computing Power

My PhD relies on volumetric visualizations which require a lot of computing power due to the raymarching algorithms. While AR-HMDs are gaining more computing power through time about the same rate as multi-touch tablets do, users might tend to display later visualizations with higher resolutions, i. e., resolutions desktop machines currently have, which consume again more computing power and will again reach the limits of the hardware. At the time I am writing this thesis, my system can handle the display of four simultaneous (i. e., visible in the user's FoV) visualizations about $25 \cdot 25 \cdot 25\text{cm}^3$ in the physical space for which datasets comprise approximately $370 \times 280 \times 50 = 5,180,000$ cells before being unresponsive. This reaches the limit of the *multi-stack*[1] visualizations for two users (one public visualization plus two subjective instances). Designs combining stacked and linked visualizations might then be inappropriate for AR-HMDs which rely on even more visualization instances. Nonetheless, the shared stacked visualization may not need to be rendered in full fidelity during independent work phases, and the linked copies are not needed in full fidelity during discussions. So an effective switching mechanism for the current work focus would be ideal.

[1] We use the term *multi-stack* to mean that we explode (see the concept of explosion diagrams [256]), along one axis, a subjective instance that possesses n sets of visualization parameters into n stacked visual instances, each with its own set of visualization parameters.

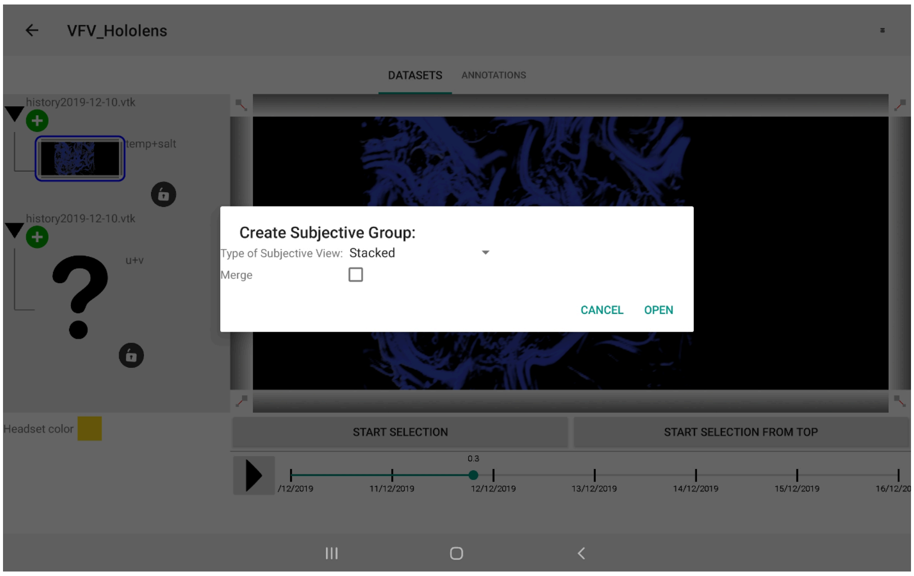


Figure 6.2: Interface of the tablet to create a *visualization context*. Users can set the type of the context and whether stacked visualizations are merged or not. The selected visual instance will be used as the public visualization of the *visualization context*.

(a) (b)

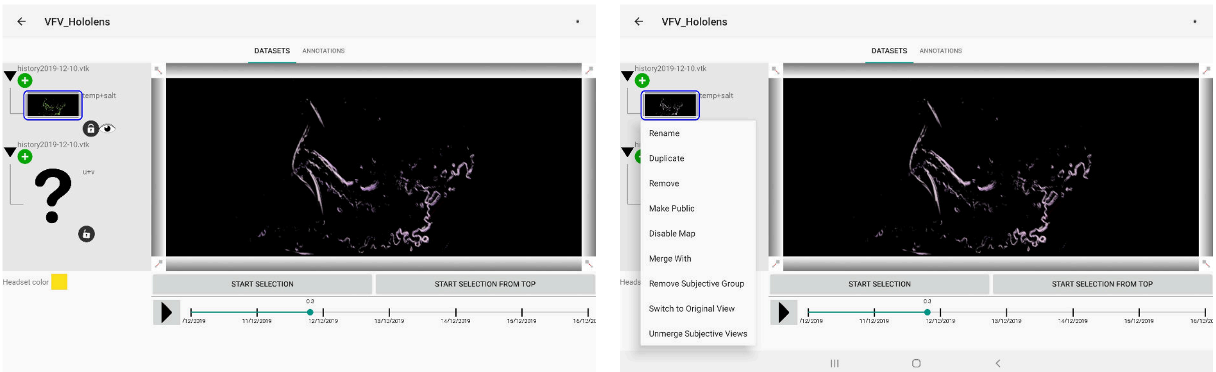
Figure 6.3: Implementation of the subjective views as defined by Figure 6.1b viewed through the AR-HMDs. The highest instance is viewed differently by user A (Figure 6.3a) and by user B (Figure 6.3b). In the implementation, the system creates n personal instances which have the same position, orientation, and scaling. A given user sees only the visual representation he or she owns to not have multiple visible visual representations overlapping each other. I display here the temperature field of the “Cabo Verde” dataset HZG provided me. In the screenshots, some eddies are visually extracted.

Despite these limitations, I would like to emphasize that these five design considerations might be applied for VR setups which rely more on heavy desktop machines compared to AR-HMDs. However, the usual trade-off between data resolutions and number of instances will likely always be present.

6.5 System

The specific implementation of the subjective views conceptual framework is based on the basic system design I described in Chapter 3. Next, I explain what is specific to the implementation of this framework in the first-person voice as I implemented the system alone after the publication of the extended abstract.

Figure 6.4: Screenshot from one HoloLens of the implementation of the *visualization context* with the stacked+linked design (Figure 6.1d). All the stacked instances are visible to all the collaborators, but the details of the private linked instances are only visible by their owners. It is to note that this configuration is reaching the limits of the current hardware (i. e., displaying simultaneously four visual instances).



(a) (b)

Figure 6.5: Screenshots of the tablet interface when users manipulate their personal views (a). An "eye" icon shows that this view is private and personal. A long-press shows a sub-menu where users can switch back to the public view and remove the subjective view group (b).

I implemented the five designs shown in Figure 6.1. The user starts the creation of *visualization context* by performing a long-press on a public visualization instance, which would serve as the frame of reference, on the multi-touch tablet. A menu then appears where the user can click on "Create Subjective Views" (Figure 3.5) to create the *visualization context*. A prompt-window then asks for specific parameters (linked vs. stacked only vs. both; Merging of the personal views or not. See Figure 6.2). The corresponding stacked and linked visualization instances are then created. The links are anchored on one corner of each related cube. The algorithm selects the pair of corners that is the shortest. Only the personal visualizations of those asking for it are created to not clutter the workspace more than needed, as users can be separated into smaller groups, and thus the *visualization context* might only be useful to some users. By having linked or stacked visualizations, one should be aware that a *visualization context* is created. More research is however needed to understand which awareness cues are useful. Figures 6.3 and 6.4 show screenshots of the implementation from the AR-HMDs point of view.

For a given context, I render in the list of the created instances (left part of Figure 6.5) only the "current" visualization of the *visualization contexts* which the given user manipulates, and

I hide the other related instances. If the user again long-presses on a 3D visualization, he or she can "switch to [the] public" and "switch to [the] personal" visualization instance. If the personal object is selected and the *visualization context* layout possesses linked visualizations (Figures 6.1c, d, and e), the FI3D widgets manipulate the geometric (position, rotation, scaling) parameters of this linked visualization.

Each visualization is a specific 3D object with a pair of IDs (i. e., which dataset this visualization is instance of, and which instance ID is this visualization for this given dataset). Instead of having one object with multiple visualization parameters, I create as many visualization instances as necessary (at a maximum of two for each user) and place them accordingly to the chosen layout. It means that the visually *merged* visualization object, for *merged* layout (Figures 6.1b and e), is composed of n multiple objects in the scene graph, n being the number of participants having a *personal* visualization. For those objects, only the personal visualization is visible for a given user. This was simpler to implement in the on-going project I developed throughout this PhD. It also allows me to combine multiple other layouts. For example, if one wants also to display multiple timesteps per visualization parameters (i. e., performing a small-multiple of timesteps), the number of possibilities would be hard to manage in the software. However, by relying on an equivalent of layout systems such as the ones 2D Graphical User Interfaces use, we can encapsulate and merge multiple layouts together seamlessly.

This also results in design considerations for "Post-WIMP" [164, 275][2] frameworks. "Post-WIMP" designers might inspire themselves to what current GUI frameworks are doing, as the design issues might be similar in both cases.

6.6 Scenarios

Next, we give examples to better illustrate the scenarios in which those concepts of *visualization contexts* and *subjective views* potentially could be useful. Let us consider three scenarios where two users (U_a and U_b) work together.

6.6.1 Relating multiple properties

Section 3.2 shows that scientists are investigating multiple properties. Section 3.6 also shows that some properties might be better than others to find suitable patterns (e. g., eddies). Now let us consider a situation where users see an eddy with the salinity property and want to check the same behavior within the temperature and the velocity properties.

In such a scenario, U_a would spawn a *visualization context* associated with the salinity property (called d_s) which would serve as a frame of reference. He would use a merged–vertical layout (see Figure 6.1b). U_a would then possess now a subjective instance (called d_v for the velocity property). U_b would then create her own subjective instance (called d_t for the temperature property). Having subjective instances will allow U_a and U_b to relate their subjective instances to the original dataset (d_s), as all visual instances (with the two subjective ones sharing the same geometry) would be vertically-aligned with nothing between the original and the subjective instances. This relationship would help U_a and U_b to relate their resulting visualizations with a baseline, without the need to duplicate d_s. Working with the same d_s would, moreover, allow both users to still continue to collaborate together and share ideas while they are exploring their subjective instances.

[2] Post-WIMP interfaces are all human computer interfaces that go beyond the classical "Windows, Icons, Menus, Pointers" metaphors that most GUIs rely on. Those interfaces are generally based on a subset of pen/touch interfaces, natural language interfaces, and mid-air gestures and other immersive interfaces (e. g., gaze); see also Lee et al.'s monograph [164] on Post-WIMP interfaces for the field of information visualization.

Now let us consider that U_a found in d_v a match in the eddy behavior, but not U_b within d_t. Since both users know that they are in a subjective mode, it might be wise to now *multi-stack* the view to have the three visualizations stacked together (see Figure 6.1a) to better discuss, together, the case of the temperature and see if the missing behavior is expected.

We can also imagine a scenario where pointing to missing features might be interesting, as both users are expected to explore different sets of properties together. For instance, U_a can say (and use non-verbal communication cues) that he has found the eddy in the mid-range velocity scalar (i. e., the velocity magnitude) values, which might correspond to, e. g., mid-range temperature (instead of extrema) scalar values. With this information, U_b might check these ranges to see if she is able to find the eddy or not, while U_a continues to refine his visualization.

6.6.2 Searching Other Patterns

The above use case concerned two users collaborating jointly. Now let us consider a case where searching for behaviors would take more time. In such a scenario, users might want to work separately.

Let us consider that two visualizations are in the public space, which contain multiple interesting insights (e. g., multiple eddies that are not available in a singular visualization). Using the design we described in Figure 6.1e, users would first search for additional eddies or visually refine in other properties of the dataset the 3D size of the eddies both users were able to capture. U_a might want to work alone in his private instance, while U_b would at one point move towards the public workspace to relate her current subjective instance with the instance serving as reference, but also with the second instance put in the public workspace.

At the end, both users would join and might want to *multi-stack* the view (see Figure 6.1d) to discuss together about the new information they have gathered.

6.6.3 Tight Link Between Virtual And Real World

While those two scenarios we described above concern the oceanographic domain area, *visualization contexts* can also be applied to other expertise. Let us consider a use case where the physical and the virtual worlds are tightly linked together.

For instance, we can imagine a conference room scenario in which a company, which works on some project (e. g., river dam, building a new power plant), uses a physical model of the site as a reference frame visible to all. Then the subjective views are placed on top of this physical reference frame, and everyone can easily relate between what they virtually see (e. g., electric schema, pipe schema, foundation schema, radiation simulations for power plants) and the physical model that serves as a support. This physical model might fulfill the same purpose as the shared reference all our *visualization context* designs rely on. In such a case, designers might consider the parameterization of the shared reference used by the *visualization context*, e. g., its visibility status. Additionally, such a support might be useful to add physical annotations onto, or for other non-domain experts to understand what is going on.

Those three scenarios show how designs that combine public, personal, and subjective views might support users in their exploratory tasks. More extensive user experiments might confirm or not those possible scenarios and might yield other scenarios and hypotheses that we did not think about.

6.7 Discussion and Conclusion

The subjective view framework and the resulting workspace designs use the capacity of the AR-HMDs to merge the social and the visualization spaces, while displaying custom contents per user depending on their roles. While only the designs represented in Figures 6.1b and e really use the basic definition of *subjective views*, i. e., a spatial object that is encoded differently per user, we argue that this strict definition suggests that the common frame of reference is the only shared property (i. e., position, orientation, and sometimes scaling) of a visualization. In our *visualization context* concept we use thus a common, public visualization as a shared frame of reference, and rely on stacking or linking to relate other personal, private, or *subjective views* to it. Moreover, while the other designs do not directly use the basic subjective property, it is worth noting that a personal instance belonging to a particular user does not necessarily need to be rendered to all the other collaborators. Our proposed designs can then be altered by changing the visibility rights of the instances of the *visualization context*.

Evaluations are required to examine how users would use and interact with the *visualization context* and the embedded visualizations, in particular how they would use and share insights from subjective views. By using tablets, multiple interactions such as filtering and spatial interactions can be done remotely. These need to be communicated well as they do not give deictic cues to the other collaborators. At the same time we need to avoid cluttering the common workspace with proxies or disturbing others with direct manipulations. Moreover, as it is true for wall-sized displays, users cannot both have a clear overview of all the stacked instances and have details about a specific instance, especially with AR-HMDs for which the resolution of the screen and the data (due to computer power) cannot match what other VR systems and large displays offer.

We did not study those functionalities because of the COVID-19 pandemic which hindered a lot users' experiments. However, we discussed and showed the functionalities to my colleagues at HZG in August 2021, two months before I am handing this thesis over. With two users collaborating with each other, it seems that vertical stackings are less effective than horizontal ones because users cannot see clearly the highest visualization, despite having the possibility to rotate all the stacked instances simultaneously. I do not have any other feedback for the moment. This research needs to be continued by myself or by another researcher.

CHAPTER 7

DISCUSSION AND CONCLUSION

This thesis discussed the use of AR-HMDs for collaborative 3D data explorations. It first surveyed the junction of the CSCW and Augmented Reality (Chapter 2) research areas, then explored two fundamental interaction techniques (Chapter 4 and Chapter 5), in addition to give primary designs of subjective views (Chapter 6). All those projects are part of a proof-of-concept application that allows collaborative exploration of 3D oceanographic datasets (Chapter 3).

In this chapter, I summarize my contributions and aim at giving general answers to my research questions. I close by posing more research questions and discussing future work.

7.1 Multi-touch tablets plus AR-HMDs (RQ1)

I had stated my first research question (**RQ1**) as follows in Chapter 1:

> **RQ1** Compared to standard existing approaches, what are the advantages and disadvantages of the combined use of multi-touch tablets and AR-HMDs in collaborative 3D data exploration scenarios?

In the software I created throughout this PhD, I mostly use the multi-touch tablets as input devices and the AR-HMDs as output devices. While the two research projects I described in Chapter 4 and Chapter 5 rely on mid-air interactions, a lot more input capabilities described in Chapter 3 happen on the touch screens of the multi-touch tablets. This setup is different to past ones which combined AR-HMDs with workstations [50, 281, 284]. I hypothesize that such an environment is suitable for collaborative work as the multi-touch tablets do not hinder users' mobility, compared to workstations, while giving usual 2D Graphical User Interfaces (GUI) to users. Those multi-touch tablets also can serve as tangible devices (Chapter 5) without active focus transitions between the devices.

However, this comes at a cost of efficiency with respect to the user interfaces. Indeed, users are used to and are more performant with workstations than with multi-touch tablets as the former possess the usual mouse+keyboard interface which is useful, for example, for scripting (e. g., through MatLab or Python frameworks) capabilities that are needed to effectively filter datasets.

