Supporting Collaboration in Large Interactive Spaces

Cédric Fleury

Habilitation à diriger des recherches

université PARIS-SACLAY

Supporting Collaboration in Large Interactive Spaces

Favoriser la collaboration dans les grands espaces interactifs

Habilitation à diriger des recherches de l'Université Paris-Saclay

présentée et soutenue à Gif-sur-Yvette, le 14 février 2024, par

Cédric FLEURY

Composition du jury

Anatole LÉCUYER Directeur de recherche, Inria Rennes, France	Rapporteur
Carman NEUSTAEDTER Professeur, Simon Fraser University, Canada	Rapporteur
Anthony STEED Professeur, University College London, Royaume-Uni	Rapporteur
Laurence NIGAY Professeur, Université Grenoble Alpes, France	Examinatrice
Michel BEAUDOUIN-LAFON Professeur, Université Paris-Saclay, France	Examinateur
Anastasia BEZERIANOS Professeur, Université Paris-Saclay, France	Présidente

ACKNOWLEDGMENTS

First, I want to thank all the members of my Habilitation committee. I am extremely honored that they have accepted to be part of this committee. I would especially like to thank the three reviewers: Anatole Lécuyer, Carman Neustaedter and Anthony Steed for the time they spent reading the manuscript, and for their detailed and insightful reports. Special thanks to Carman for waking up so early to attend the defense because of the time difference. I would also like to thank the other members: Laurence Nigay for her challenging questions, Anastasia Bezerianos for accepting to chair the defense, and Michel Beaudouin-Lafon for mentoring me through the process, but also for his support and kindness during all these years.

Next, I want to deeply thank all my colleagues at Université Paris-Saclay, where I carried out most of the research that led to this Habiliation. Special thanks to Wendy Mackay for welcoming me as part of the InSitu and ExSitu teams, and for making these teams so lively and dynamic. I had the pleasure not only to share my office, but also to co-supervise a PhD student with Theophanis Tsandilas. I would like to thank Patrick Bourdot for inviting me to collaborate with the VENISE team and for co-supervising two PhD students with me. Michel, Wendy, Theophanis and Patrick definitely shaped the way I conducted my research. Thanks also to the past and current members of the InSitu, ExSitu and VENISE teams with whom I had the chance to work, and especially to Caroline Appert, Baptiste Caramiaux, Olivier Chapuis, Sarah Fdili Alaoui, Nicolas Ferey, Stéphane Huot, Huyen Nguyen, and Jeanne Vézien. I am extremely grateful to the engineers, Olivier Gladin, Nicolas Ladevèze and Alexandre Kabil, for their technical support on many projects and for helping me set up the technical system for my defense. As teaching was an important part of my work over these years, I would like to thank my colleagues at the Polytech Paris-Saclay engineering school for their dedicated involvement.

Then, I want to thank my new colleagues at IMT Atlantique with whom I have collaborated on research and teaching in the recent years. Special thanks to Thierry Duval, with whom I started my academic career, for his warm welcome at IMT Atlantique and his support. Thanks to Gilles Coppin, Mathieu Chollet and Cédric Dumas for co-supervising PhD students with me. I also thank Guillaume Moreau, Etienne Peillard and all the members of the INUIT team for our interesting discussions about research and many other topics.

Beyond these two institutions, I would like to thank my colleagues in France and abroad: Bruno Arnaldi and Valérie Gouranton with whom I started my career, and Tat-Jen Cham, Morten Fjeld, Henry Fuchs, Khanh-Duy Le, Tiberiu Popa, Bruce Thomas, and James Walsh with whom I collaborated on various projects.

Most importantly, I would like to warmly thank the PhD students I have had the pleasure of supervising over the years: Ignacio Avellino, Yujiro Okuya, Yiran Zhang, Arthur Fages, Thomas Rinnert and Aurélien Léchappé.

Finally, I want to thank my parents and my sister Delphine for their support all these years and for attending my defense. I would also like to thank my partner Delphine for her patience and unconditional support on a daily basis, and my daughter Clémence for distracting me from writing this manuscript.

CONTENTS

1 INTRODUCTION

The intensive use of digital technology leads to an exponential increase in the quantity, complexity and variety of data produced by our society. Many fields, such as science, industry, business or health, generate more and more data every year. For example, the telescope used for the Legacy Survey of Space and Time (LSST) will capture about a thousand 3.2-gigapixel images of the sky every night for ten years. The total amount of data collected over the ten years will reach about 60 petabytes of raw data [Obs22]. In healthcare, the amount of data collected each year is growing exponentially due to the accumulation of a wide variety of digital information including personal medical records, radiology images, clinical trial data and human genetics. New forms of data, such as 3D imaging, genomic sequences or biometric sensor recordings, further accentuate this trend. For instance, US healthcare data was estimated to be about 150 exabytes in 2011, but is now probably up to several zettabytes [RR14; Gui+16]. As a last example, the computer-aided design (CAD) model of a Boeing 777 is composed of 6 million parts and connectors, 350 million triangular faces to display, and 12 gigabytes of geometry data to store [XZ15].