But I am not considering here the computing power of the multi-touch tablets to be a limitation, compared to what workstations have. Indeed, the HoloLens 2^{nd} gen. I am using possesses similar computing power than the multi-touch tablets I am associating them with. As such, the computing power to render complex 3D visualizations is as limited for the AR-HMDs as it is for the multi-touch tablets. Moreover, in my system (Chapter 3), the AR-HMDs render multiple 3D visualizations in stereoscopy simultaneously, while the multi-touch tablets show only one visualization at a time without stereoscopic rendering. All the complex computations, such as 3D volumetric selections (Chapter 5) happen finally on the server and not on the multi-touch tablet, allowing the latest to focus solely on the user interface. As such, it is the AR-HMDs that limit the resolution and the number of 3D visualizations users can look at simultaneously and not the multi-touch tablets nor the remote server.

Next, I am explaining the possible benefits of separating the input and output devices.

7.1.1 One Input Device and One Output Device

By looking back to my collaborative immersive analytics prototype tested with two persons (Chapter 3), it seems that separating the interaction modalities and the visualizations in two separated classes of devices brings new kind of human interactions. Indeed, during this test, one user was not trained with the multi-touch tablet and thus could not use it at the beginning of the session. While I was training this user (U_b) to use the tablet, his collaborator (U_a) was manipulating the AR environment. As such consequence, U_b put down the tablet and forgot about it. Instead, he was following the discussion and communicating with his collaborator using social interaction inside the AR space (e. g., pointing toward objects and features).

We can thus imagine scenarios where one person handles the *public* environment while multiple other persons follow the discussion, without them understanding how things work behind the scene. This can be useful when the majority is spectator of a minority's actions (which corresponds to the Isenberg et al. [136]'s first categorization of visual applications based on users' engagement), e. g., for teaching purposes (classes, museum) and for presentations.

While most of the interactions happen on the multi-touch tablets, I also used in this PhD the AR-HMDs as input devices. In Chapter 4, my co-authors and I tested multiple 3D point specification techniques which start on the multi-touch tablets but rely, once started, on the AR-HMDs using hand gestures. Handling simultaneously two input devices might lead to possible conflicts as I discuss next.

7.1.2 Conflicting Interaction Modalities?

By looking back at my first research project that focuses on specifying 3D points (Chapter 4), multi-touch tablets do not hinder the input modalities (mid-air gestures, voice commands, and eye-gaze) of AR-HMDs when those can rely on no-to-one hand.

Moreover, AR-HMDs do not hinder the capabilities of the multi-touch tablets: the tablets are *usually* visible and their touch input can be separated from the mid-air gestures of the AR-HMDs. Moreover, the tablets can be used as tangible devices inside the AR space (Chapter 5), which is compatible with the paradigm of using mainly mid-air gestures for AR-HMDs.

However, there are some limitations. Interactions based on two hands are not possible with the AR-HMDs because users are expected to hold their tablets in at least one hand. Moreover, the screen of a tablet is only visible when it is not pointing towards, or is not inside, an AR 3D visualization. Finally, mid-air gestures based on hands can be misinterpreted when the user manipulates their tablet, e. g., some participants saw the startup menu of the HoloLens 2^{nd} gen. (which requires the user to tap their wrist with their first finger) showing up unexpectedly in the users' experiment studying our implementation of Tangible Brush (Chapter 5).

Let us consider the "point specification" interaction. While specifying a point, a user should not manipulate the tablet until the interaction is closed, as the AR-HMD can misunderstand gestures that it should not, e. g., the "confirmation" gesture (which requires a "Tap" gesture) can be recognized when users manipulate the touch screen of their tablet. We can thus derive the design guideline that interactions that require simultaneously touch input and mid-air gestures input based on hands should be avoided.

7.1.3 New Research Question: Size of 2D Personal and Public Screens

One research project my co-authors and I reviewed in Chapter 2 combined AR-HMDs, not with multi-touch tablets, but with large interactive displays [227]. Such a large display gives a large public 2D canvas that collaborators can use as a support for their analysis but also to communicate with other collaborators.

In my system, the multi-touch tablets are private by design, as users are expected to hold them in their hands, with the screens visible only by their owners. One can thus imagine a system that combines (1) AR-HMDs for 3D visualizations, (2) a large display for a public canvas, and (3) multi-touch tablets for private canvases and 2D user interfaces. Compared to Reipschläger et al.'s work [227], the 2D Graphical User Interfaces (e. g., buttons) can then be exported to the multi-touch tablets instead of being on the large display. The large display then can fully act as a public canvas for 2D information and as a support for the 3D visualizations by, e. g., rendering 2D annotations, 2D maps, and satellite images in an oceanographic use case such as the one I described in Chapter 3. Such a combination might be interested to pursue in future work.

7.2 Specifying Points: Solo Against Collaboration (RQ2)

In Chapter 4, my co-authors and I studied 3D point specifications. This concerns my second research question (**RQ2**) stated as:

> **RQ2** To specify points in a volumetric dataset, what are the implications of the different available metaphors on the users' understanding, co-presence, and performance, in a collaborative AR environment?

It seems that direct manipulations best support users' performance and co-presence. However, for what concerns users' communication, we saw in our first experiment that having a ray supports users in their communication cues (e. g., adding meanings to the "forward" cue). This echoes to multiple research projects [149, 217, 259] focusing on remote collaboration that show that augmenting pointing gestures enhance strongly the communication between users. It might be interesting to search, in later research, for the effects of pointing augmentation cues in co-located settings. Still, raycasting techniques (e. g., our AR Go-Go technique) seem to not be accurate enough to specify points and should only be used as a communication support.

Adding a second frame of reference for remote interaction—our World-In-Miniature (WIM) interaction—, however, can be preferred for parallel tasks. While we did not study such a scenario, direct manipulations make one move inside the public domain which might be used by multiple other persons simultaneously. While the WIM is accurate, with respect to its size, it should not be used as a technique by default as the creation of two frames of reference might disturb co-workers when they are working together.

In sum then, a system may need to propose multiple pointing interaction techniques based on what users intend to do: specifying a 3D position known in advance in common tasks or in private objects (Manual); searching for a suitable position inside an area where discussion is mandatory because the position is yet unknown prior to start the search (AR Go-Go); or work in parallel (WIM). Still, adding multiple possibilities (here three different modalities) for the same high-level task (specifying 3D points) might be confusing for users. Designers should then consider the trade-off between adding functionalities and their usefulness.

7.3 Volumetric Selection: The Impact of Directness (RQ3)

In Chapter 5, my co-authors and I studied volumetric selections using a tangible multi-touch tablet for solo-tasks. This concerns my third research question (**RQ3**) stated as:

> **RQ3** To specify regions of interest, what are the implications of a tangible multi-touch tablet where its 3D position has meanings in the AR space compared to the original Tangible Brush which decouples the input and output spaces?

This question is complex as the two experiments described in Chapter 5 gathered many insights which I am summarizing below.

7.3.1 Direct Interaction not Preferred in AR

In Chapter 5, my co-authors and I found three possible mappings (one direct and two remote mappings) to manipulate remotely a 2D plane using a tangible device. In this project, users needed to draw 2D shapes on their multi-touch tablets and to extrude those shapes by moving their tablets around using the different mappings we provided.

We showed in this project that users did not prefer direct interactions for this task for two reasons. The first one concerns the lack of overview when users are in front of the 3D visualizations with respect to the AR space. The second one is the visibility conflict: the tablet is no more visible once it points toward or is inside an AR visualization. However, we note that most of the datasets we used for these experiments had (1) large areas to be selected, and (2) contained dense visualizations. Using a World-In-Miniature metaphor might give large overviews to users in the most direct metaphor and change the results. More research is then required for smaller selections with sparse visualizations and smaller required overviews, for which results might be different.

The main difference between the two remote mappings concerns the alignment with respect to the output device. The first mapping (RA) uses the same orientation between the physical tablet and the virtual tablet with respect to the AR space. The second one (RF) aligns, in orientation, the physical tablet and the virtual tablet with respect to the screen of the tablet. While we thus expected users to focus on the respective screen for which the mapping is aligned with, we showed that users mainly focused on the 3D AR space. Moreover, we saw that some participants took benefits of the two separated coordinate systems created by RF when they focused on the AR space. They mostly put themselves in more suitable positions to have better depth perceptions and, by extension, to be more accurate.

Based on the second experiment, it seems that the dimensionality between the input modalities (here the tablet) and the output modalities (AR-HMDs vs. external 2D screens) have a strong impact on which screen users focus on as I explain next.

7.3.2 The Impact of Similar Dimensions Between the Output and Input Devices

In this research project, users move their tablets in their 3D physical space. In the *AR* condition, users see in 3D their selections inside their environment. However, in the *2D* condition, they can only see their selections on 2D screens (using either the tablet or the external 2D screen).

We saw in this project that participants strongly focused on the AR 3D view in the *AR* condition compared to the *2D* condition where they strongly focused on their multi-touch tablet (which is what they manipulated and where the mapping “Relative-Full” was defined for). Moreover, users, to draw accurately their 2D lassos, are expected to focus on their tablet which is a 2D screen. Having an n-dimensional input modality seems then to invite users to focus on an output device with a similar dimension (2D tablet for the 2D lasso, 3D AR space for the 3D extrusions).

One may say that directness influences which screen users focus on. While this might be true at some extent, Table 5.1 shows that, for the original Tangible Brush technique, focusing on the 2D screen leads to higher directness than focusing on the multi-touch tablet. Moreover, the level of directness is similar for users focusing on the 2D external screen (in the original implementation; see Figure 5.17) and focusing on the AR-HMD in the RF mapping (in the AR condition, see Figure 5.9). But, in the former case, we have strong evidence that users focused on the multi-touch tablet (the 3D interactive device), while, for the AR-HMD, users focused mainly on the AR-HMDs (the 3D output display). The difference in dimensions between the input and output devices might lead to bigger shifts in focus compared to the difference in the level of directness. However, we had not envisioned this relationship as part of our hypotheses. More research is needed to validate this primary result, to understand the implications of such behavior, and to find when and where this statement no longer stands.

7.4 Remote vs. Direct (RQ4)

RQ2 and **RQ3** discuss about the direct nature of interactions (specifying points and regions of interest). This leads to my fourth research question (**RQ4**) that I stated as:

> **RQ4** As a side effect (post-hoc research question) of the interaction modalities I studied, what are the main benefits and limitations of direct interaction mappings compared to remote ones for one-user and collaborative environments?

7.4.1 A Matter of Object Size for Solo Tasks?

The most important difference between those two interactions concerns their "volume space." Theoretically, a 3D point has no volume, which is not the case of Regions of Interest (ROIs).

This might explain why users were more performant (in a one-user environment) with direct interactions for specifying points, compared to ROIs where users were more performant with remote interactions. Indeed, to specify a point, a user might need less overviews than to define a volume. However, this is true only if the user knows what point to specify. If a given user needs to browse a broader region before specifying a point, remote interactions can either be used for the browsing step of the interaction, or for the whole interaction.

I hypothesize thus that remote interaction might be preferred for interactions that require a large overview and regions, and direct interaction might be preferred when users focus only in small areas. Investigating this hypothesis might result in possible insights about the users' preference between egocentric and exocentric interactions. The specification of points should then be studied again, in both solo tasks, but also in collaborative tasks, where users need a global overview before specifying a point close to their target. Results might be different than what I answered for **RQ2** originally. I hypothesize that users might prefer the "World-In-Minature" interaction for its remote and exocentric nature if those users need large overviews.

7.4.2 And if We Consider Collaborative Tasks?

Using Bruckner et al.'s model [57] (see Figure 1.4), I said that Class 1 interactions (the most direct form of interactions) can theoretically merge the input ($\mathcal{M}$ and $\mathcal{I}$), output ($\mathcal{O}$ and $\mathcal{V}$), and collaborative spaces inside the users' physical space. These collaborators are expected to better understand another one's doings without much effort.

However, we saw that users preferred remote interactions over the direct one for large volumetric selections with dense datasets (Chapter 5), compared to the specification of 3D points which required small overviews (Chapter 4) where users preferred the direct interaction over the remote ones. Having an interaction where remote interactions perform better than direct ones might change the collaborative metrics for tasks where users work together (e. g., co-presence and mutual understanding) and for parallel independent tasks. As such, more experiments are required to study all those different use cases (large overviews/small overviews; dense datasets/sparse datasets; remote interactions being most performant/direct interactions being most performant) for both tasks requiring users to work together and for parallel tasks.

7.5 The AR Multiverse of Subjective Views (RQ5)

My latest research question is stated as:

RQ5 What are the advantages and disadvantages of subjective views during the collaboration following the *modifier* and *appearance* dimensions for volumetric scientific datasets? Which interaction techniques and visualizations support them best with regard to the users' understanding, co-presence, and performance?

I was not able to fully answer this research question. However, in Chapter 6, I gave multiple design considerations for subjective views of volumetric scientific datasets where they are mostly filtered by altering significantly the transparency of data points. All those designs have in common that they rely on a public object which is rendered the same to all users. I hypothesize that this gives users a common ground which would support the collaborators' discussion. It remains yet to study those designs in a real use-case scenario.

Specifically, I would be interested in studying what visual instances users would use (e. g., their own instances or the public one) for designs involving real *subjective views* as the literature define them (Figure 6.1b and Figure 6.1e), or designs that multi-stack the subjective views (Figure 6.1a and Figure 6.1d). I would also be interested about how much users would focus on their private views for designs that involve private–linked instances (Figure 6.1c, Figure 6.1d, and Figure 6.1e) during the whole "exploratory" steps, and whether users would feel more comfortable with those private instances. Finally, usual questions about co-presence, mutual understanding, and more importantly performance remain to explore.

7.6 Conclusion

In this thesis, I first have reviewed the junction between AR and CSCW. With my co-authors, we saw that 2D and 3D data exploration were not extensively studied in AR compared to the other immersive platforms (e. g., VR). I thus first framed what kind of environment would suite collaborative 3D volumetric data exploration using AR as a support. I arrived at the conclusion that the combination of multiple multi-touch tablets and AR-HMDs would fulfill such a scenario (**RQ1**), even if their combination is not yet fully understood (e. g., What are the implications of this environment being combined with a public large interactive display? How to manage the conflicts of using the AR-HMDs' and the tangible multi-touch tablets' input capabilities simultaneously?)

Then, because of the lack of fundamental knowledge for what concerns exploratory tools, I first studied the interaction of specifying 3D points (**RQ2**) for one-user and collaborative environments. I then studied the combination of a tangible multi-touch tablet with an AR-HMD to select volumes inside a given 3D visualization (**RQ3**). Those two research projects diverge from what users preferred (remote vs. direct interactions) and how much overview they required. These yield numerous research questions (**RQ4**).

I then focused on one of the particularities of AR-HMDs which concerns their faculty to alter an object's rendering and visual parameters per user (**RQ5**), and I finalized a proof-of-concept to visualize 3D VTK_STRUCTURED_POINTS (i. e., a data format where each cell is part of a 3D structured grid). Because of the COVID-19 pandemic, however, I could not study those functionalities and this proof-of-concept together. I can therefore only give primary results based on a pilot study and the design processes.

While I cannot give a general answer for what concerns the benefits and limitations of AR-HMDs for collaborative 3D data exploration, I can, however, state some of them. The benefits of AR-HMDs I found concern the merging of the interactive, output, and collaborative spaces inside the users' physical environment. Users can use their physical environment for guiding cues (e. g., "towards me", "towards the chair", etc.) and structure their environment the way they want, e. g., "this part is my working area and that part is yours." Moreover, 3D 6-DoFs interactions seem more appealing to users that use AR-HMDs compared to 2D non-stereoscopic environments. The AR-HMDs also seem to lower substantially the users' mental demand compared to 2D environments for 3D interactions, at least for what concerns 3D volumetric selections. Moreover, practitioners requiring powerful software such as those we found only on workstations can still use their tools and AR-HMDs simultaneously to get the best of both worlds [281, 284]. For scenarios requiring less powerful software, AR-HMDs can be combined with multi-touch tablets (or similar, e. g., smartphones) to (1) have access to 3D visualizations through AR-HMDs, while (2) to have access to usual graphical user interfaces through the multi-touch tablets, and (3) to give additional tangible interactions for which physical objects can improve the users' performance.

As for the limitations, I can name the AR-HMDs' computing power compared to workstations and that optical see-through AR-HMDs cannot render black colors. While video see-through can render black colors, they have other limitations such as additional latencies due to the cameras capturing the real world, and the lower resolution of the screens and cameras of the headsets compared to the almost "infinite" resolution of the users' eyes in the real world. These limitations are inherent to the technology and would always exist compared to (1) workstations which are expected to be more powerful, and (2) have better screen quality.

These benefits and limitations together make AR-HMDs more practical than workstations (if used separately) only for collaborative work if the purpose is to share basic knowledge and where users do not need all their usual exploratory tools (e. g., scripting).