This digital data provides unique opportunities to support scientific discovery, improve industrial processes, help decision-making, foster creation or encourage learning. However, enabling humans to handle and explore such a large quantity and variety of data is more than ever a major challenge. Although research work, especially in the domain of artificial intelligence (AI), focuses on the automatic processing of large amounts of such data, it is mandatory to keep humans involved in the analysis process. Users must keep control over how the data is processed and need to be able to understand the results provided by computers.

In this context, the need for computer-mediated collaboration has never been so high. On the one hand, processing or exploring complex data sets usually requires multiple collaborators to combine their expertise or share their knowledge. On the other hand, the current societal context, including pandemic situations and the green transition, requires groups of users to work remotely while intensively using digital tools. The COVID-19 pandemic has significantly increased our reliance on computer-mediated collaboration tools (+44%, according to Gartner [Rim22]), and some experts predict that this situation will largely persist after the pandemic [Blo22]. As a consequence, supporting collaboration among co-located and remote users when analyzing large amounts of data is also a key challenge.

My research investigates how large interactive spaces, such as wall-sized displays [And+11; Bea+12], immersive virtual reality systems [Cru+92] or augmented reality spaces [BK02], can foster collaboration on complex data sets. The ability of such systems to display large amounts of information, potentially in 3D, and to spatially organize that information offers new alternatives for interacting with digital content. These technologies start to be used in a wide range of application domains, such as scientific data analysis [**Fle+12**; PBC17; Kam+18], review of computer-aided design (CAD) models [Bou+10], monitoring of processes in control rooms [Sch+12], scheduling of complex events [Liu15] or pathology diagnosis [Rud+16]. They also

provide large collaborative spaces for interaction and communication between multiple users, which can be useful for brainstorming and combining ideas during product design [**Oku+20**], crisis management [PBC18] or creative work [Str+99].

Nevertheless, large interactive spaces require us to rethink how users interact with computers and collaborate through them. In particular, they offer the opportunity to develop new forms of interaction and new solutions to foster collaboration by taking advantage of the large visualization space and physical space available to users. My research contributes to this evolution in four ways:

1. I propose new interaction paradigms to handle large displays and make use of the large physical space surrounding users. These paradigms contrast with traditional mouse-keyboard or touch-based interfaces by allowing multiple users to interact simultaneously, while moving freely within the system.

2. I explore collaborative interaction among multiple users within a shared interactive space. Such interaction provides users with dedicated features to enrich collaborative activities, while managing conflicts and interference that may arise when interacting in the same space or with the same content.

3. I create interactive systems that connect remote users across heterogeneous interactive spaces. These systems rely on specific technical solutions to synchronize complex data and transmit communication cues, including spatialized audio and 3D user representations.

4. I investigate video-mediated communication among remote collaborators in large interactive spaces. Such spaces involve specific constraints to deploy telepresence capabilities, but also offer unique possibilities to enhance remote users' perception and non-verbal communication.

1.1 TERMINOLOGY AND SCOPE

My research work is at the crossroads of Human-Computer Interaction (HCI), Virtual and Augmented Reality (VR/AR) and Computer-Supported Cooperative Work (CSCW). I study how humans communicate with computers, but also how humans communicate among them through computers. Although part of my research focuses on collaboration through immersive technologies including virtual and augmented reality, my contributions are broader in the HCI domain including work on interaction design and video-mediated communication.

Next, I define the key terminology used in this manuscript and hence clarify the scope of my research work:

Large interactive space. I choose this term to encompass both immersive and non-immersive systems that provide users with a vast physical space for interaction. I further define large interactive spaces and give examples in Section 2.1.

Mixed reality. Among the various terminologies that unify virtual and augmented reality, I opt to use the term "mixed reality" as originally defined by Milgram et al. [Mil+95]. The mixed reality continuum, recently revised by Skarbez et al. [SSW21], categorizes augmented reality (AR) and virtual reality (VR) systems according to the proportion of the real and virtual world perceived by users. I employ this term equivalently to "extended reality" in the manuscript.

Physical workspace. This workspace refers to the physical space where users interact within the system. It is typically defined by the available space in front of the displays or by the limits of the tracking system.

Virtual workspace. In mixed reality systems, the physical workspace is mapped onto a specific region of the virtual environment, referred to as the virtual workspace. This virtual workspace is the virtual counterpart of the physical workspace: users can travel everywhere in this virtual workspace by walking in the physical workspace.

Telepresence. I use the term "telepresence" to designate video-mediated communication systems that enhance users' feeling of being present in the same space. These systems usually rely on large or immersive displays and advanced sound synthesis. They contrast with conventional videoconferencing systems, using standard computers or mobile devices equipped with a single camera and a relatively small screen. Examples of such telepresence systems are presented in Section 2.3.1.

Collaborative virtual environment. I define collaborative virtual environments (CVEs) as distributed systems that enable remote users to meet in a shared mixed reality environment. This definition encompasses systems using either virtual reality or augmented reality technologies. It corresponds to what is sometimes referred to as social virtual reality (SVR). Section 2.3.2 describes collaborative virtual environments in more depth.

Awareness. When referring to awareness in this manuscript, I specifically focus on the awareness among collaborators. It includes all the perceptual cues that help users understand the position, actions and intentions of their collaborators, along with the information they aim to communicate.

User representation. I employ this terminology to refer to embodied representations that enable users to perceive remote collaborators in an interactive space. These representations can use either live video or 3D avatars to provide users with appropriate awareness of each other, depending on the context.

1.2 CONTEXT AND INSPIRATION

This manuscript presents my research activities since 2012. However, before that, my PhD work already focused on remote collaboration across immersive VR systems. While one part concentrated on distributed architecture for synchronizing virtual environments, another part explored collaborative interaction between users with heterogeneous devices. This work was part of a national project, named *Collaviz*, that provided valuable remote collaboration scenarios for scientific data analysis.

During my post-doctoral position (2012-2013), I explored 3D head reconstruction for telepresence at the *BeingThere Centre* [FSB14]. The ambition of this joint international laboratory was to connect remote rooms through stereoscopic wall-sized displays, which thus become glass windows between the rooms. Although my work focused on technological aspects, it enabled me to gain a better understanding of non-verbal communication, including facial expressions and eye gaze.

My position as an associate professor at Université Paris-Saclay (2013-2021) provided me with the opportunity to explore these research themes in various projects.

First, I collaborated on the DIGISCOPE project[1], which created a network of ten interconnected platforms for interactive visualization of large datasets. This project was a unique occasion to design and test collaborative systems connecting remote platforms. It also provided various application domains, including scientific research, computer-aided design, decision support systems, and education. Second, I participated in projects involving engineers from the VR center of the PSA automotive company (now Stellantis[2]). It was a unique opportunity to access real collaborative design scenarios and interview engineers on their current practices.

In 2021, I joined IMT Atlantique as an associate professor. I continue to investigate collaboration in large interactive spaces. Part of this work is conducted in the context of CONTINUUM[3], a follow up to the DIGISCOPE project at a national level.

1.3 RESEARCH METHODOLOGY

The contributions described in this manuscript are supported by a wide range of empirical results. In particular, my work applied several HCI methods to investigate research questions and analyze findings. I learned this HCI methodology all along my research career, but I considerably improved my expertise during my position at Université Paris-Saclay. Beyond the traditional controlled experiments used to evaluate the proposed techniques or systems, I integrated user-centered methods involving potential users in the design of these techniques or systems whenever possible. These methods include participatory design, interviews and qualitative observations with low or high-fidelity prototypes. For example, we interviewed engineers from the PSA automotive company for the work on computer-aided design. We conducted qualitative observations and interviews with civil engineering students for the design of *ShapeCompare* (Section 3.1.2). We ran preliminary observations using low-fidelity prototypes for the design of *CamRay* (Section 4.2.1.2). When controlled experiments were not suitable to assess our solutions, we conducted user studies on more open-ended tasks to observe how users appropriate tools and compare their different strategies, as for evaluating *ARgus* (Section 4.2.3).

Although my training and expertise are more related to computer science, I have endeavored to ground my work with foundations in psychology and sociology. In particular, we designed collaborative systems based on the concept of *grounding in communication* proposed by Clark and Brennan [CB91]. This concept refers to the communication process required to build a *common ground* between users, including mutual knowledge, beliefs, and assumptions about the collaborative situations [CM81; CSB83]. Collaborative systems should either provide support to enhance the establishment of common ground, or compensate for communication cues that are lost in computer-mediated communication. In addition, our contributions on collaborative design build on specific previous work that studied collaborative design practices. We drew inspiration from the work of De Bono [De 67] on *lateral thinking*, and Stempfle and Badke-Schaub [SB02] on the thinking processes of design teams. Détienne's work [Déto6] also provided us with valuable insights on managing task interdependencies and multiple perspectives in design.

1 http://www.digiscope.fr/en/
2 https://www.stellantis.com/en/
3 https://www.lri.fr/~mbl/CONTINUUM/en/

1.4 MANUSCRIPT OVERVIEW

This section describes the remaining chapters of the manuscript, and acknowledges the students and collaborators who contributed to this research work. After a brief chapter on related work, I divide my contributions into two parts: the first is related to interaction and co-located collaboration in large interactive spaces, while the second concerns remote collaboration across such spaces. For each part, I provide representative papers that illustrate my contributions to the related topic. These papers are attached at the end of the manuscript. Publications I have co-authored are highlighted in [**Bold**] in this manuscript.

Related work on large interactive spaces (Chapter 2). This chapter defines large interactive spaces and illustrates how users interact and collaborate in such systems. It also describes previous work that explores remote collaboration across large interactive spaces, including telepresence and collaborative virtual environments.

From interaction to collaboration in a shared interactive space (Chapter 3). Allowing each user to interact in large interactive spaces is a necessary step to support collaboration. This chapter introduces new interaction paradigms that provide users with the ability to master the unusual characteristics of these systems. Beyond individual interaction, it investigates how such systems can foster co-located collaboration by providing appropriate collaborative interaction among users.