As a side note, I would like to emphasize that this thesis demonstrated an engineering work for AR 3D visualizations. I showed that such visualizations are now possible with the current technology. Moreover, current AR-HMDs embed so many sensors and computing power that, even if the provided API does not answer a developer's needs or purposes, he or she can still use the sensors' data to implement his or her own algorithm just as I did with OpenCV (Chapter 4) to track the users' hands. While Unity is probably the most current used development platform for the Microsoft's HoloLens, one can still use C# or C++ and use the Microsoft's DirectX API to better optimize the rendering.

I do hope that this thesis would give a general picture to researchers and industrials interested at understanding in which scenarios would AR-HMDs support users for 3D data exploratory tasks. For such an environment to be adopted inside a workspace, however, a dedicated server and a well-designed network infrastructure seem necessary. Moreover, a faster way to load datasets inside all the devices (server, multi-touch tablets, and AR-HMDs) is also required for users that work primarily with workstations to, e. g., make simulations. More engineering work and thoughts are then necessary and would extend this thesis well to have a working environment that can be adopted by practitioners.

Bibliography

[1] ACM, "Acm conference on computer supported cooperative work," https://dl.acm.org/conference/cscw, accessed: December 16, 2021.

[2] M. J. Adams, Y. J. Tenney, and R. W. Pew, "Situation awareness and the cognitive management of complex systems," *Human Factors*, vol. 37, no. 1, pp. 85–104, Mar. 1995. doi: 10.1518/001872095779049462

[3] M. Adcock, D. Feng, and B. Thomas, "Visualization of off-surface 3D viewpoint locations in spatial augmented reality," in *Proc. SUI*. New York: ACM, 2013, pp. 1–8. doi: 10.1145/2491367.2491378

[4] M. Agrawala, A. C. Beers, I. McDowall, B. Fröhlich, M. Bolas, and P. Hanrahan, "The two-user responsive workbench: Support for collaboration through individual views of a shared space," in *Proc. SIGGRAPH*. New York: ACM, 1997, pp. 327—332. doi: 10.1145/258734.258875

[5] M. Akçayır and G. Akçayır, "Advantages and challenges associated with augmented reality for education: A systematic review of the literature," *Educational Research Review*, vol. 20, pp. 1–11, Feb. 2017. doi: 10.1016/j.edurev.2016.11.002

[6] D. Akers, A. Sherbondy, R. Mackenzie, R. Dougherty, and B. Wandell, "Exploration of the brain's white matter pathways with dynamic queries," in *Proc. VIS*. Los Alamitos: IEEE Computer Society, 2004, pp. 377–384. doi: 10.1109/VISUAL.2004.30

[7] P.-A. Albinsson and S. Zhai, "High precision touch screen interaction," in *Proc. CHI*. New York: ACM, 2003, pp. 105–112. doi: 10.1145/642611.642631

[8] L. Alem and W. Huang, *Recent trends of mobile collaborative augmented reality systems*. New York: Springer, 2011. doi: 10.1007/978-1-4419-9845-3

[9] V. Amrhein, S. Greenland, and B. McShane, "Scientists rise up against statistical significance," *Nature*, vol. 567, no. 7748, pp. 305–307, 2019. doi: 10.1038/d41586-019-00857-9

[10] V. Amrhein, D. Trafimow, and S. Greenland, "Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication," *The American Statistician*, 2018. doi: 10.1080/00031305.2018.1543137

[11] F. Argelaguet and C. Andujar, "A survey of 3D object selection techniques for virtual environments," *Computer & Graphics*, vol. 37, no. 3, pp. 121–136, May 2013. doi: 10.1016/j.cag.2012.12.003

[12] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. Fitzmaurice, "Experimental evaluation of sketching on surfaces in VR," in *Proc. CHI*. New York: ACM, 2017, pp. 5643–5654. doi: 10.1145/3025453.3025474

[13] S. Attfield, G. Kazai, M. Lalmas, and B. Piwowarski, "Towards a science of user engagement (position paper)," in *WSDM Workshop on User Modeling for Web Applications*, 2011.

[14] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, Nov. 2001. doi: 10.1109/38.963459

[15] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, Aug. 1997. doi: 10.1162/pres.1997.6.4.355

[16] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister, "The hologram in my hand: How effective is interactive exploration of 3D visualizations in immersive tangible augmented reality?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 457–467, Jan. 2018. doi: 10.1109/TVCG.2017.2745941

[17] R. M. Baecker, *Readings in groupware and computer-supported cooperative work*. San Francisco: Morgan Kaufmann Publishers, 1993.

[18] T. Baguley, "Standardized or simple effect size: What should be reported?" *British Journal of Psychology*, vol. 100, no. 3, pp. 603–617, Aug. 2009. doi: 10.1348/000712608X377117

[19] R. Balakrishnan, ""beating" fitts' law: virtual enhancements for pointing facilitation," *International Journal of Human-Computer Studies*, vol. 61, no. 6, pp. 857–874, 2004. doi: 10.1016/j.ijhcs.2004.09.002

[20] R. Ballagas, T. E. Dugan, G. Revelle, K. Mori, M. Sandberg, J. Go, E. Reardon, and M. Spasojevic, "Electric agents: Fostering sibling joint media engagement through interactive television and augmented reality," in *Proc. CSCW*. New York: ACM, 2013, pp. 225–236. doi: 10.1145/2441776.2441803

[21] M. Baloup, T. Pietrzak, and G. Casiez, "RayCursor: A 3D pointing facilitation technique based on raycasting," in *Proc. CHI*. New York: ACM, 2019, pp. 101:1–101:12. doi: 10.1145/3290605.3300331

[22] J. Barbier, P. Kenny, J. Young, J.-M. Normand, M. T. Keane, M. O'Sullivan, and A. Ventresque, "MAAP annotate: When archaeology meets augmented reality for annotation of megalithic art," in *Proc. VSMM*. Los Alamitos: IEEE Computer Society, 2017, pp. 283–291. doi: 10.1109/VSMM.2017.8346282

[23] E. Barsom, M. Graafland, and M. Schijven, "Systematic review on the effectiveness of augmented reality applications in medical training," *Surgical endoscopy*, vol. 30, no. 10, pp. 4174–4183, Oct. 2016. doi: 10.1007/s00464-016-4800-6

[24] A. U. Batmaz and W. Stuerzlinger, "Effects of 3D rotational jitter and selection methods on 3D pointing tasks," in *Proc. VR*. Los Alamitos: IEEE Computer Society, Mar. 2019, pp. 1687–1692. doi: 10.1109/VR.2019.8798038

[25] D. Belcher, M. Billinghurst, S. E. Hayes, and R. Stiles, "Using augmented reality for visualizing complex graphs in three dimensions," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2003, pp. 84–93. doi: 10.1109/ISMAR.2003.1240691

[26] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, "Understanding and constructing shared spaces with mixed-reality boundaries," *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 3, pp. 185–223, Sep. 1998. doi: 10.1145/292834.292836

[27] H. Benko and S. Feiner, "Balloon selection: A multi-finger technique for accurate low-fatigue 3D selection," in *Proc. 3DUI*. Los Alamitos: IEEE Computer Society, 2007, pp. 79–86. doi: 10.1109/3DUI.2007.340778

[28] H. Benko, E. W. Ishak, and S. Feiner, "Collaborative mixed reality visualization of an archaeological excavation," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2004, pp. 132–140. doi: 10.1109/ISMAR.2004.23

[29] H. Benko, A. D. Wilson, and P. Baudisch, "Precise selection techniques for multi-touch screens," in *Proc. CHI*. New York: ACM, 2006, pp. 1263–1272. doi: 10.1145/1124772.1124963

[30] H. Benko, A. D. Wilson, and F. Zannier, "Dyadic projected spatial augmented reality," in *Proc. UIST*. New York: ACM, 2014, pp. 645–655. doi: 10.1145/2642918.2647402

[31] H. Benko, A. Wilson, and R. Jota, "MirageTable: Freehand interaction on a projected augmented reality tabletop," in *Proc. CHI*. New York: ACM, 2012, pp. 199–208. doi: 10.1145/2207676.2207704

[32] L. Besançon, "An interaction continuum for 3D dataset visualization," Ph.D. dissertation, Université Paris-Saclay, Dec. 2017.

[33] L. Besançon, M. Ammi, and T. Isenberg, "Pressure-based gain factor control for mobile 3D interaction using locally-coupled devices," in *Proc. CHI*. New York: ACM, 2017, pp. 1831–1842. doi: 10.1145/3025453.3025890

[34] L. Besançon and P. Dragicevic, "La différence significative entre valeurs p et intervalles de confiance (The significant difference between p-values and confidence intervals)," in *Conférence Francophone sur l'Interaction Homme-Machine*. Poitiers, France: AFIHM, Aug. 2017, in French, English translation in [32].

[35] L. Besançon and P. Dragicevic, "The continued prevalence of dichotomous inferences at CHI," in *Proc. CHI*. New York: ACM, May 2019, pp. 1–11. doi: 10.1145/3290607.3310432

[36] L. Besancon, P. Issartel, M. Ammi, and T. Isenberg, "Hybrid tactile/tangible interaction for 3D data exploration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 881–890, Jan. 2017. doi: 10.1109/TVCG.2016.2599217

[37] L. Besançon, P. Issartel, M. Ammi, and T. Isenberg, "Mouse, tactile, and tangible input for 3D manipulation," in *Proc. CHI*. New York: ACM, 2017, pp. 4727–4740. doi: 10.1145/3025453.3025863

[38] L. Besançon, **Mickael Sereno**, M. Ammi, L. Yu, and T. Isenberg, “Hybrid touch/tangible spatial 3D data selection,” *Computer Graphics Forum*, vol. 38, no. 3, pp. 553–567, Jun. 2019. doi: 10.1111/cgf.13710

[39] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg, “The state of the art of spatial interfaces for 3d visualization,” *Computer Graphics Forum*, vol. 40, no. 1, pp. 293–326, Feb. 2021. doi: 10.1111/cgf.14189

[40] P. Bhattacharyya, R. Nath, Y. Jo, K. Jadhav, and J. Hammer, “Brick: Toward a model for designing synchronous colocated augmented reality games,” in *Proc. CHI*. New York: ACM, 2019, pp. 323:1–323:9. doi: 10.1145/3290605.3300553

[41] M. Billinghurst, J. Bowskill, and J. Morphett, “WearCom: A wearable communication space,” in *Proc. CVE*, vol. 98, 1998.

[42] M. Billinghurst, M. Cordeil, A. Bezerianos, and T. Margolis, “Collaborative immersive analytics,” in *Immersive Analytics*. Springer, 2018, pp. 221–257. doi: 10.1007/978-3-030-01388-2_8

[43] M. Billinghurst and H. Kato, “Collaborative mixed reality,” in *Mixed Reality – Merging Real and Virtual Worlds*. Springer-Verlag Berlin Heidelberg, 1999, pp. 261–284.

[44] M. Billinghurst, H. Kato, and I. Poupyrev, “The MagicBook: a transitional AR interface,” *Computers & Graphics*, vol. 25, no. 5, pp. 745–753, Oct. 2001, mixed realities - beyond conventions. doi: 10.1016/S0097-8493(01)00117-0

[45] M. Billinghurst and B. H. Thomas, “Mobile collaborative augmented reality,” in *Recent trends of mobile collaborative augmented reality systems*. Springer, 2011, pp. 1–19. doi: 10.1007/978-1-4419-9845-3

[46] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*. Massachusetts, USA: A. K. Peters, Ltd., 2005.

[47] F. Biocca, C. Harms, and J. K. Burgoon, “Toward a more robust theory and measure of social presence: Review and suggested criteria,” *Presence: Teleoperators & virtual environments*, vol. 12, no. 5, pp. 456–480, Oct. 2003. doi: 10.1162/105474603322761270

[48] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, “Immersive virtual environment technology as a methodological tool for social psychology,” *Psychological Inquiry*, vol. 13, no. 2, pp. 103–124, 2002. doi: 10.1207/S15327965PLI1302_01

[49] R. A. Bolt, ““put-that-there”: Voice and gesture at the graphics interface,” *ACM SIGGRAPH Computer Graphics*, vol. 14, no. 3, pp. 262–270, Jul. 1980. doi: 10.1145/965105.807503

[50] A. Bornik, R. Beichel, E. Kruijff, B. Reitinger, and D. Schmalstieg, “A hybrid user interface for manipulation of volumetric medical data,” in *Proc. 3DUI*. Los Alamitos: IEEE Computer Society, 2006, pp. 29–36. doi: 10.1109/VR.2006.8

[51] D. Bowman, C. Wingrave, J. Campbell, and V. Ly, “Using pinch gloves™ for both natural and abstract interaction techniques in virtual environments,” Department of Computer Science (0106) Virginia Tech, Blacksburg, VA 24061, USA, Tech. Rep. TR-01-23, 2001.

[52] D. A. Bowman and R. P. McMahan, “Virtual reality: How much immersion is enough?” *Computer*, vol. 40, no. 7, pp. 36–43, Jul. 2007. doi: 10.1109/MC.2007.257

[53] D. A. Bowman and L. F. Hodges, “An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments,” in *Proc. I3D*. New York: ACM, 1997, pp. 35–38. doi: 10.1145/253284.253301

[54] D. A. Bowman, D. B. Johnson, and L. F. Hodges, “Testbed evaluation of virtual environment interaction techniques,” in *Proc. VRST*. New York: Association for Computing Machinery, 1999, pp. 26—-33. doi: 10.1145/323663.323667

[55] K. W. Brodlie, L. A. Carpenter, R. A. Earnshaw, J. R. Gallop, R. J. Hubbold, A. M. Mumford, C. D. Osland, and P. Quarendon, *Scientific visualization: techniques and applications*. Springer-Verlag, 1992. doi: 10.1007/978-3-642-76942-9

[56] R. Brouet, R. Blanch, and M.-P. Cani, “Understanding hand degrees of freedom and natural gestures for 3D interaction on tabletop,” in *Human-Computer Interaction–INTERACT*, P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 297–314. doi: 10.1007/978-3-642-40483-2_20

[57] S. Bruckner, T. Isenberg, T. Ropinski, and A. Wiebel, "A model of spatial directness in interactive visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 8, pp. 2514–2528, Aug. 2019. doi: 10.1109/TVCG.2018.2848906

[58] G. Bruder, F. Steinicke, and W. Stürzlinger, "Effects of visual conflicts on 3D selection task performance in stereoscopic display environments," in *Proc. 3DUI*. Los Alamitos: IEEE Computer Society, Mar. 2013, pp. 115–118. doi: 10.1109/3DUI.2013.6550207

[59] W. Büschel, J. Chen, R. Dachselt, S. Drucker, T. Dwyer, C. Görg, T. Isenberg, A. Kerren, C. North, and W. Stuerzlinger, "Interaction for immersive analytics," in *Immersive Analytics*, K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas, Eds. Berlin/Heidelberg: Springer, 2018, ch. 4, pp. 95–138. doi: 10.1007/978-3-030-01388-2_4

[60] W. Büschel, A. Lehmann, and R. Dachselt, "MIRIA: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data," in *Proc. CHI*. New York: ACM, 2021, pp. 1–14. doi: 10.1145/3411764.3445651

[61] W. Büschel, A. Mitschick, T. Meyer, and R. Dachselt, "Investigating smartphone-based pan and zoom in 3D data spaces in augmented reality," in *Proc. MobileHCI*. New York: ACM, 2019, pp. 2:1–2:13. doi: 10.1145/3338286.3340113

[62] A. Butz, T. Höllerer, S. Feiner, B. MacIntyre, and C. Beshers, "Enveloping users and computers in a collaborative 3D augmented reality," in *Proc. IWAR*. Los Alamitos: IEEE Computer Society, 1999, pp. 35–44. doi: 10.1109/IWAR.1999.803804

[63] T. P. Caudell and D. W. Mizell, "Augmented reality: an application of heads-up display technology to manual manufacturing processes," in *Proc. HICSS*. Los Alamitos: IEEE Computer Society, 1992, pp. 659–669. doi: 10.1109/HICSS.1992.183317

[64] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung, "Touching the void: Direct-touch interaction for intangible displays," in *Proc. CHI*. New York: ACM, 2010, pp. 2625–2634. doi: 10.1145/1753326.1753725

[65] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson, "Immersive analytics," in *Proc. BDVA*. IEEE, 2015, pp. 73–80. doi: 10.1109/BDVA.2015.7314296

[66] M. Chen, S. J. Mountford, and A. Sellen, "A study in interactive 3-D rotation using 2-D control devices," *ACM SIGGRAPH Computer Graphics*, vol. 22, no. 4, pp. 121–129, Aug. 1988. doi: 10.1145/378456.378497

[67] Z. Chen, W. Zeng, Z. Yang, L. Yu, C.-W. Fu, and H. Qu, "LassoNet: Deep lasso-selection of 3D point clouds," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 195–204, Jan. 2020. doi: 10.1109/TVCG.2019.2934332

[68] D. Clergeaud, J. S. Roo, M. Hachet, and P. Guitton, "Towards seamless interaction between physical and virtual locations for asymmetric collaboration," in *Proc. VRST*. New York: ACM, 2017, pp. 17:1–17:4. doi: 10.1145/3139131.3139165

[69] A. Cockburn, P. Dragicevic, L. Besançon, and C. Gutwin, "Threats of a replication crisis in empirical computer science," *Commun. ACM*, vol. 63, no. 8, p. 70–79, Jul. 2020. doi: 10.1145/3360311

[70] A. Cockburn, C. Gutwin, and A. Dix, "HARK no more: On the preregistration of CHI experiments," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–12. doi: 10.1145/3173574.3173715

[71] R. Coe, "It's the effect size, stupid: What effect size is and why it is important," in *Proc. Annual Conf. British Educational Research Assoc.*, 2002.