The first section focuses on the large visualization space, studying how users can interact with 3D virtual objects on a wall-sized display and collaboratively explore a large number of these objects. The work on 3D interaction [**LF16**] was part of the master's thesis of J.-B. Louvet. The work on collaborative exploration [**Oku+20**] was part of the PhD thesis of Y. Okuya, co-supervised with P. Bourdot (CNRS senior researcher), and involved O. Gladin and N. Ladévèze (both engineers).

The second section concentrates on interaction in a 3D space, investigating 3D object deformation with haptic interaction and collaborative sketching in augmented reality. The first part on 3D object deformation [**Oku+18a**; **Oku+21**] was carried out during the PhD of Y. Okuya, co-supervised with P. Bourdot, in collaboration with N. Ladévèze. The second part on collaborative sketching [**FFT23**] was part of the PhD thesis of A. Fages, co-supervised with T. Tsandilas (Inria researcher).

The third section presents multiple navigation techniques that leverage the large physical space to maximize physical displacements and allow tangible interaction. It also investigates how these techniques can be extended to collaborative navigation involving two co-located users. All these contributions [**Zha+19**; **Zha+20**; **Zha+21**; **Zha+22**] were based on the PhD work of Y. Zhang, co-supervised with P. Bourdot, and involved S.-T. Ho (master's intern), T.T.H. Nguyen (associate professor at Université Paris-Saclay), and N. Ladévèze (engineer).

Representative papers:

- Interaction: *J.-B. Louvet, C. Fleury (2016). Combining Bimanual Interaction and Teleportation for 3D Manipulation on Multi-Touch Wall-sized Displays. Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST), 8 pages.*

- Collaboration: *Y. Okuya, O. Gladin, N. Ladévèze, C. Fleury, P. Bourdot (2020). Investigating Collaborative Exploration of Design Alternatives on a Wall-Sized Display. Proceedings of the ACM conference on Human Factors in Computing Systems (CHI), 12 pages.*
- Navigation: *Y. Zhang, N. Ladévèze, H. Nguyen, C. Fleury, P. Bourdot (2020). Virtual Navigation considering User Workspace: Automatic and Manual Positioning before Teleportation. Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST), 10 pages.*
- Collaborative navigation: *Y. Zhang, T.T.H. Nguyen, N. Ladévèze, C. Fleury, P. Bourdot (2022). Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs, Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR), 9 pages.*

Collaboration and awareness across remote spaces (Chapter 4). Remote collaboration is becoming crucial due to major changes in our society, including new work organization and the green transition. This chapter presents how large interactive spaces can foster collaboration among remote users. In particular, such collaboration requires technical solutions to connect heterogeneous platforms, as well as appropriate awareness and communication cues among the remote collaborators.

The first section focuses on the technical aspects of connecting remote users across heterogeneous systems. A distributed architecture synchronizing computer-aided design data [**Oku+18b**; **Oku+21**] was created during the PhD of Y. Okuya, co-supervised with P. Bourdot, in collaboration with N. Ladévèze and O. Gladin. As part of the DIGISCOPE project, we designed a system to explore 3D audio mappings between remote platforms [**Fyf+18**] with L. Fyfe (engineer), O. Gladin, and M. Beaudouin-Lafon (professor at Univ. Paris-Saclay). During my post-doc, I proposed a method to reconstruct remote users' head in 3D [**Fle+14**], supervised by H. Fuchs (professor at Univ. of North Carolina) and T.J. Cham (associate professor at NTU Singapore), and with the collaboration of T. Popa (post-doc at ETH Zurich).

The second section concentrates on video-mediated communication across large interactive spaces. The work on deictic gestures perception [**AFB15**] and telepresence across wall-sized displays [**Ave+17**] was part of the PhD work of I. Avellino, co-supervised with M. Beaudouin-Lafon, and involved W. Mackay (Inria senior researcher). The work about one-to-many telepresence [**Le+19**] was a collaboration with K.-D. Le (PhD student at Chalmers Univ. of Technology), I. Avellino, M. Fjeld (professor at Chalmers Univ. of Technology), and A. Kunz (professor at ETH Zurich). The work on multi-view collaboration with AR users [**FFT22b**] was part of the PhD thesis of A. Fages, co-supervised with T. Tsandilas.

Representative papers:

- Perceptual study: *I. Avellino, C. Fleury, M. Beaudouin-Lafon (2015). Accuracy of deictic gestures to support telepresence on wall-sized displays. Proceedings of the ACM conference on Human Factors in Computing Systems (CHI), 4 pages.*
- One-to-one telepresence: *I. Avellino, C. Fleury, W. Mackay, M. Beaudouin-Lafon (2017). CamRay: Camera Arrays Support Remote Collaboration on Wall- Sized Displays. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 12 pages.*
- One-to-many telepresence: *K.-D. Le, I. Avellino, C. Fleury, M. Fjeld, A. Kunz (2019). GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration. Proc. of the IFIP TC13 Conference on Human-Computer Interaction (INTERACT), 21 pages.*

- Telepresence with AR users: *A. Fages, C. Fleury, T. Tsandilas (2022). Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users, Proc. of the ACM Conference on Computer-Supported Cooperative Work (CSCW), 27 pages.*

Future perspectives and closing remarks (Chapter 5). This chapter presents future directions for my research and concludes with some remarks regarding the future of collaboration in large interactive spaces.

2

RELATED WORK ON LARGE INTERACTIVE SPACES

The main objective of this chapter is to situate the context of my research work. As a consequence, it is not intended to be an exhaustive overview of the related work, but rather to provide examples that define and illustrate the foundations of my work. The first section proposes a definition of large interactive spaces and describes the different systems used in the work presented in this manuscript. The second section introduces some previous work that illustrates how users interact and collaborate in such systems. Finally, the last section presents related work addressing collaboration across remote interactive systems.

2.1 LARGE INTERACTIVE SPACES

Large interactive spaces include interactive systems that provide users with a large physical space, enabling them to move and interact in 3D. This large space can often accommodate multiple users, making these systems particularly well-suited to collaboration. By definition, they contrast with personal computers and mobile devices, which are generally designed for single users who remain static relative to the display. The key characteristics of large interactive spaces include:

- A **large visualization space**, which displays digital content in a large portion of the available 3D space, such as on a full wall or all around the users.
- A **large physical space**, which allows users to physically move around to explore the digital content and perform 3D interaction such as pointing, grabbing or manipulating virtual objects.
- Various **interaction devices**, which enable users to interact with the digital content from multiple locations in the 3D space. These devices range from touch devices to VR controllers.
- A **tracking system**, which detects the positions of both the users and the interaction devices in the 3D space.

With this definition, I decide to group immersive and non-immersive systems together, as they share many similarities in terms of interaction and collaboration. While hybrid systems do exist, non-immersive systems typically encompass 2D wall-sized displays and large digital tabletops, whereas immersive systems include all mixed reality devices. All these systems empower users to visualize and manipulate large volumes of complex data. They support physical navigation and 3D interaction to explore such data. Additionally, they can accommodate multiple users within the same interactive space and connect remote users.

In this section, I classify large interactive spaces into three categories that illustrate the wide range of systems available: (i) wall-sized displays, (ii) immersive virtual reality systems and (iii) augmented reality spaces. These three categories cover all the devices used in the research work described in this manuscript.

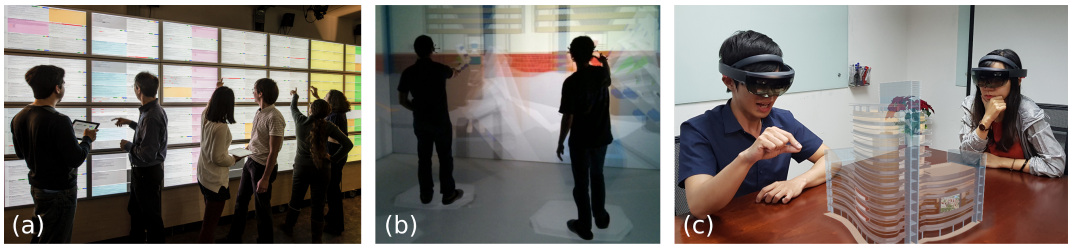

Figure 2.1: Large interactive spaces: (a) users preparing the CHI 2013 conference program on the WILD wall-sized display (© Inria), (b) users exploring a virtual factory in the EVE immersive system (© CNRS - VENISE), and (c) users reviewing a building 3D mockup in an augmented reality space (© Hoshinim - CC-BY-SA 4.0).

2.1.1 *Wall-sized displays*

Wall-sized displays typically consist of a large, ultra high-resolution display, as described by Andrews et al. [And+11] and Beaudouin-Lafon et al. [Bea+12]. They are often built using a large number of high-resolution screens, although a few use multiple projectors. Some previous studies have demonstrated their ability to improve performance in various tasks when compared to standard desktop computers. Czerwinski et al. [Cze+03] showed that larger screen space increases productivity and user satisfaction in everyday computer use, particularly when working with multiple windows. Bi and Balakrishnan [BB09] and Grudin [Gru01] corroborated these findings by showing that peripheral awareness facilitates tasks involving multiple windows. For sense-making tasks, Andrews et al. [AEN10] demonstrated that large screens enable users to employ a distributed cognitive process, which improves performance by allowing users to associate content meaning with spatial locations. Lischke et al. [Lis+15] proved the benefits of wall-sized displayed for information search tasks. Finally, other studies have shown that physical navigation provided by wall-sized displays improves spatial memory [JSH19] and user performance in visual search [BNB07] and data manipulation [Liu+14].

I used two wall-sized displays in my work: WILD (Figure 2.1-a) and WILDER. The WILD (Wall-sized Interaction with Large Datasets) platform was composed of an 8×4 grid of 30" *Apple Cinema Displays* screens at the time we conducted the work presented in this manuscript. It measured $5.5m \times 1.8m$, with a resolution of 20480×6400 pixels. More recently, the WILD platform has been upgraded with $8K$ screens increasing the resolution to over 1 Gigapixel (61441×17240). It is controlled by a cluster of 16 computers, each managing two screens. A *VICON* infrared tracking system[1] can track the position and orientation of the users and interaction devices in front of the display.