[72] M. Cordeil, B. Bach, Y. Li, E. Wilson, and T. Dwyer, "Design space for spatio-data coordination: Tangible interaction devices for immersive information visualisation," in *Proc. PacificVis*. Los Alamitos: IEEE Computer Society, 2017, pp. 46–50. doi: 10.1109/PACIFICVIS.2017.8031578

[73] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the CAVE," in *Proc. SIGGRAPH*. New York: ACM, 1993, pp. 135–142. doi: 10.1145/166117.166134

[74] M. Csikszentmihalyi and I. Csikszentmihalyi, *Beyond boredom and anxiety*. Jossey-Bass San Francisco, 1975, vol. 721.

[75] G. Cumming, "The new statistics: Why and how," *Psychological Science*, vol. 25, no. 1, pp. 7–29, Jan. 2014. doi: 10.1177/0956797613504966

[76] J. J. Cummings and J. N. Bailenson, "How immersive is enough? a meta-analysis of the effect of immersive technology on user presence," *Media Psychology*, vol. 19, no. 2, pp. 272–309, 2016. doi: 10.1080/15213269.2015.1015740

[77] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Perception of space and motion*. Elsevier, 1995, pp. 69–117. doi: 10.1016/B978-012240530-3/50005-5

[78] T. Czauderna, J. Haga, J. Kim, M. Klapperstück, K. Klein, T. Kuhlen, S. Oeltze-Jafra, B. Sommer, and F. Schreiber, "Immersive analytics applications in life and health sciences," in *Immersive Analytics*. Cham: Springer, 2018, pp. 289–330. doi: 10.1007/978-3-030-01388-2_10

[79] F. Daiber, E. Falk, and A. Krüger, "Balloon selection revisited: Multi-touch selection techniques for stereoscopic data," in *Proc. AVI*. New York: ACM, 2012, pp. 441–444. doi: 10.1145/2254556.2254641

[80] G. de Haan, M. Koutek, and F. H. Post, "Towards intuitive exploration tools for data visualization in VR," in *Proc. VRST*. New York: ACM, 2002, pp. 105–112. doi: 10.1145/585740.585758

[81] N. J. Dedual, O. Oda, and S. K. Feiner, "Creating hybrid user interfaces with a 2D multi-touch tabletop and a 3D see-through head-worn display," in *Proc. ISMAR*. IEEE, 2011, pp. 231–232. doi: 10.1109/ISMAR.2011.6092391

[82] A. DeVincenzi, L. Yao, H. Ishii, and R. Raskar, "Kinected conference: Augmenting video imaging with calibrated depth and audio," in *Proc. CSCW*. New York: ACM, 2011, pp. 621–624. doi: 10.1145/1958824.1958929

[83] M. Dolata, D. Agotai, S. Schubiger, and G. Schwabe, "Pen-and-paper rituals in service interaction: Combining high-touch and high-tech in financial advisory encounters," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. CSCW, Nov. 2019. doi: 10.1145/3359326

[84] S. Dong, A. H. Behzadan, F. Chen, and V. R. Kamat, "Collaborative visualization of engineering processes using tabletop augmented reality," *Advances in Engineering Software*, vol. 55, pp. 45–55, Jan. 2013. doi: 10.1016/j.advengsoft.2012.09.001

[85] P. Dragicevic, "Fair statistical communication in HCI," in *Modern Statistical Methods for HCI*, J. Robertson and M. Kaptein, Eds. Cham, Switzerland: Springer International Publishing, 2016, ch. 13, pp. 291–330. doi: 10.1007/978-3-319-26633-6_13

[86] P. Dragicevic, F. Chevalier, and S. Huot, "Running an HCI experiment in multiple parallel universes," in *CHI Extended Abstracts*. New York: ACM, 2014, pp. 607–618. doi: 10.1145/2559206.2578881

[87] M. Dunleavy, C. Dede, and R. Mitchell, "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning," *Journal of Science Education and Technology*, vol. 18, no. 1, pp. 7–22, Feb. 2009. doi: 10.1007/s10956-008-9119-1

[88] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schreiber, W. Stuerzlinger, and B. Thomas, "Immersive analytics: An introduction," in *Immersive Analytics*. Springer, 2018, pp. 1–23. doi: 10.1007/978-3-030-01388-2_1

[89] S. Eivazi, H. Afkari, R. Bednarik, V. Leinonen, M. Tukiainen, and J. E. Jääskeläinen, "Analysis of disruptive events and precarious situations caused by interaction with neurosurgical microscope," *Acta Neurochirurgica*, vol. 157, no. 7, pp. 1147–1154, Jul. 2015. doi: 10.1007/s00701-015-2433-5

[90] N. ElSayed, B. Thomas, K. Marriott, J. Piantadosi, and R. Smith, "Situated analytics," in *Proc. BDVA*. Los Alamitos: IEEE Computer Society, 2015, pp. 96–103. doi: 10.1109/BDVA.2015.7314302

[91] N. A. ElSayed, B. H. Thomas, K. Marriott, J. Piantadosi, and R. T. Smith, "Situated analytics: Demonstrating immersive analytical tools with augmented reality," *Journal of Visual Languages & Computing*, vol. 36, pp. 13–23, Oct. 2016. doi: 10.1016/j.jvlc.2016.07.006

[92] C. Elvezio, M. Sukan, O. Oda, S. Feiner, and B. Tversky, "Remote collaboration in AR and VR using virtual replicas," in *Proc. SIGGRAPH*. New York: ACM, 2017, pp. 13:1–13:2. doi: 10.1145/3089269.3089281

[93] B. Ens, J. D. Hincapié-Ramos, and P. Irani, "Ethereal planes: A design framework for 2D information space in 3D mixed reality environments," in *Proc. SUI*. New York: ACM, 2014, pp. 2–12. doi: 10.1145/2659766.2659769

[94] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst, "Revisiting collaboration through mixed reality: The evolution of groupware," *International Journal of Human-Computer Studies*, vol. 131, pp. 81–98, Nov. 2019. doi: 10.1016/j.ijhcs.2019.05.011

[95] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon, "Windows on the world: 2D windows for 3D augmented reality," in *Proc. UIST*. New York: ACM, 1993, pp. 145–155. doi: 10.1145/168642.168657

[96] G. W. Fitzmaurice, "Graspable user interfaces," Ph.D. dissertation, University of Toronto, Canada, 1996.

[97] G. W. Fitzmaurice, H. Ishii, and W. A. S. Buxton, "Bricks: Laying the foundations for graspable user interfaces," in *Proc. CHI*. New York: ACM, 1995, pp. 442–449. doi: 10.1145/223904.223964

[98] P. J. Fleming and J. J. Wallace, "How not to lie with statistics: The correct way to summarize benchmark results," *Communication of the ACM*, vol. 29, no. 3, pp. 218–221, Mar. 1986. doi: 10.1145/5666.5673

[99] A. Fonnet and Y. Prié, "Survey of immersive analytics," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019, early Access. doi: 10.1109/TVCG.2019.2929033

[100] G. Fontaine, "The experience of a sense of presence in intercultural and international encounters," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 4, pp. 482–490, Fall 1992. doi: 10.1162/pres.1992.1.4.482

[101] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-touch vs. mouse input for tabletop displays," in *Proc. CHI*. New York: ACM, 2007, pp. 647–656. doi: 10.1145/1240624.1240726

[102] A. Forsberg, J. Chen, and D. Laidlaw, "Comparing 3D vector field visualization methods: A user study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1219–1226, Nov. 2009. doi: 10.1109/TVCG.2009.126

[103] A. Forsberg, K. Herndon, and R. Zeleznik, "Aperture based selection for immersive virtual environments," in *Proc. UIST*. New York: ACM, 1996, pp. 95–96. doi: 10.1145/237091.237105

[104] E. Foxlin, M. Harrington, and G. Pfeifer, "Constellation: A wide-range wireless motion-tracking system for augmented reality and virtual set applications," in *Proc. SIGGRAPH*. New York: ACM, 1998, pp. 371–378. doi: 10.1145/280814.280937

[105] J. Franz, M. Alnusayri, J. Malloch, and D. Reilly, "A comparative evaluation of techniques for sharing AR experiences in museums," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. CSCW, Nov. 2019. doi: 10.1145/3359226

[106] C.-W. Fu, W.-B. Goh, and J. A. Ng, "Multi-touch techniques for exploring large-scale 3D astrophysical simulations," in *Proc. CHI*. New York: ACM, 2010, pp. 2213–2222. doi: 10.1145/1753326.1753661

[107] K.-C. GAN and E. R. HOFFMANN, "Geometrical conditions for ballistic and visually controlled movements," *Ergonomics*, vol. 31, no. 5, pp. 829–839, 1988, pMID: 3402428. doi: 10.1080/00140138808966724

[108] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer, "Integrating the physical environment into mobile remote collaboration," in *Proc. MobileHCI*. New York: ACM, 2012, pp. 241–250. doi: 10.1145/2371574.2371610

[109] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation," in *Proc. VRST*. New York: ACM, 2014, pp. 197–205. doi: 10.1145/2671015.2671016

[110] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "World-stabilized annotations and virtual scene navigation for remote collaboration," in *Proc. UIST*. New York: ACM, 2014, pp. 449–459. doi: 10.1145/2642918.2647372

[111] G. Gigerenzer, "Statistical rituals: The replication delusion and how we got there," *Advances in Methods and Practices in Psychological Science*, p. 2515245918771329, 2018. doi: 10.1177/2515245918771329

[112] S. Gosset, **Mickael Sereno**, L. Besançon, and T. Isenberg, "Tangible volumetric brushing in augmented reality," in *Proc. Vis*. Los Alamitos: IEEE, 2020, extended Abstract.

[113] T. Grossman and R. Balakrishnan, "The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area," in *Proc. CHI*. New York: ACM, 2005, pp. 281–290. doi: 10.1145/1054972.1055012

[114] T. Grossman and R. Balakrishnan, "The design and evaluation of selection techniques for 3D volumetric displays," in *Proc. UIST*. New York: ACM, 2006, pp. 3–12. doi: 10.1145/1166253.1166257

[115] J. Grudin, "Computer-supported cooperative work: history and focus," *Computer*, vol. 27, no. 5, pp. 19–26, May 1994. doi: 10.1109/2.291294

[116] H. Guo, H. Xiao, and X. Yuan, "Scalable multivariate volume visualization and analysis based on dimension projection and parallel coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1397–1410, 2012. doi: 10.1109/TVCG.2012.80

[117] K. Gupta, G. A. Lee, and M. Billinghurst, "Do you see what I see? the effect of gaze tracking on task space remote collaboration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2413–2422, Nov. 2016. doi: 10.1109/TVCG.2016.2593778

[118] C. Gutwin and S. Greenberg, "A descriptive framework of workspace awareness for real-time groupware," *Computer Supported Cooperative Work*, vol. 11, no. 3, pp. 411–446, Sep. 2002. doi: 10.1023/A:1021271517844

[119] R. Harmon, W. Patterson, W. Ribarsky, and J. Bolter, "The virtual annotation system," in *Proc. VR*. Los Alamitos: IEEE Computer Society, Mar. 1996, pp. 239–245. doi: 10.1109/VRAIS.1996.490533

[120] C. Harms and F. Biocca, "Internal consistency and reliability of the networked minds measure of social presence," in *Seventh Annual International Workshop: Presence*, 2004.

[121] S. G. Hart, "Nasa-task load index (NASA-TLX); 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006. doi: 10.1177/154193120605000909

[122] C. Heeter, "Being there: The subjective experience of presence," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 2, pp. 262–271, Spring 1992. doi: 10.1162/pres.1992.1.2.262

[123] J. Heinrich and D. Weiskopf, "Continuous parallel coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1531–1538, 2009. doi: 10.1109/TVCG.2009.131

[124] J. Helske, S. Helske, M. Cooper, A. Ynnerman, and L. Besancon, "Can visualization alleviate dichotomous thinking effects of visual representations on the cliff effect," *IEEE Transactions on Visualization & Computer Graphics*, vol. 63, no. 8, pp. 70–79, Apr. 5555. doi: 10.1109/TVCG.2021.3073466

[125] S. J. Henderson and S. Feiner, "Opportunistic controls: Leveraging natural affordances as tangible user interfaces for augmented reality," in *Proc. VRST*. New York: ACM, 2008, pp. 211–218. doi: 10.1145/1450579.1450625

[126] S. J. Henderson and S. K. Feiner, "Augmented reality in the psychomotor phase of a procedural task," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2011, pp. 191–200. doi: 10.1109/ISMAR.2011.6092386

[127] K. P. Herndon, A. van Dam, and M. Gleicher, "The challenges of 3D interaction: A CHI '94 workshop," *SIGCHI Bull.*, vol. 26, no. 4, pp. 36–43, Oct. 1994. doi: 10.1145/191642.191652

[128] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "Consumed endurance: A metric to quantify arm fatigue of mid-air interactions," in *Proc. CHI*. New York: ACM, 2014, pp. 1063–1072. doi: 10.1145/2556288.2557130

[129] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell, "Passive real-world interface props for neurosurgical visualization," in *Proc. CHI*. New York: ACM, 1994, pp. 452–458. doi: 10.1145/191666.191821

[130] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell, "Usability analysis of 3D rotation techniques," in *Proc. UIST*. New York: ACM, 1997, pp. 1–10. doi: 10.1145/263407.263408

[131] W. Huang, M. Billinghurst, L. Alem, and S. Kim, "HandsInTouch: Sharing gestures in remote collaboration," in *Proc. OzCHI*. New York: ACM, 2018, pp. 396–400. doi: 10.1145/3292147.3292177

[132] C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier, and R. Vuillemot, "FiberClay: Sculpting three dimensional trajectories to reveal structural insights," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 704–714, Jan 2019. doi: 10.1109/TVCG.2018.2865191

[133] D.-N. T. Huynh, K. Raveendran, Y. Xu, K. Spreen, and B. MacIntyre, "Art of defense: A collaborative handheld augmented reality board game," in *Proc. SIGGRAPH*. New York: ACM, 2009, pp. 135–142. doi: 10.1145/1581073.1581095

[134] A. Irlitti, R. T. Smith, S. Von Itzstein, M. Billinghurst, and B. H. Thomas, "Challenges for asynchronous collaboration in augmented reality," in *Proc. ISMAR-Adjunct*. IEEE, 2016, pp. 31–35. doi: 10.1109/ISMAR-Adjunct.2016.0032

[135] A. Irlitti, S. Von Itzstein, R. T. Smith, and B. H. Thomas, "Performance improvement using data tags for handheld spatial augmented reality," in *Proc. VRST*. New York: ACM, 2014, p. 161–165. doi: 10.1145/2671015.2671116

[136] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative visualization: Definition, challenges, and research agenda," *Information Visualization*, vol. 10, no. 4, pp. 310–326, Oct. 2011. doi: 10.1177/1473871611412817

[137] T. Isenberg, "Interactive exploration of three-dimensional scientific visualizations on large display surfaces," in *Collaboration Meets Interactive Spaces*, C. Anslow, P. Campos, and J. Jorge, Eds. Berlin/Heidelberg: Springer, 2016, ch. 6, pp. 97–123. doi: 10.1007/978-3-319-45853-3_6

[138] T. Isenberg and M. Hancock, "*Gestures* vs. postures: 'Gestural' touch interaction in 3D environments," in *Proc. 3DCHI*, 2012, pp. 53–61.

[139] H. Ishii and B. Ullmer, "Tangible bits: Towards seamless interfaces between people, bits and atoms," in *Proc. CHI*. New York: ACM, 1997, pp. 234–241. doi: 10.1145/258549.258715

[140] P. Issartel, L. Besançon, T. Isenberg, and M. Ammi, "A tangible volume for portable 3D interaction," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, Sep. 2016, pp. 215–220. doi: 10.1109/ISMAR-Adjunct.2016.0079

[141] B. Jackson, T. Y. Lau, D. Schroeder, K. C. Toussaint, and D. F. Keefe, "A lightweight tangible 3D interface for interactive visualization of thin fiber structures," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2802–2809, Dec. 2013. doi: 10.1109/TVCG.2013.121

[142] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani, "Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion," in *Proc. CHI*. New York: ACM, 2017, pp. 3328–3339. doi: 10.1145/3025453.3025523

[143] Y. Jansen, P. Dragicevic, and J.-D. Fekete, "Evaluating the efficiency of physical visualizations," in *Proc. CHI*. New York: ACM, Apr. 2013, pp. 2593–2602. doi: 10.1145/2470654.2481359

[144] C. R. Johnson, "Top scientific visualization research problems," *IEEE Computer Graphics and Applications*, vol. 24, no. 4, pp. 13–17, Jul./Aug. 2004. doi: 10.1109/MCG.2004.20

[145] P. Kán, A. Dünser, M. Billinghurst, and C. Schönauer, "The effects of direct and global illumination on presence in augmented reality," in *Challenging Presence: Proceedings of 15th International Conference on Presence*. Vienna, Austria: Facultas Verlags-und Buchhandels AG, 2014, pp. 223–230.

[146] S. Kasahara, V. Heun, A. S. Lee, and H. Ishii, "Second surface: Multi-user spatial collaboration system based on augmented reality," in *Proc. SIGGRAPH*. New York: ACM, 2012, p. 1–4. doi: 10.1145/2407707.2407727

[147] D. F. Keefe and T. Isenberg, "Reimagining the scientific visualization interaction paradigm," *Computer*, vol. 46, no. 5, pp. 51–57, May 2013. doi: 10.1109/MC.2013.178

[148] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch, "Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008-2017)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2947–2962, Nov. 2018. doi: 10.1109/TVCG.2018.2868591

[149] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in *Proc. ISMAR*. IEEE, 2014, pp. 83–92. doi: 10.1109/ISMAR.2014.6948412

[150] C. Kirner and T. G. Kirner, "A data visualization virtual environment supported by augmented reality," in *Proc. SMC*. Los Alamitos: IEEE Computer Society, 2006, pp. 97–102. doi: 10.1109/ICSMC.2006.384365

[151] K. Kiyokawa, H. Takemura, and N. Yokoya, "A collaboration support technique by integrating a shared virtual reality and a shared augmented reality," in *Proc. SMC*. IEEE, 1999, pp. 48–53. doi: 10.1109/ICSMC.1999.816444

[152] M. Klapperstuck, T. Czauderna, C. Goncu, J. Glowacki, T. Dwyer, F. Schreiber, and K. Marriott, "ContextuWall: Peer collaboration using (large) displays," in *Proc. BDVA*. Los Alamitos: IEEE Computer Society, 2016, pp. 1–8. doi: 10.1109/BDVA.2016.7787047

[153] T. Klein, F. Guéniat, L. Pastur, F. Vernier, and T. Isenberg, "A design study of direct-touch interaction for exploratory 3D scientific visualization," *Computer Graphics Forum*, vol. 31, no. 3pt3, pp. 1225–1234, Jun. 2012. doi: 10.1111/j.1467-8659.2012.03115.x

[154] J. Kniss, S. Premoze, M. Ikits, A. Lefohn, C. Hansen, and E. Praun, "Gaussian transfer functions for multi-field volume visualization," in *Proc. Visualization*. Los Alamitos: IEEE Computer Society, 2003, pp. 497–504. doi: 10.1109/VISUAL.2003.1250412

[155] R. Komiyama, T. Miyaki, and J. Rekimoto, "JackIn space: Designing a seamless transition between first and third person view for effective telepresence collaborations," in *Proc. AH*. New York: ACM, 2017, pp. 14:1–14:9. doi: 10.1145/3041164.3041183

[156] M. Krzywinski and N. Altman, "Points of significance: Error bars," *Nature Methods*, vol. 10, no. 10, pp. 921–922, Oct. 2013. doi: 10.1038/nmeth.2659

[157] O.-H. Kwon, C. Muelder, K. Lee, and K.-L. Ma, "A study of layout, rendering, and interaction methods for immersive graph visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, pp. 1802–1815, Jul. 2016. doi: 10.1109/TVCG.2016.2520921

[158] W. S. Lages and D. A. Bowman, "Walking with adaptive augmented reality workspaces: Design and usage patterns," in *Proc. IUI*. ACM, 2019, pp. 356—366. doi: 10.1145/3301275.3302278

[159] R. Langner, M. Satkowski, W. Büschel, and R. Dachselt, "MARVIS: Combining mobile devices and augmented reality for visual data analysis," in *Proc. CHI*. New York: ACM, 2021, pp. 1–16. doi: 10.1145/3411764.3445593

[160] J. J. LaViola, Jr., "A discussion of cybersickness in virtual environments," *SIGCHI Bulletin*, vol. 32, no. 1, pp. 47–56, Jan. 2000. doi: 10.1145/333329.333344

[161] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev, *3D user interfaces: theory and practice*. Boston: Addison-Wesley Professional, 2017.

[162] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "The stretchable arms for collaborative remote guiding," in *Proc. ICAT - EGVE*. Aire-la-Ville, Switzerland: Eurographics Association, 2015, pp. 147–150. doi: 10.2312/egve.20151322

[163] M. Le Goc, P. Dragicevic, S. Huron, J. Boy, and J.-D. Fekete, "A better grasp on pictures under glass: Comparing touch and tangible object manipulation using physical proxies," in *Proc. AVI*. New York: ACM, 2016, pp. 76–83. doi: 10.1145/2909132.2909247

[164] B. Lee, A. Srinivasan, P. Isenberg, and J. Stasko, "Post-WIMP interaction for information visualization," *Foundations and Trends in Human–Computer Interaction*, vol. 14, no. 1, pp. 1–95, 2021. doi: 10.1561/1100000081

[165] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "Mixed reality collaboration through sharing a live panorama," in *Proc. SA*. New York: ACM, 2017, pp. 14:1–14:4. doi: 10.1145/3132787.3139203

[166] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "A user study on MR remote collaboration using live 360 video," in *Proc. ISMAR*. IEEE, 2018, pp. 153–164. doi: 10.1109/ISMAR.2018.00051

[167] K. M. Lee, "Presence, explicated," *Communication Theory*, vol. 14, no. 1, pp. 27–50, Feb. 2004. doi: 10.1111/j.1468-2885.2004.tb00302.x

[168] S. Lee and S. Zhai, "The performance of touch screen soft buttons," in *Proc. CHI*. New York: ACM, 2009, pp. 309–318. doi: 10.1145/1518701.1518750

[169] N. H. Lehment, D. Merget, and G. Rigoll, "Creating automatically aligned consensus realities for AR videoconferencing," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2014, pp. 201–206. doi: 10.1109/ISMAR.2014.6948428

[170] K.-C. Lien, B. Nuernberger, T. Höllerer, and M. Turk, "PPV: Pixel-point-volume segmentation for object referencing in collaborative augmented reality," in *Proc. ISMAR*. IEEE, 2016, pp. 77–83. doi: 10.1109/ISMAR.2016.21

[171] H. Limerick, D. Coyle, and J. W. Moore, "The experience of agency in human-computer interactions: a review," *Frontiers in Human Neuroscience*, vol. 8, p. 643, 2014. doi: 10.3389/fnhum.2014.00643

[172] P. Lincoln, G. Welch, A. Nashel, A. Ilie, A. State, and H. Fuchs, "Animatronic shader lamps avatars," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2009, pp. 27–33. doi: 10.1109/ISMAR.2009.5336503

[173] S. Liu, D. Cheng, and H. Hua, "An optical see-through head mounted display with addressable focal planes," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2008, pp. 33–42. doi: 10.1109/ISMAR.2008.4637321

[174] J. Looser, M. Billinghurst, R. Grasset, and A. Cockburn, "An evaluation of virtual lenses for object selection in augmented reality," in *Proc. GRAPHITE*. New York: ACM, 2007, pp. 203–210. doi: 10.1145/1321261.1321297

[175] D. López, L. Oehlberg, C. Doger, and T. Isenberg, "Towards an understanding of mobile touch navigation in a stereoscopic viewing environment for 3D data exploration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 5, pp. 1616–1629, May 2016. doi: 10.1109/TVCG.2015.2440233

[176] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa, "Collaboration in augmented reality," *Computer Supported Cooperative Work*, vol. 24, no. 6, pp. 515–525, Dec. 2015. doi: 10.1007/s10606-015-9239-0

[177] A. MacWilliams, C. Sandor, M. Wagner, M. Bauer, G. Klinker, and B. Bruegge, "Herding sheep: Live system development for distributed augmented reality," in *Proc. ISMAR*. IEEE, 2003, pp. 123–132. doi: 10.1109/ISMAR.2003.1240695

[178] T. Mahmood, W. Fulmer, N. Mungoli, J. Huang, and A. Lu, "Improving information sharing and collaborative analysis for remote GeoSpatial visualization using mixed reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, Oct. 2019, pp. 236–247. doi: 10.1109/ISMAR.2019.00021

[179] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2011, pp. 137–146. doi: 10.1109/ISMAR.2011.6092379

[180] G. E. Marai, A. G. Forbes, and A. Johnson, "Interdisciplinary immersive analytics at the electronic visualization laboratory: Lessons learned and upcoming challenges," in *Workshop on Immersive Analytics*. Los Alamitos: IEEE Computer Society, 2016, pp. 54–59. doi: 10.1109/IMMERSIVE.2016.7932384

[181] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, R. N. Henry, T. Itoh, W. Stuerzlinger, and B. H. Thomas, *Immersive Analytics*. Berlin/Heidelberg: Springer, 2018. doi: 10.1007/978-3-030-01388-2

[182] P. Marshall, R. Fleck, A. Harris, J. Rick, E. Hornecker, Y. Rogers, N. Yuill, and N. S. Dalton, "Fighting for control: Children's embodied interactions when using physical and digital representations," in *Proc. CHI*. New York: ACM, 2009, pp. 2149–2152. doi: 10.1145/1518701.1519027

[183] A. Martinet, G. Casiez, and L. Grisoni, "The effect of DOF separation in 3D manipulation tasks with multi-touch displays," in *Proc. VRST*. New York: ACM, 2010, pp. 111–118. doi: 10.1145/1889863.1889888

[184] A. Martinet, G. Casiez, and L. Grisoni, "Integrality and separability of multi-touch interaction techniques in 3D manipulation tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 3, pp. 369–380, Jan. 2012. doi: 10.1109/TVCG.2011.129

[185] S. Matsutomo, T. Miyauchi, S. Noguchi, and H. Yamashita, "Real-time visualization system of magnetic field utilizing augmented reality technology for education," *IEEE Transactions on Magnetics*, vol. 48, no. 2, pp. 531–534, Feb. 2012. doi: 10.1109/TMAG.2011.2174208

[186] J. P. McIntire, P. R. Havig, and E. E. Geiselman, "Stereoscopic 3D displays and human performance: A comprehensive review," *Displays*, vol. 35, no. 1, pp. 18–26, Jan. 2014. doi: 10.1016/j.displa.2013.10.004

[187] B. B. McShane and D. Gal, "Statistical significance and the dichotomization of evidence," *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 885–895, 2017. doi: 10.1080/01621459.2017.1289846

[188] D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. E. Smith, "Optimality in human motor performance: ideal control of rapid aimed movements." *Psychological review*, vol. 95, no. 3, pp. 340–370, 1988. doi: 10.1037/0033-295X.95.3.340

[189] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transactions on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, Dec. 1994.

[190] R. A. Montano-Murillo, C. Nguyen, R. H. Kazi, S. Subramanian, S. DiVerdi, and D. Martinez-Plasencia, "Slicing-volume: Hybrid 3D/2D multi-target selection technique for dense virtual environments," in *Proc. VR*. Los Alamitos: IEEE Computer Society, 2020, pp. 53–62. doi: 10.1109/VR46266.2020.00023

[191] S. Mora, A. Boron, and M. Divitini, "CroMAR: Mobile augmented reality for supporting reflection on crowd management," *International Journal of Mobile Human Computer Interaction*, vol. 4, no. 2, pp. 88—-101, Apr. 2012. doi: 10.4018/jmhci.2012040107

[192] K. Moreland, "Diverging color maps for scientific visualization," in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, Y. Kuno, J. Wang, R. Pajarola, P. Lindstrom, A. Hinkenjann, M. L. Encarnação, C. T. Silva, and D. Coming, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 92–103.

[193] S. Mori, S. Ikeda, and H. Saito, "A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, p. 17, Jun. 2017. doi: 10.1186/s41074-017-0028-1

[194] J. Müller, R. Rädle, and H. Reiterer, "Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing," in *Proc. CHI*. New York: ACM, 2017, pp. 6481–6486. doi: 10.1145/3025453.3025717

[195] M. A. Nacenta, P. Baudisch, H. Benko, and A. Wilson, "Separability of spatial manipulations in multi-touch interfaces," in *Proc. GI*, ser. GI '09. Toronto, Ontario, Canada: Canadian Human-Computer Communications Society, 2009, pp. 175–182.

[196] S. Nilsson, B. Johansson, and A. Jonsson, "Using AR to support cross-organisational collaboration in dynamic tasks," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2009, pp. 3–12. doi: 10.1109/ISMAR.2009.5336522

[197] D. A. Norman, "Cognitive engineering," in *User Centered System Design*, D. Norman and S. Draper, Eds., 1986, pp. 31–61.

[198] H. L. O'Brien and E. G. Toms, "What is user engagement? a conceptual framework for defining user engagement with technology," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 6, pp. 938–955, Apr. 2008. doi: 10.1002/asi.20801

[199] O. Oda and S. Feiner, "3D referencing techniques for physical objects in shared augmented reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2012, pp. 207–215. doi: 10.1109/ISMAR.2012.6402558

[200] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky, "Virtual replicas for remote assistance in virtual and augmented reality," in *Proc. UIST*. New York: ACM, 2015, pp. 405–415. doi: 10.1145/2807442.2807497

[201] O. Oda and S. Feiner, "Interference avoidance in multi-user hand-held augmented reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2009, pp. 13–22. doi: 10.1109/ISMAR.2009.5336507

[202] K. O'Hara, G. Gonzalez, A. Sellen, G. Penney, A. Varnavas, H. Mentis, A. Criminisi, R. Corish, M. Rouncefield, N. Dastur, and T. Carrell, "Touchless interaction in surgery," *Communications of the ACM*, vol. 57, no. 1, pp. 70–77, Jan. 2014. doi: 10.1145/2541883.2541899

[203] I. C. Olson, Z. Atrash Leong, U. Wilensky, and M. S. Horn, ""It's just a toolbar!" using tangibles to help children manage conflict around a multi-touch tabletop," in *Proc. TEI*. New York: ACM, 2011, pp. 29–36. doi: 10.1145/1935701.1935709

[204] S. K. Ong, M. L. Yuan, and A. Y. C. Nee, "Augmented reality applications in manufacturing: a survey," *International Journal of Production Research*, vol. 46, no. 10, pp. 2707–2742, Mar. 2008. doi: 10.1080/00207540601064773

[205] S. Ong and Y. Shen, "A mixed reality environment for collaborative product design and development," *CIRP Annals*, vol. 58, no. 1, pp. 139–142, 2009. doi: 10.1016/j.cirp.2009.03.020

[206] J. Ou, S. R. Fussell, X. Chen, L. D. Setlock, and J. Yang, “Gestural communication over video stream: Supporting multimodal interaction for remote collaborative physical tasks,” in *Proc. ICMI*. New York: ACM, 2003, pp. 242–249. doi: 10.1145/958432.958477

[207] K. S. Park, A. Kapoor, and J. Leigh, “Lessons learned from employing multiple perspectives in a collaborative virtual environment for visualizing scientific data,” in *Proc. CVE*. New York: ACM, 2000, pp. 73—-82. doi: 10.1145/351006.351015

[208] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson, “Room2Room: Enabling life-size telepresence in a projected augmented reality environment,” in *Proc. CSCW*. New York: ACM, 2016, pp. 1716–1725. doi: 10.1145/2818048.2819965

[209] K. Perlin and E. M. Hoffert, “Hypertexture,” in *Proc. SIGGRAPH*. New York: ACM, 1989, pp. 253–262. doi: 10.1145/74333.74359