The WILDER platform consists of a 15×5 grid of 21.6" *Planar* screens[2] with $3mm$ ultra-thin bezels. It measures $5.9m \times 2m$, with a resolution of 14400×4800 pixels. It is controlled by a cluster of 10 computers, each managing a row of 8 or 7 screens. The platform also integrates a *VICON* tracking system, along with a *PQLabs* infrared frame[3] surrounding the display to provide multitouch capability.

1 https://www.vicon.com/
2 https://www.planar.com/
3 https://www.pqlabs.com/

2.1.2 *Immersive virtual reality systems*

Virtual reality (VR) systems have been used for several decades to immerse users in virtual environments. They provide users with multi-sensory feedback, enabling them to perceive the virtual environment through multiple sensory cues, including visual, audio and haptic cues. This virtual environment is a digital world simulated by computers, which can present a large variety of information to the users. A more complete definition of virtual reality can be found in Fuchs et al. [FMG11]. When focusing on visual cues, the key characteristics of immersive VR systems include a wide field of regard, a wide field of view, high-resolution displays, and depth perception. The field of regard corresponds to the amount of the physical space surrounding users in which images are displayed, while the field of view refers to the viewing angle instantaneously perceived by users. Depth cues are essential for immersion, as they give users the feeling of being present in the 3D space of the virtual environment. In current VR systems, depth perception mainly relies on 3D stereoscopic vision and motion parallax. Motion parallax refers to the depth information conveyed by the relative movements of virtual objects in response to changes in the viewer's position. Providing users with the ability to move in the physical space of immersive systems is crucial for perceiving motion parallax.

A wide range of devices has been used to immerse users in virtual environments, including head-mounted displays, CAVE-like systems and various types of stereoscopic screens. The *CAVE* (Cave Automatic Virtual Environment), introduced in 1992 by Cruz-Neira et al. [Cru+92], was the first system using multiple projected screens surrounding users to increase the field of regard. Since then, the technology has improved considerably, and this type of platform has been deployed in many research laboratories and large companies, such as automotive and aerospace manufacturing companies. At the same time, head-mounted displays have also evolved dramatically, and a wide range of high-quality devices is now available to the general public.

In my research work, I used a CAVE-like system, named Eve (Figure 2.1-b), and head-mounted displays, including *HTC Vive* and *HTC Vive Pro Eye* headsets[4]. The Eve (Environnement Virtuel Evolutif) system is composed of four back-projected stereoscopic screens, measuring $4.8m \times 2.7m$ (front & floor) and $2.7m \times 2.7m$ (left & right). Each screen has a resolution of 1920×1080 pixels. The projectors are able to achieve both active stereovision and polarization image multiplexing, enabling two users to have their own stereoscopic view of the virtual environment. The system can thus support collaborative interaction between these two users, while providing both with correct motion parallax. Applications are executed on a server that distributes the rendering to four computers, each connected to a projector. An *ART* infrared tracking system[5] is used to capture the position and orientation of the users and interaction devices in the system. A *Scale-One* haptic device from Haption[6] is also available to interact with the virtual environment. This device consists of a *Virtuose* haptic arm[7] mounted on a 4-degree-of-freedom carrier, allowing users to interact anywhere in the physical space despite the limited range of the haptic arm.

4 https://www.vive.com/
5 https://ar-tracking.com/
6 https://www.haption.com/en/products-en/scale-one-en.html
7 https://www.haption.com/en/products-en/virtuose-6d-en.html

2.1.3 *Augmented reality spaces*

Augmented reality (AR) has the potential to transform any physical space into a large interactive space, as it can display virtual content anywhere in the physical space and allow users to interact with this content (Figure 2.1-c). Early work on augmented reality appeared at the beginning of the 1990s with systems, such as *KARMA* proposed by Feiner et al. [FMS93]. Early applications included industrial applications [CM92], medical applications [Fuc+96] or human-robot interaction [Mil+93]. Since then, AR hardware and software have drastically improved, as highlighted by Billinghurst et al. in their survey [BCL15]. Modern systems are now able to accurately track device positions, sense the physical world geometry in real time, and display content on lightweight wearable headsets. Virtual content can thus be seamlessly integrated into the real world with stable placement and convenient ways to interact with it. Moreover, augmented reality inherently supports collaboration between multiple users sharing the same physical space, since they can see each other in the real world. As a consequence, the number of AR applications has exploded, and they now reach a very wide range of domains, including marketing, medicine, education, entertainment, and architecture, as detailed in [BCL15].

With the large development of mixed reality headsets available to the general public, there is an increasing number of devices capable of delivering high-quality augmented reality experiences. These devices can be categorized into two types: optical and video see-through devices. Optical see-through devices enable users to see the real world through semitransparent screens, onto which the virtual content is overlaid. In contrast, video see-through devices display video feeds from cameras located on the device, and integrate the virtual content inside these video feeds. In the work presented in this manuscript, I used optical see-through devices, and more specifically the *Hololens 2* headsets from Microsoft[8].

2.2 INTERACTION IN LARGE INTERACTIVE SPACES

Large interactive spaces require specific techniques to interact with digital content due to their unique characteristics. In particular, they involve going beyond conventional 2D interfaces and keyboard-mouse interaction, as users must handle large visualization spaces while moving within the system. In this section, I present some previous work illustrating how users can interact with both 2D and 3D content in large interactive spaces. I also detail other related work that explores how multiple users can interact in the same interactive space.