[210] D. N. E. Phon, M. B. Ali, and N. D. A. Halim, “Collaborative augmented reality in education: A review,” in *Proc. LaTiCE*. Los Alamitos: IEEE Computer Society, 2014, pp. 78–83. doi: 10.1109/LaTiCE.2014.23

[211] W. Piekarski and B. Thomas, “ARQuake: The outdoor augmented reality gaming system,” *Communications of the ACM*, vol. 45, no. 1, pp. 36–38, Jan. 2002. doi: 10.1145/502269.502291

[212] J. S. Pierce, B. C. Stearns, and R. Pausch, “Voodoo dolls: Seamless interaction at multiple scales in virtual environments,” in *Proc. I3D*. New York: ACM, 1999, pp. 141–145. doi: 10.1145/300523.300540

[213] B. Piper, C. Ratti, and H. Ishii, “Illuminating clay: a 3D tangible interface for landscape analysis,” in *Proc. CHI*. New York: ACM, 2002, pp. 355–362. doi: 10.1145/503376.503439

[214] A. Pister, P. Buono, J.-D. Fekete, C. Plaisant, and P. Valdivia, “Integrating prior knowledge in mixed initiative social network clustering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1775–1785, Feb. 2021. doi: 10.1109/TVCG.2020.3030347

[215] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst, “The effects of sharing awareness cues in collaborative mixed reality,” *Frontiers in Robotics and AI*, vol. 6, pp. 5:1–5:18, Feb. 2019. doi: 10.3389/frobt.2019.00005

[216] T. Piumsomboon, G. A. Lee, B. Ens, B. H. Thomas, and M. Billinghurst, “Superman vs giant: A study on spatial perception for a multi-scale mixed reality flying telepresence interface,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2974–2982, Nov. 2018. doi: 10.1109/TVCG.2018.2868594

[217] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, “Mini-Me: An adaptive avatar for mixed reality remote collaboration,” in *Proc. CHI*. New York: ACM, 2018, pp. 46:1–46:13. doi: 10.1145/3173574.3173620

[218] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst, “On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction,” in *Proc. CHI*. New York: ACM, 2019, pp. 228:1–228:17. doi: 10.1145/3290605.3300458

[219] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst, “CoVAR: A collaborative virtual and augmented reality system for remote collaboration,” in *Proc. SIGGRAPH Asia*. New York: ACM, 2017, pp. 3:1–3:2. doi: 10.1145/3132818.3132822

[220] R. Poelman, O. Akman, S. Lukosch, and P. Jonker, “As if being there: Mediated reality for crime scene investigation,” in *Proc. CSCW*. New York: ACM, 2012, pp. 1267–1276. doi: 10.1145/2145204.2145394

[221] L. Poretski, J. Lanir, and O. Arazy, “Normative tensions in shared augmented reality,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 142:1–142:22, Nov. 2018. doi: 10.1145/3274411

[222] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, “The go-go interaction technique: Non-linear mapping for direct manipulation in VR,” in *Proc. UIST*. New York: ACM, 1996, pp. 79–80. doi: 10.1145/237091.237102

[223] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst, “Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques,” *Computer Graphics Forum*, vol. 17, no. 3, pp. 41–52, Sep. 1998. doi: 10.1111/1467-8659.00252

[224] E. Prytz, S. Nilsson, and A. Jönsson, "The importance of eye-contact for collaboration in AR systems," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2010, pp. 119–126. doi: 10.1109/ISMAR.2010.5643559

[225] R. Raskar, G. Welch, and H. Fuchs, "Spatially augmented reality," in *Proc. IWAR*. Natick, MA, USA: A. K. Peters, Ltd., 1999, pp. 63–72.

[226] P. Reipschläger and R. Dachselt, "DesignAR: Immersive 3D-modeling combining augmented reality with interactive displays," in *Proc. ISS*. New York: ACM, 2019, pp. 29–41. doi: 10.1145/3343055.3359718

[227] P. Reipschläger, T. Flemisch, and R. Dachselt, "Personal augmented reality for information visualization on large interactive displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1182–1192, 2020. doi: 10.1109/TVCG.2020.3030460

[228] J. Rekimoto, "Transvision: A hand-held augmented reality system for collaborative design," in *Proc. VSMM*, 1996, pp. 85–90.

[229] P. Rheingans, "Are we there yet? Exploring with dynamic visualization," *IEEE Computer Graphics and Applications*, vol. 22, no. 1, pp. 6–10, Jan. 2002. doi: 10.1109/38.974511

[230] H. Ro, S. Chae, I. Kim, J. Byun, Y. Yang, Y. Park, and T. Han, "A dynamic depth-variable ray-casting interface for object manipulation in AR environments," in *Proc. SMC*. Los Alamitos: IEEE Computer Society, Oct. 2017, pp. 2873–2878. doi: 10.1109/SMC.2017.8123063

[231] T. Rodden, "A survey of CSCW systems," *Interacting with Computers*, vol. 3, no. 3, pp. 319–353, December 1991. doi: 10.1016/0953-5438(91)90020-3

[232] J. S. Roo, R. Gervais, and M. Hachet, "Inner garden: An augmented sandbox designed for self-reflection," in *Proc. TEI*. New York: ACM, 2016, pp. 570–576. doi: 10.1145/2839462.2856532

[233] J. S. Roo and M. Hachet, "One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities," in *Proc. UIST*. New York: ACM, 2017, pp. 787–795. doi: 10.1145/3126594.3126638

[234] J. S. Roo and M. Hachet, "Towards a hybrid space combining spatial augmented reality and virtual reality," in *Proc. 3DUI*. Los Alamitos: IEEE Computer Society, 2017, pp. 195–198. doi: 10.1109/3DUI.2017.7893339

[235] D. A. W. Rosa and H. H. Nagel, "Selection techniques for dense and occluded virtual 3D environments, supported by depth feedback: Double, bound and depth bubble cursors," in *Proc. SCCC*. Los Alamitos: IEEE Computer Society, 2010, pp. 218–225. doi: 10.1109/SCCC.2010.51

[236] L. Sambrooks and B. Wilkinson, "Comparison of gestural, touch, and mouse interaction with fitts' law," in *Proc. OzCHI*. New York: ACM, 2013, pp. 119–122. doi: 10.1145/2541016.2541066

[237] F. Samsel, M. Petersen, G. Abram, T. L. Turton, D. Rogers, and J. Ahrens, "Visualization of ocean currents and eddies in a high-resolution global ocean-climate model," in *Proceedings of the international conference on high performance computing, networking, storage and analysis*, 2015.

[238] J. Sauro and J. R. Lewis, "Average task times in usability tests: What to report?" in *Proc. CHI*. New York: ACM, 2010, pp. 2347–2350. doi: 10.1145/1753326.1753679

[239] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L. M. Encarnação, M. Gervautz, and W. Purgathofer, "The Studierstube augmented reality project," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 1, pp. 33–54, Feb. 2002. doi: 10.1162/105474602317343640

[240] D. Schmalstieg, A. Fuhrmann, Z. Szalavari, and M. Gervautz, "Studierstube—An environment for collaboration in augmented reality," in *Proc. CVE*, 1996.

[241] A. Sherbondy, D. Akers, R. Mackenzie, R. Dougherty, and B. Wandell, "Exploring connectivity of the brain's white matter with dynamic queries," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 4, pp. 419–430, Jul. 2005. doi: 10.1109/TVCG.2005.59

[242] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009. doi: 10.1098/rstb.2009.0138

[243] G. Smith and J. Mariani, "Using subjective views to enhance 3D applications," in *Proc. VRST*. New York: ACM, 1997, p. 139–146. doi: 10.1145/261135.261161

[244] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "BeThere: 3D mobile collaboration with spatial input," in *Proc. CHI*. New York: ACM, 2013, pp. 179–188. doi: 10.1145/2470654.2470679

[245] P. Song, W. B. Goh, C.-W. Fu, Q. Meng, and P.-A. Heng, "WYSIWYF: Exploring and annotating volume data with a tangible handheld device," in *Proc. CHI*. New York: ACM, 2011, pp. 1333–1342. doi: 10.1145/1978942.1979140

[246] R. R. Springmeyer, M. M. Blattner, and N. L. Max, "A characterization of the scientific data analysis process," in *Proc. VIS*. Los Alamitos: IEEE Computer Society, 1992, pp. 235–242. doi: 10.1109/VISUAL.1992.235203

[247] A. Stafford, W. Piekarski, and B. Thomas, "Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2006, pp. 165–172. doi: 10.1109/ISMAR.2006.297809

[248] R. Stenholt, "Efficient selection of multiple objects on a large scale," in *Proc. VRST*. New York: ACM, 2012, pp. 105—112. doi: 10.1145/2407336.2407357

[249] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual reality on a WIM: Interactive worlds in miniature," in *Proc. CHI*. New York: ACM, 1995, pp. 265–272. doi: 10.1145/223904.223938

[250] W. Stuerzlinger and C. A. Wingrave, *The Value of Constraints for 3D User Interfaces*. Vienna: Springer Vienna, 2011, pp. 203–223. doi: 10.1007/978-3-211-99178-7_11

[251] N. Sugano, H. Kato, and K. Tachibana, "The effects of shadow representation of virtual objects in augmented reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, Oct. 2003, pp. 76–83. doi: 10.1109/ISMAR.2003.1240690

[252] H. B. Surale, A. Gupta, M. Hancock, and D. Vogel, "TabletInVR: Exploring the design space for using a multi-touch tablet in virtual reality," in *Proc. CHI*. New York: ACM, 2019, pp. 13:1–13:13. doi: 10.1145/3290605.3300243

[253] I. E. Sutherland, "A head-mounted three dimensional display," in *Proc. AFIPS*. New York: ACM, 1968, pp. 757–764. doi: 10.1145/1476589.1476686

[254] Z. Szalavári, E. Eckstein, and M. Gervautz, "Collaborative gaming in augmented reality," in *Proc. VRST*. New York: ACM, 1998, pp. 195–204. doi: 10.1145/293701.293740

[255] Z. Szalavári, D. Schmalstieg, A. Fuhrmann, and M. Gervautz, """Studierstube": An environment for collaboration in augmented reality," *Virtual Reality*, vol. 3, no. 1, pp. 37–48, Mar. 1998. doi: 10.1007/BF01409796

[256] M. Tatzgern, D. Kalkofen, and D. Schmalstieg, "Multi-perspective compact explosion diagrams," *Computers & Graphics*, vol. 35, no. 1, pp. 135–147, Feb. 2011. doi: 10.1016/j.cag.2010.11.005

[257] R. M. Taylor, T. C. Hudson, A. Seeger, H. Weber, J. Juliano, and A. T. Helser, "VRPN: A device-independent, network-transparent VR peripheral system," in *Proc. VRST*. New York: ACM, 2001, pp. 55—-61. doi: 10.1145/505008.505019

[258] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock, "Mixed reality remote collaboration combining 360 video and 3D reconstruction," in *Proc. CHI*. New York: ACM, 2019, pp. 201:1–201:14. doi: 10.1145/3290605.3300431

[259] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock, "Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration," in *Proc. OzCHI*. New York: ACM, 2018, pp. 406–410. doi: 10.1145/3292147.3292200

[260] **Mickäel Francisco Sereno**, B. Köhler, and B. Preim, "Comparison of divergence-free filters for cardiac 4D PC-MRI data," in *Bildverarbeitung für die Medizin 2018*, A. Maier, T. M. Deserno, H. Handels, K. H. Maier-Hein, C. Palm, and T. Tolxdorff, Eds. Springer Berlin Heidelberg, 2018, pp. 139–144.

[261] **Mickael Sereno**, M. Ammi, T. Isenberg, and L. Besançon, "Tangible brush: Performing 3D selection with portable and position-aware devices," in *Proc. Vis*. Los Alamitos: IEEE, 2016, extended Abstract.

[262] **Mickael Sereno**, L. Besançon, and T. Isenberg, "Supporting volumetric data visualization and analysis by combining augmented reality visuals with multi-touch input," *Computer Graphics Forum*, 2019, extended Abstract.

[263] **Mickael Sereno**, L. Besançon, and T. Isenberg, "Point specification in collaborative visualization for 3d scalar fields using augmented reality," *Virtual Reality*, pp. 1–19, 2021, to appear. doi: 10.1007/s10055-021-00614-2

[264] **Mickael Sereno**, S. Gosset, L. Besançon, and T. Isenberg, "Hybrid touch/tangible spatial selection in augmented reality," in *Proc. EuroVis*. Los Alamitos: IEEE, 2022, pp. 1–12, in submission.

[265] **Mickael Sereno** and T. Isenberg, "Subjective views in co-located augmented reality — initial design," in *Proc. Vis*. Los Alamitos: IEEE, 2020, extended Abstract.

[266] **Mickael Sereno**, S. Lupone, M. Debiossac, N. Kalashnyk, and P. Roncin, "Active correction of the tilt angle of the surface plane with respect to the rotation axis during azimuthal scan," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 382, pp. 123–126, 2016. doi: 10.1016/j.nimb.2016.05.001

[267] **Mickael Sereno**, X. Wang, L. Besançon, M. J. Mcguffin, and T. Isenberg, "Collaborative work in augmented reality: A survey," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–20, 2020, in-press. doi: 10.1109/TVCG.2020.3032761

[268] S. Thanyadit, P. Punpongsanon, and T. Pong, "ObserVAR: Visualization system for observing virtual reality users using augmented reality," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, Oct. 2019, pp. 258–268. doi: 10.1109/ISMAR.2019.00023

[269] J. J. Thomas and K. A. Cook, *Illuminating the path: The research and development agenda for visual analytics*. United States. Department of Homeland Security, 2005.

[270] A. Top, G. Hamarneh, and R. Abugharbieh, "Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3D image segmentation," in *In Proc. MCV*. Berlin, Heidelberg: Springer, 2011, pp. 204–213. doi: 10.1007/978-3-642-18421-5_20

[271] M. Tory and T. Möller, "Human factors in visualization research," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 1, pp. 72–84, Jan. 2004. doi: 10.1109/TVCG.2004.1260759

[272] J. W. Tukey, *Exploratory data analysis*. London: Pearson, 1977, vol. 2.

[273] A. Ulinski, C. Zanbaka, Z. Wartell, P. Goolkasian, and L. F. Hodges, "Two handed selection techniques for volumetric data," in *Proc. 3DUI*. Los Alamitos: IEEE Computer Society, 2007, pp. 107–114. doi: 10.1109/3DUI.2007.340782

[274] J. C. Valentine, A. M. Aloe, and T. S. Lau, "Life after NHST: How to describe your data without "p-ing" everywhere," *Basic and Applied Social Psychology*, vol. 37, no. 5, pp. 260–273, Aug. 2015. doi: 10.1080/01973533.2015.1060240

[275] A. van Dam, "Post-wimp user interfaces," *Communication of the ACM*, vol. 40, no. 2, pp. 63—-67, Feb. 1997. doi: 10.1145/253671.253708

[276] G. R. VandenBos, Ed., *Publication Manual of the American Psychological Association*, 6th ed. Washington: American Psychological Association, 2009.

[277] M. Veit, A. Capobianco, and D. Bechmann, "Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments," in *Proc. VRST*. New York: ACM, 2009, pp. 51–58. doi: 10.1145/1643928.1643942

[278] X. Wang and P. S. Dunston, "User perspectives on mixed reality tabletop visualization for face-to-face collaborative design review," *Automation in Construction*, vol. 17, no. 4, pp. 399–412, May 2008. doi: 10.1016/j.autcon.2007.07.002

[279] X. Wang and P. S. Dunston, "Comparative effectiveness of mixed reality-based virtual environments in collaborative design," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 3, pp. 284–296, May 2011. doi: 10.1109/TSMCC.2010.2093573

[280] X. Wang, P. E. Love, M. J. Kim, and W. Wang, "Mutual awareness in collaborative design: An augmented reality integrated telepresence system," *Computers in Industry*, vol. 65, no. 2, pp. 314 – 324, Feb. 2014. doi: 10.1016/j.compind.2013.11.012

[281] X. Wang, "Augmented reality environments for the interactive exploration of 3D data," Theses, Université Paris-Saclay, Dec. 2020.

[282] X. Wang, L. Besançon, M. Ammi, and T. Isenberg, "Augmenting tactile 3D data navigation with pressure sensing," *Computer Graphics Forum*, vol. 38, no. 3, pp. 635–647, Jun. 2019. doi: 10.1111/cgf.13716

[283] X. Wang, L. Besançon, F. Guéniat, **Mickael Sereno**, M. Ammi, and T. Isenberg, "A vision of bringing immersive visualization to scientific workflows," in *Workshop CHI-IA*, Glasgow, UK, 2019.