2.2.1 *Interaction with 2D content*

In this subsection, I concentrate on how to interact with 2D digital content displayed on large screens, such as wall-sized displays or digital tabletops. A first solution is to interact at close proximity to the display through direct touch. However, this solution requires users to travel long distances along the screen and can lead to difficulties in accessing some areas, such as the very top or bottom of the screen. To solve this issue, Bezerianos and Balakrishnan [BB05a] introduced a dedicated

8 https://www.microsoft.com/hololens/

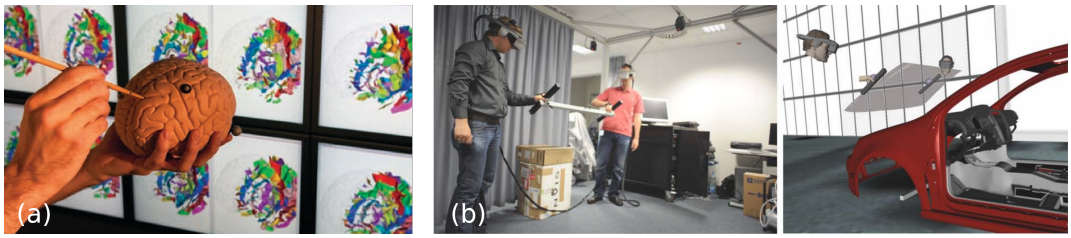

Figure 2.2: Tangible interaction in large interactive spaces: (a) a 3D brain model is used to control brain scans on a wall-sized display (Figure from Beaudouin-Lafon [Bea11]), and (b) a shared prop is used to manipulate a virtual car windshield by co-located users immersed in VR (Figure from Salzmann et al. [SJF09]).

widget on the display to attract far away items close to users. In a similar spirit, *Canvas portals* [BB05b] enable users to interact from distant parts of the display, using interactive portals that replicate the digital content of remote display areas.

To fully take advantage of wall-sized display characteristics, it is common to step away from the screen to get an overview of the displayed data. However, users still need to interact with the display content even when located at a distance. Previous work has explored interaction techniques to achieve mid-air pointing on wall-sized displays. Nancel et al. [Nan+15] provide a comprehensive overview of these techniques. Other studies have investigated the use of mobile devices to interact at a distance with wall-sized displays, ranging from smartwatches [Hor+18] to tablets [Kis+17] or tabletops [Bea11]. For example, *Smarties* [CBF14] proposes a generic solution for using touch on mobile devices to interact with wall-sized displays. It provides an interface on the mobile device, as well as a communication protocol for sending input to a wall-sized display to control interaction cursors or other specific elements. Lastly, 3D interaction with tangible props can also be used to interact with digital content on wall-sized displays. For instance, Beaudouin-Lafon [Bea11] described a scenario in which neuroscientists manipulate a stick in relation to a physical 3D brain model (Figure 2.2-a). The stick mimics a camera pointing at the brain and controls the orientation of brain scans distributed on the wall-sized display. *WallTokens* [CAC21] also proposes to use tangible objects that can be slid and attached to a wall-sized display to manipulate the digital content.

2.2.2 *Interaction with 3D content*

A large number of interaction techniques have been designed to interact with 3D content in mixed reality environments. An exhaustive list of 3D interaction techniques can be found in the book by LaViola et al. [LaV+17]. These techniques are usually classified in three categories, as proposed by Hand [Han97]: (i) selection and manipulation, (ii) navigation and (iii) system control. System control consists of sending commands to the application for requesting specific actions, changing the interaction mode, or modifying some parameters.

To select and manipulate virtual objects, we can use two approaches as for the interaction with 2D content. The first one involves traveling close to the virtual objects and performing direct manipulation. The simplest solution to achieve this is to use a virtual hand mimicking the movements of a user's real hand, as studied by Jacoby et al. [JFH94]. However, the travel actions required to reach

the objects can be inconvenient and increase interaction complexity. The second approach thus takes advantage of being in a virtual environment to interact with objects at a distance with, for example, the *"Go-Go" technique* [Pou+96] or a virtual ray [Min95]. Argelaguet and Andujarb [AA13] presented an extensive survey of pointing techniques for 3D object selection in mixed reality.

A wide variety of navigation techniques have been proposed for traveling in virtual environments. An obvious approach is to let users walk, with their physical movements mapped to virtual displacements. This has many benefits [LaV+17], such as being easy to learn and use, increasing immersion as studied by Usoh et al. [Uso+99], reducing motion sickness, and promoting spatial understanding. However, as the physical space is limited, users can only cover short distances. Redirect walking, originally proposed by Razzaque et al. [RKW01], guides users away from the physical boundaries, enabling them to travel longer distances in the virtual environment. This technique either deforms the users' spatial perception, making them follow a circular path in the physical space [Ste+10] or reorients the virtual environment while users are looking at distractors [CF17]. Nevertheless, real walking is not always feasible or desired, especially in small physical spaces. In such cases, steering metaphors offer an alternative, allowing users to continuously control their direction and speed in the virtual environment. The direction and speed can be defined by a simple device such as a gamepad or based on users' body position [Che+13; KRF11]. Selection-based metaphors can also be used to choose a destination and reach it instantaneously, saving travel time. For example, teleportation techniques using a virtual ray to select the destination are now common in many VR headset applications, as they reduce motion sickness [Wei+18; Hab+18].