[284] X. Wang, L. Besançon, D. Rousseau, **Mickael Sereno**, M. Ammi, and T. Isenberg, "Towards an understanding of augmented reality extensions for existing 3D data analysis tools," in *Proc. CHI*. New York: ACM, 2020, pp. 1–13. doi: 10.1145/3313831.3376657

[285] C. Ware and R. Arsenault, "Frames of reference in virtual object rotation," in *Proc. APGV*. New York: ACM, 2004, pp. 135–141. doi: 10.1145/1012551.1012576

[286] C. Ware and P. Mitchell, "Visualizing graphs in three dimensions," *ACM Transactions on Applied Perception*, vol. 5, no. 1, pp. 2:1–2:15, Jan. 2008. doi: 10.1145/1279640.1279642

[287] M. Whitlock, S. Smart, and D. A. Szafir, "Graphical perception for immersive analytics," in *Proc. VR*. Los Alamitos: IEEE, 2020, pp. 616–625. doi: 10.1109/VR46266.2020.00084

[288] A. Wiebel, F. M. Vos, D. Foerster, and H.-C. Hege, "WYSIWYP: What you see is what you pick," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2236–2244, 2012. doi: 10.1109/TVCG.2012.292

[289] L. Wilkinson, *The Grammar of Graphics*. Springer-Verlag New York, 2005. doi: 10.1007/0-387-28695-0

[290] J. Williamson, R. Murray-Smith, B. Blankertz, M. Krauledat, and K.-R. Müller, "Designing for uncertain, asymmetric control: Interaction design for brain–computer interfaces," *International Journal of Human-Computer Studies*, vol. 67, no. 10, pp. 827–841, 2009. doi: 10.1016/j.ijhcs.2009.05.009

[291] A. D. Wilson and H. Benko, "Combining multiple depth cameras and projectors for interactions on, above and between surfaces," in *Proc. UIST*. New York: ACM, 2010, pp. 273–282. doi: 10.1145/1866029.1866073

[292] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, Jun. 1998. doi: 10.1162/105474698565686

[293] H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang, "Current status, opportunities and challenges of augmented reality in education," *Computers & education*, vol. 62, pp. 41–49, Mar. 2013. doi: 10.1016/j.compedu.2012.10.024

[294] H. Xia, S. Herscher, K. Perlin, and D. Wigdor, "Spacetime: Enabling fluid individual and collaborative editing in virtual reality," in *Proc. UIST*. New York: ACM, 2018, pp. 853–866. doi: 10.1145/3242587.3242597

[295] L. Xie, A. N. Antle, and N. Motamedi, "Are tangibles more fun? comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces," in *Proc. TEI*. New York: ACM, 2008, pp. 191–198. doi: 10.1145/1347390.1347433

[296] L. Yu, K. Efstathiou, P. Isenberg, and T. Isenberg, "Efficient structure-aware selection techniques for 3D point cloud visualizations with 2DOF input," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2245–2254, Dec. 2012. doi: 10.1109/TVCG.2012.217

[297] L. Yu, K. Efstathiou, P. Isenberg, and T. Isenberg, "CAST: Effective and efficient user interaction for context-aware selection in 3D particle clouds," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 886–895, Jan. 2016. doi: 10.1109/TVCG.2015.2467202

[298] L. Yu, P. Svetachov, P. Isenberg, M. H. Everts, and T. Isenberg, "FI3D: Direct-touch interaction for the exploration of 3D scientific visualization spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1613–1622, Nov./Dec. 2010. doi: 10.1109/TVCG.2010.157

[299] N. Yuill and Y. Rogers, "Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces," *ACM Transactions on Computer-Human Interaction*, vol. 19, no. 1, pp. 1:1–1:25, May 2012. doi: 10.1145/2147783.2147784

[300] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR," in *Proc. ISMAR*. Los Alamitos: IEEE Computer Society, 2008, pp. 193–202. doi: 10.1109/ISMAR.2008.4637362

APPENDIX A

System – Additional Materials

I provide next the questions my co-authors and I designed and broadcast to domain experts to help us build the overall system (Chapter 3), and the corresponding answers from the three users who answered it..

In this survey we are interested in understanding potential data analysis meetings you hold with a small group of team members (i. e., <5-6 members). For example, these could be meetings where you analyze or discuss data you gathered during an expedition or any data related to it.

First, few questions about yourself:

1. What is your specific expertise in the larger project team?

As part of your job you regularly attend meetings with smaller groups of team members to analyze or discuss data. Please tell us about these meetings

2. Please describe these small meetings: how often and for how long you meet & who you meet with in terms of the number of meeting participants and their role/position. If you have different types of small group meetings, please describe them separately (for example: meetings with students vs. meetings with other team members).
3. Why do you meet and what would be the result of an ideal small-group data analysis meeting?
4. What is a typical outcome of the meeting and how does it differ from an ideal outcome?
5. What issues may keep you or your collaborators from reaching an ideal meeting outcome?

We want to learn more about how you work with data during these meetings:

6. In which form is data shared during these small group meetings, and how do you work with these data? E. g., spreadsheets, numerical summaries, visualizations?
7. What materials are typically created during the meeting? E. g., new visualizations, analyses, sketches, notes, etc.
8. What kind of physical material do you use during the meeting? E. g., laptops, whiteboards, flipcharts, paper, notebooks, etc.

Next, we are interested in the data you typically work with:

9. What kind of weather metrics do you typically analyse? E. g., temperature, velocity, vorticity, pressure, etc.

10. How broad and deep are your analyses during the meeting? Do you focus on just a specific subset of the data? Do you attempt to bring different aspects of the data together?

11. How do you spatially relate one kind of data, e. g., simulation forecast data, with another one, e. g., actual measurements made during the expedition?

12. How do you temporally relate one kind of data, e. g., data captured by the plane, with another one, e. g., actual measurements made during the expedition?

13. Describe the possible difficulties you have encountered during the small group meetings to relate spatial data together, to relate temporal data together, and the possible mismatch between these two dimensions.

Next, we want to learn about how you work with visualizations of data:

14. What is the role or importance of visualizations in your work?

15. When do you create visualizations of data?

16. How do you use visualizations during meetings?

17. How many visualizations are you likely to compare and look at during a meeting?

18. What difficulties do you have comparing and combining different representations of your data for analysis?

Finally, two questions about your ideal small group analysis meeting:

19. Please describe your ideal collaborative data analysis software/hardware setup that would allow you and your small team to produce the most effective analysis outcomes. Please, do not mention specific software (Excel, etc.) but describe functionalities an ideal system would have.

20. Please describe your ideal physical meeting environment (e. g., room-size, the use of traditional tools such as whiteboards or papers, the possibility to separate between sub-groups, to work alone for a time, etc.) that would allow you and your small team to produce the most effective analysis outcomes.

What is your specific expertise in the larger project team?	Please describe these small meetings: how often and for how long you meet & who you meet with in terms of the number of meeting participants and their role/position.If you have different types of small group meetings, please describe them separately (for example: meetings with students vs. meetings with other team members).
Ocean gliders and their collected data.	On average about once a week, as a mix between scientists and PhD students. It also typically depends on the work we're currently doing.
Plankton and particle image analysis to determine spatio-temmporal variability in abundance and distribution	Weekly 1on1 meetings with my boss to discuss ongoing science; weekly group meetings with the department to discuss admin and science topics; sporadic project-specific meetings to discuss projects (status updates of data analysis and discussion of results) such as M160 in depth.
Biological Carbon Pump, Plankton, Carbon Flux	Manuscript meetings with up to 12 sciences (on average) to show, compare and combine datasets

Why do you meet and what would be the result of an ideal small-group data analysis meeting?	**What is a typical outcome of the meeting and how does it differ from an ideal outcome?**
To exchange ideas, understand what things are going on and finding ways of how to proceed.	That we still have no clue.
We meet to brainstorm and update each other on on-going work. An ideal small-group data analysis meeting would divide tasks according to each group members strengths and with a combination of group and individual work, there would be more results/clearer results at the end of the meeting than at the start.	The typical meetings are often more about updating each other on the current status than about advancing it.
Usually data comparisons and manuscript preparations - understand other researchers datasets and how data can support and contribute to own results and observations	basic understanding - could be better since experts with a certain expertise are not always good in explaining results

What issues may keep you or your collaborators from reaching an ideal meeting outcome?	**In which form is data shared during these small group meetings, and how do you work with these data? E.g. spreadsheets, numerical summaries, visualizations?**
I think sometimes the level detail required to understand the problem and each of brings in different areas of expertise.	Usually in printed graphs, some times conceptually on white boards.
time constraints	Data shared during meetings is most often in the form of graphs, to be discussed. numerical data may be shared with each other during the meeting for future work but is rarely adressed during meetings.
data is often hard to grasp - a visual solution could help here	images, spreadsheets - data is then reproduced and plotted with own routines

What materials are typically created during the meeting? E.g., new visualizations, analyses, sketches, notes, etc.	What kind of physical material do you use during the meeting? E.g. laptops, whiteboards, flipcharts, paper, notebooks, etc.
Notes and sketches.	Paper (mostly), whiteboards, sometimes laptops. During covid solely laptops.
Mostly just notes.	Due to covid it's all online, so laptops, zoom/ skype, doodling tools and screensharing options within video conferencing software.
notes, spreadsheets	currently due to pandemic virtual meetings with notation function - in person meetings with whiteboards and presentations

What kind of weather metrics do you typically analyse? E.g., temperature, velocity, vorticity, pressure, etc.	**How broad and deep are your analyses during the meeting? Do you focus on just a specific subset of the data? Do you attempt to bring different aspects of the data together?**
temperature, velocity, pressure, turbulence, turbidity, density, buoyancy frequency	Usually focussed on specific subsets of the data.
Wave heights and storm frequency	Mostly superficial, focussing on subsets of data (due to time constraints).
temperature, turbidity, turbulence	usually subsets and pre-processed datasets (we are working with data from underwater cameras including thousands of images)

How do you spatially relate one kind of data, e.g., simulation forecast data, with another one, e.g., actual measurements made during the expedition?	How do you temporally relate one kind of data, e.g., data captured by the plane, with another one, e.g., actual measurements made during the expedition?
Transects in time or space.	Transects in time
I prefer side-by-sides to integrated plots of different kinds of data like that as an integration often smooths over the weaknesses of either or both data sets and can lead to risky assumptions. As I work mostly with depth profiles, pressure and/or depth measurements are useful to equi-scale different kinds of measurements.	Ideally both would have accurate time stamps.
actual measurements and distribution patterns of plankton and particles	our data has timestamps and can be temporally and spatially combined with other data.

Describe the possible difficulties you have encountered during the small group meetings to relate spatial data together, to relate temporal data together, and the possible mismatch between these two dimensions.	**What is the role or importance of visualizations in your work?**
In particular with glider data there is the issue where time and space dimensions are mingled, because the gliders go relatively slow. We can sometimes project the data relative to a water body if sufficient data exist.	Showing relationships between varies parameters
Time and depth stamps can vary if measurements were not made accurately or if instruments were not callibrated correctly. This needs to be corrected before bringing the data together.	Visualisations aid our own understanding and capability to interpret the data, but also help to present it effectively and clearly to others.
we can produce 3 dimensional (including time and space) datasets which are often hard to relate to pure 2D datasets / plane plots of hydrography	very important to present our high resolution results

When do you create visualizations of data?	How do you use visualizations during mee-tings?
	Mostly as 2D graphs.
During data processing, data analsyis, and for presentations.	Most often in teh form of shared plots which can then be discussed. Ideally these are self-explana-tory.
we try to do this as good and often as possible	presentations of distribution patterns

How many visualizations are you likely to compare and look at during a meeting?	What difficulties do you have comparing and combining different representations of your data for analysis?
Varies	Time and space mingling that occurs in glider data. Often data are also to be interpreted in 3D space, which is ambiguous in 2D.
Depends on the length of the meeting and the relevance of the visualization.	Formatting the data in the right way to plot it effectively can be time consuming.
usually this is a very common and often used method to better understand datasets	an ideal presentation of a 3d environment including hundredthousands of datapoints...

Please describe your ideal collaborative data analysis software/hardware setup that would allow you and your small team to produce the most effective analysis outcomes. Please, do not mention specific software (Excel, etc.) but describe functionalities an ideal system would have.	Please describe your ideal physical meeting environment (e.g., room-size, the use of traditional tools such as whiteboards or papers, the possibility to separate between sub-groups, to work alone for a time, etc.) that would allow you and your small team to produce the most effective analysis outcomes.
We usually write our own software tools.	Ideal physical meeting room would have soft music, a bar, and a pool.
The ideal system should be flexible, useable with different programming languages (different members of my team code in different languages), able to work with data input in different formats, allow easy data formatting within the system, and allow output of data and visualisations in different formats. Most importantly: it should be easy to learn as a system that requires a lengthy training process will be unlikely to convince people to move away from what they know. Especially seeing as nobody has the time for lengthy training, a way to start the new program and be able to work with it right away would be vital.	I think for a data meeting, a group size of more than 5 complicates things and leads to some people not opening their mouths at all during the meeting. A combination of screens and whiteboards works best in my opinion. If the meeting is with a larger group, sub-groups should be built, and if an advance is discussed that can be reached within a short time, it should be possible to pause the meeting and get there in either group or individual work, and then re-convene.
a digital twin of the observed region including different datasets	large group meeting for overall discussion and presentation, then breakout groups (size up to 10) and afterwards a synthesis in a large group again

B

APPENDIX

The Specification of Points – Additional Materials

I provide next the questionnaire my co-authors and I designed and gave to the participants of the "Specification of Points" project (see Chapter 4).

Pair ID: __ Participant ID: __

Questionnaire

Instructions

Please, cross the kind of graph shown on the right **ON THE VERTICAL LINES**. See the example on the right.

Bad — Good

General questions

At what frequency do you use AR headset?

Never — Once / Year — Monthly — Weekly — Daily

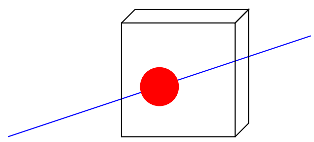
At what frequency do you use VR headset?

Never — Once / Year — Monthly — Weekly — Daily

Was the tracking of your hand accurate?

Very Bad — Very Good

Go-Go technique



Performance

How successful were you in accomplishing the task using the Go-Go technique?

Very Bad — Very Good

How was the task mentally demanding using the Go-Go technique?

Very Low — Very High

How was the task physically demanding using the Go-Go technique?

Very Low — Very High

How hurried or rushed was the pace of the task using the Go-Go technique?

Very Low — Very High

How hard did you have to work to accomplish your level of performance using the Go-Go technique?

Very Low — Very High

How irritated, stressed, and annoyed did you feel during the task using the Go-Go technique?

Very Low — Very High

Collaborative Co-Presence

I noticed my partner in the Go-Go technique session.

Strongly Disagree | Disagree | Slightly Disagree | Neutral | Slightly Agree | Agree | Strongly Agree

My partner noticed me in the Go-Go technique session.

Strongly Disagree — Strongly Agree

My partner's presence was obvious to me in the Go-Go technique session.

Strongly Disagree — Strongly Agree

My presence was obvious to my partner in the Go-Go technique session.

Strongly Disagree — Strongly Agree

My partner caught my attention in the Go-Go technique session.

Strongly Disagree — Strongly Agree

I caught my partner's attention in the Go-Go technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

Collaborative Message Understanding

My thoughts were clear to my partner in the Go-Go technique session.

○ Strongly Disagree ○ Disagree ○ Slightly Disagree ○ Neutral ○ Slightly Agree ○ Agree ○ Strongly Agree

My partner's thoughts were clear to me in the Go-Go technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

It was easy to understand my partner in the Go-Go technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

My partner found it easy to understand me in the Go-Go technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

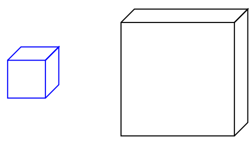
Understanding my partner was difficult in the Go-Go technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

My partner had difficulty understanding me in the Go-Go technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

WIM technique



Performance

How successful were you in accomplishing the task using the WIM technique?

Very Bad — Very Good

How was the task mentally demanding using the WIM technique?

Very Low — Very High

How was the task physically demanding using the WIM technique?

Very Low — Very High

How hurried or rushed was the pace of the task using the WIM technique?

Very Low — Very High

How hard did you have to work to accomplish your level of performance using the WIM technique?

Very Low — Very High

How irritated, stressed, and annoyed did you feel during the task using the WIM technique?

Very Low — Very High

Collaborative Co-Presence

I noticed my partner in the WIM technique session.

Strongly Disagree — Disagree — Slightly Disagree — Neutral — Slightly Agree — Agree — Strongly Agree

My partner noticed me in the WIM technique session.

Strongly Disagree — Strongly Agree

My partner's presence was obvious to me in the WIM technique session.

Strongly Disagree — Strongly Agree

My presence was obvious to my partner in the WIM technique session.

Strongly Disagree — Strongly Agree

My partner caught my attention in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

I caught my partner's attention in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

Collaborative Message Understanding

My thoughts were clear to my partner in the WIM technique session.

Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
○	○	○	○	○	○	○

My partner's thoughts were clear to me in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

It was easy to understand my partner in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

My partner found it easy to understand me in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

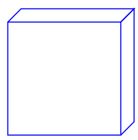
Understanding my partner was difficult in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

My partner had difficulty understanding me in the WIM technique session.

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

Manual technique



Performance

How successful were you in accomplishing the task using the Manual technique?

Very Bad ——————————— Very Good

How was the task mentally demanding using the Manual technique?

Very Low — Very High

How was the task physically demanding using the Manual technique?

Very Low — Very High

How hurried or rushed was the pace of the task using the Manual technique?

Very Low — Very High

How hard did you have to work to accomplish your level of performance using the Manual technique?

Very Low — Very High

How irritated, stressed, and annoyed did you feel during the task using the Manual technique?

Very Low — Very High

Collaborative Co-Presence

I noticed my partner in the Manual technique session.

Strongly Disagree | Disagree | Slightly Disagree | Neutral | Slightly Agree | Agree | Strongly Agree

My partner noticed me in the Manual technique session.

Strongly Disagree — Strongly Agree

My partner's presence was obvious to me in the Manual technique session.

Strongly Disagree — Strongly Agree

My presence was obvious to my partner in the Manual technique session.

Strongly Disagree — Strongly Agree

My partner caught my attention in the Manual technique session.	Strongly Disagree ○—○—○—○—○—○—○ Strongly Agree
I caught my partner's attention in the Manual technique session.	Strongly Disagree ○—○—○—○—○—○—○ Strongly Agree

Collaborative Message Understanding

	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
My thoughts were clear to my partner in the Manual technique session.	○	○	○	○	○	○	○
My partner's thoughts were clear to me in the Manual technique session.	○	○	○	○	○	○	○
It was easy to understand my partner in the Manual technique session.	○	○	○	○	○	○	○
My partner found it easy to understand me in the Manual technique session.	○	○	○	○	○	○	○
Understanding my partner was difficult in the Manual technique session.	○	○	○	○	○	○	○
My partner had difficulty understanding me in the Manual technique session.	○	○	○	○	○	○	○

Rank each technique according to your preferences regarding speed (1 is best):

__ The Go-Go technique
__ The WIM technique
__ The Manual technique

Rank each technique according to your preferences regarding accuracy (1 is best):

__ The Go-Go technique
__ The WIM technique
__ The Manual technique

Rank each technique according to your preferences regarding co-presence (1 is best):

__ The Go-Go technique
__ The WIM technique
__ The Manual technique

Rank each technique according to your preferences regarding understanding what was happening during the collaboration (1 is best):

__ The Go-Go technique
__ The WIM technique
__ The Manual technique

Rank what technique you would prefer if both people would work on different tasks simultaneously (1 is best):

__ The Go-Go technique
__ The WIM technique
__ The Manual technique

Rank each technique according to your preferences in general (1 is best):

__ The Go-Go technique
__ The WIM technique
__ The Manual technique

Any comments to add (optional)?

Appendix C

The Specification of Regions – Additional Materials

I provide next the two questionnaires my co-authors and I designed and gave to the participants of the "Specification of Regions" project (see Chapter 5).

Participant ID: __ Age: __
Gender (M/F): __ Profession: _____

Questionnaire – First Experiment

Instructions

Please, cross the kind of graph shown on the right <u>**ON THE VERTICAL LINES**</u>. See the example on the right.

Bad — Good

General questions

At what frequency do you use AR headset?

Never — Once / Year — Monthly — Weekly — Daily

At what frequency do you use VR headset?

Never — Once / Year — Monthly — Weekly — Daily

Was the tracking of the multi-touch tablet accurate?

Very Bad — Very Good

Absolute mapping

How **successful** were you in accomplishing the task using the Absolute mapping?

Very Bad — Very Good

How was the task **mentally demanding** using the Absolute mapping?

Very Low — Very High

How was the task **physically demanding** using the Absolute mapping?

Very Low — Very High

How **hurried or rushed was the pace** of the task using the Absolute mapping?

Very Low — Very High

How **hard** did you have to work to accomplish your level of performance using the Absolute mapping?

Very Low — Very High

How **irritated, stressed, and annoyed** did you feel during the task using the Absolute mapping?

Very Low — Very High

In which screen did you mostly focus while extruding 3D volumetric selection shapes using the Absolute mapping?

For the Galaxy Dataset	AR view ○	Tablet ○
For the Cylinder Dataset	AR view ○	Tablet ○
For the Spring Dataset	AR view ○	Tablet ○

Relative-Aligned mapping

How **successful** were you in accomplishing the task using the Relative-Aligned mapping?

Very Bad — Very Good

How was the task **mentally demanding** using the Relative-Aligned mapping?

Very Low — Very High

How was the task **physically demanding** using the Relative-Aligned mapping?

Very Low — Very High

How **hurried or rushed was the pace** of the task using the Relative-Aligned mapping?

Very Low | Very High

How **hard** did you have to work to accomplish your level of performance using the Relative-Aligned mapping?

Very Low | Very High

How **irritated, stressed, and annoyed** did you feel during the task using the Relative-Aligned mapping?

Very Low | Very High

In which screen did you mostly focus while extruding 3D volumetric selection shapes using the Relative-Aligned mapping?

For the Galaxy Dataset	AR view ○	Tablet ○
For the Cylinder Dataset	AR view ○	Tablet ○
For the Spring Dataset	AR view ○	Tablet ○

Relative-Full mapping

How **successful** were you in accomplishing the task using the Relative-Full mapping?

Very Bad | Very Good

How was the task **mentally demanding** using the Relative-Full mapping?

Very Low | Very High

How was the task **physically demanding** using the Relative-Full mapping?

Very Low | Very High

How **hurried or rushed was the pace** of the task using the Relative-Full mapping?

Very Low — Very High

How **hard** did you have to work to accomplish your level of performance using the Relative-Full mapping?

Very Low — Very High

How **irritated, stressed, and annoyed** did you feel during the task using the Relative-Full mapping?

Very Low — Very High

In which screen did you mostly focus while extruding 3D volumetric selection shapes using the Relative-Full mapping?

For the Galaxy Dataset	AR view ○	Tablet ○
For the Cylinder Dataset	AR view ○	Tablet ○
For the Spring Dataset	AR view ○	Tablet ○

Rank each technique according to your preferences regarding **speed** (1 is best):

__ The Absolute mapping
__ The Relative-Aligned mapping
__ The Relative-Full mapping

Rank each technique according to your preferences regarding **accuracy** (1 is best):

__ The Absolute mapping
__ The Relative-Aligned mapping
__ The Relative-Full mapping

Rank each technique according to your preferences **in general** (1 is best):

__ The Absolute mapping
__ The Relative-Aligned mapping
__ The Relative-Full mapping

Any comments to add (optional)?

Participant ID: __ Age: __
Gender (M/F): __ Profession: _____

Questionnaire – Second Experiment

Instructions

Please, cross the kind of graph shown on the right **ON THE VERTICAL LINES**. See the example on the right.

Bad — Good

General questions

At what frequency do you use AR headset?

Never — Once / Year — Monthly — Weekly — Daily

At what frequency do you use VR headset?

Never — Once / Year — Monthly — Weekly — Daily

Was the tracking of the multi-touch tablet accurate?

Very Bad — Very Good

3D AR view

How **successful** were you in accomplishing the task using the 3D AR view?

Very Bad — Very Good

How was the task **mentally demanding** using the 3D AR view?

Very Low — Very High

How was the task **physically demanding** using the 3D AR view?

Very Low — Very High

How **hurried or rushed was the pace** of the task using the 3D AR view?

Very Low — Very High

o How **hard** did you have to work to accomplish your level of performance using the 3D AR view?

Very Low — Very High

How **irritated, stressed, and annoyed** did you feel during the task using the 3D AR view?

Very Low — Very High

In which case did you look at the 3D AR view during the **extrusion**?

In which case did you look at the tablet view during the **extrusion**?

On which screen did you mostly focus while placing the tablet's plane before drawing and extruding 2D shapes in the "3D AR" condition?

3D AR view ○ Tablet ○ Both about equally ○

How much was the 3D view useful?

Not at All ○ ○ ○ ○ ○ ○ ○ Very Much

2D view

How **successful** were you in accomplishing the task using the 2D view?

Very Bad — Very Good

How was the task **mentally demanding** using the 2D view?

Very Low — Very High

How was the task **physically demanding** using the 2D view?

Very Low — Very High

How **hurried or rushed was the pace** of the task using the 2D view?

Very Low — Very High

How **hard** did you have to work to accomplish your level of performance using the 2D view?

Very Low — Very High

How **irritated, stressed, and annoyed** did you feel during the task using the 2D view?

Very Low — Very High

In which case did you look at the AR view during the **extrusion**?

In which case did you look at the tablet view during the **extrusion**?

On which screen did you mostly focus while placing the tablet's plane before drawing and extruding 2D shapes in the "2D" condition?

External Screen ○ | Tablet ○ | Both about equally ○

How much was the 2D external screen useful?

Not at All ○ ○ ○ ○ ○ ○ ○ Very Much

Rank each technique according to your preferences for **the galaxy dataset** (1 is best):

__ 2D External Screen
__ 3D AR

Rank each technique according to your preferences for **the cylinder dataset** (1 is best):

__ 2D External Screen
__ 3D AR

Rank each technique according to your preferences for **the spring dataset** (1 is best):

__ 2D External Screen
__ 3D AR

Any comments to add (optional)?

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Exploration et Discussion Collaborative des Données Grâce à la Réalité Augmentée

Mots clés: Visualisation Collaborative Analytique, Réalité Augmentée, Visualisation Scientifique

Résumé: J'étudie les avantages et limitations des casques de réalité augmentée (RA) pour l'exploration collaborative de données 3D. Avant d'entamer mes travaux, je voyais dans ces casques des avantages liés à leurs capacités immersives : ils fusionnent les espaces interactifs, de visualisation, de collaboration et physique des utilisateurs. Plusieurs collaborateurs peuvent voir et interagir directement avec des visuels 3D ancrés dans le monde réel.

Ces casques reposent sur une vision stéréoscopique 3D qui fournit une perception de profondeur accrue par rapport aux écrans 2D, aidant les utilisateurs à mieux comprendre leurs données 3D. Laissant les utilisateurs se voir les uns les autres, il est possible de transitionner sans effort d'une phase de discussion à une phase d'exploration. Ces casques permettent aux utilisateurs d'interagir au sein de l'espace de travail de manière directe, rapide et intuitive en 3D, et donnent des indices sur les intentions d'une personne aux autres. Par exemple, le fait de déplacer un objet en le saisissant est un indice fort sur les intentions de cette personne. Enfin, en n'occultant pas le monde réel, les outils habituels mais importants tels que les postes de travail restent facilement accessibles dans cet environnement.

Cela étant, et bien qu'ils soient étudiés depuis des décennies, la puissance de calcul de ces casques avant la récente sortie de l'HoloLens en 2016 n'était pas suffisante pour une exploration efficace de données 3D telles que des données océaniques. De plus, les chercheurs précédemment étaient plus intéressés par *comment rendre la RA possible* que par *comment utiliser la RA*. Malgré toutes leurs qualités, il y a donc peu de travaux qui traitent de l'exploration de jeux de données 3D. Finalement, les casques de RA ne fournissent pas d'entrées 2D qui sont couramment utilisées avec les outils d'exploration actuels tels que ParaView et les logiciels de CAO, avec lesquels entre autres scientifiques et ingénieurs sont déjà efficaces.

Je théorise donc dans cette thèse les situations où ces casques sont préférables. Ils semblent préférables lorsque l'objectif est de partager des idées, d'explorer des modèles ensemble et lorsque les outils d'exploration peuvent être minimaux par rapport à ce que les postes de travail fournissent, et où la plupart des travaux et simulations préalables peuvent être effectués à l'avance. J'associe alors les casques de RA à des tablettes tactiles. J'utilise ces casques pour fusionner la visualisation, certaines interactions 3D et les espaces de collaboration dans l'espace physique des utilisateurs, et les tablettes pour la saisie 2D et l'interface utilisateur graphique habituelle que la plupart des logiciels proposent. J'étudie ensuite l'interaction de bas niveau nécessaire à l'exploration de données. Cela concerne la sélection de points et de régions dans des données 3D à l'aide de ce système hybride. Comme cette thèse vise à étudier les casques de RA dans des environnements collaboratifs, j'étudie également leurs capacités à adapter le visuel à chaque collaborateur pour un objet 3D ancré donné, similairement au "What-You-See-Is-What-I-See" relaxé qui permet par exemple à plusieurs utilisateurs de voir et modifier simultanément différentes parties d'un document partagé. Enfin, j'étudie en ce moment l'utilisation de mon système pour l'exploration collaborative en 3D des jeux de données océaniques sur lesquels travaillent mes collaborateurs du Helmholtz-Zentrum Geesthacht en Allemagne.

Pour résumer, cette thèse fournit un état de l'art de la RA à des fins collaboratifs, fournit un aperçu de l'impact de la directivité de l'interaction 3D sur l'exploration de donnée 3D, et donne aux concepteurs un aperçu de l'utilisation de la RA pour l'exploration collaborative de données scientifique 2D et 3D, en mettant l'accent sur le domaine océanographique.

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title: Collaborative Data Exploration and Discussion Supported by Augmented Reality

Keywords: Collaborative Immersive Analytics, Augmented Reality, Scientific Visualization

Abstract: I studied the benefits and limitations of Augmented Reality (AR) Head-Mounted Displays (AR-HMDs) for collaborative 3D data exploration. Prior to conducting any projects, I saw in AR-HMDs benefits concerning their immersive features: AR-HMDs merge the interactive, visualization, collaborative, and users' physical spaces together. Multiple collaborators can then see and interact directly with 3D visuals anchored within the users' physical space.

AR-HMDs usually rely on stereoscopic 3D displays which provide additional depth cues compared to 2D screens, supporting users at understanding 3D datasets better. As AR-HMDs allow users to see each other within the workspace, seamless switches between discussion and exploration phases are possible. Interacting within those visualizations allows for fast and intuitive 3D direct interactions, which yields cues about one's intentions to others, e. g., moving an object by grabbing it is a strong cue about what a person intends to do with that object. Those cues are important for everyone to understand what is currently going on. Finally, by not occluding the users' physical space, usual but important tools such as billboards and workstations performing simulations are still easily accessible within this environment without taking off the headsets.

That being said, and while AR-HMDs are being studied for decades, their computing power before the recent release of the HoloLens in 2016 was not enough for an efficient exploration of 3D data such as ocean datasets that my collaborators at Helmholtz-Zentrum Geesthacht, Germany, are working on. Moreover, previous researchers were more interested in *how to make AR possible* as opposed to *how to use AR*. Then, despite all those qualities one may think prior to working with AR-HMDs, there were almost no work that discusses the exploration of such 3D datasets. Moreover AR-HMDs are not suitable for 2D input which are commonly used within usual exploratory tools such as ParaView or CAD software, with which users, such as scientists and engineers, are already efficient. I then theorize in what situations are AR-HMDs preferable. They seem preferable when the purpose is to share insights with multiple collaborators and to explore patterns together, and where exploratory tools can be minimal compared to what workstations provide as most of the prior work and simulations can be done before hand. I am thus combining AR-HMDs with multi-touch tablets, where I use AR-HMDs to merge the visualizations, some 3D interactions, and the collaborative spaces within the users' physical space, and I use the tablets for 2D input and usual Graphical User Interfaces that most software provides (e. g., buttons and menus).

I then studied low-level interactions necessary for data exploration which concern the selection of points and regions inside datasets using this new hybrid system. As this PhD aims at studying AR-HMDs within collaborative environments, I also studied their capacities to adapt the visual to each collaborator for a given anchored 3D object. This is similar to the relaxed "What-You-See-Is-What-I-See" that allows, e. g., multiple users to see different parts of a shared document that remote users can edit simultaneously. Finally, I am currently (i. e., is not finished by the time I am writing this PhD) studying the use of this new system for the collaborative 3D data exploration of ocean datasets.

This PhD provides a state of the art of AR used within collaborative environments. It also gives insights about the impacts of 3D interaction directness for 3D data exploration. This PhD finally gives designers insights about the use of AR for collaborative scientific data exploration, with a focus on oceanography.