2.2.3 *Collaborative interaction*

Previous work has explored co-located collaboration in large interactive spaces. *Liveboard* [Elr+92] introduced a digital whiteboard with pen interaction and is probably one of the first vertical displays supporting group interaction. Jakobsen and Hornbæk [JH14] studied a collaborative problem-solving task on a wall-sized display and identified six distinct collaboration styles. They highlighted that physical proximity of participants is closely related to how tightly coupled they work. In later work [JH16], they assessed the impact of two input modalities on collaboration: touch input on the display and mouse interaction at a distance. They found that wall-sized displays can afford equal participation regardless of input modality. Although touch input seems well suited to collaboration allowing users to negotiate for space, it can lead to more interference and conflicts than with mouse input. Nevertheless, none of these studies provide interaction techniques designed to support collaborative activities. Liu et al. [Liu+16] studied different collaboration styles with pairs of participants in a data manipulation task on a wall-sized display. They varied the interaction and communication capabilities of each participant of the pair. They also included a technique that supports collaborative interaction by allowing participants to assist their partner with data manipulation. The results show the benefits of collaborative interaction, which enables participants to collaborate more tightly even when not in close proximity. Based on these findings, they proposed *CoReach* [Liu+17], a set of cooperative touch gestures that combine input from multiple users, allowing them to show or pass content to each other,

as well as group multiple items on the wall-sized display. *GroupTogether* [MHG12] detects collaborators' spatial formations (*F-formations*) and how they orient and tilt mobile devices (*micro-mobility*). It uses this spatial information to facilitate cross-device interaction between collaborators' tablets and a wall-sized display. This is an interesting example of how sociological constructs can be leveraged to support appropriate collaboration techniques.

Other studies have concentrated on co-located collaboration in immersive VR systems. A few CAVE-like systems are able to display multiple stereoscopic views of the virtual environment, thus supporting multiple users. For example, the Eve system (Section 2.1.2) provides perspectively correct views for two users, while the *C1x6* system [Kul+11] can accommodate up to six users. This system also explored various navigation techniques to enable all six users to travel together in the virtual environment. Aguerreche et al. [ADL10b] proposed using a reconfigurable tangible device for collaborative manipulation in a CAVE-like system. Co-located collaboration can also be beneficial when multiple users equipped with VR head-mounted displays are in the same room. Although they cannot see each other, they can still hear each other and feel the force applied by others when sharing tangible objects. Such shared tangible objects can simulate users holding virtual objects together in the virtual environment, as studied by Salzmann et al. [SJF09] with a virtual car windshield (Figure 2.2-b). Navigation can be challenging in such co-located situations, as it is mandatory to maintain the spatial relationship between users while they travel the virtual environment. *Multi-Ray Jumping* [WKF19] provides a collaborative teleportation technique that preserves this spatial relationship.

Augmented reality inherently supports co-located collaboration as users can see each other, providing mutual awareness. As a consequence, early work on augmented reality already explored collaborative systems that let co-located users see and interact with shared AR content, such as *TransVision* [Rek96], *Shared Space* [BWF98], or *Studierstube* [Sza+98]. Later, Billinghurst et al. [Bil+02] developed a collaborative AR system including tangible elements that users can manipulate to interact with AR content. They showed that participants' behaviors with this system are closer to unmediated collaboration than with a large 2D display. Kiyokawa et al. [Kiy+02] studied communication behaviors of co-located users with various AR devices and AR content placement. They found that optical see-through headsets with a task space situated between participants may produce the most natural collaboration. This can be explained by the fact that participants could better perceive non-verbal communication cues with this combination of device and placement. A recent survey by Sereno et al. [Ser+22] compiles notable research work on co-located collaboration in AR over the last decade. While all these techniques enable users to act on shared AR content, very few of them support collaborative interaction or provide solutions to mitigate interference and conflicts during interaction. Oda and Feiner [OF09] introduced a redirected motion system designed to prevent interference between two users, but it is dedicated to hand-held AR devices.

2.3 COLLABORATION ACROSS REMOTE INTERACTIVE SPACES

A vast body of research in computer-supported cooperative work has explored remote collaboration. I distinguish two categories among previous work that studied collaboration across remote large interactive spaces. The first category concerns

telepresence and video-mediated communication systems, which enable users to share video across remote locations. The second category involves collaborative virtual environments and aims to immerse remote users in a shared virtual environment using mixed reality technologies.

2.3.1 *Telepresence systems*

Video plays a crucial role in supporting non-verbal cues, turn-taking and shared understanding of the collaborative situation among remote users, as emphasized by Isaacs & Tang [IT93]. Similarly, Monk & Gale [MG02] observed that gaze awareness provides an alternative non-linguistic channel for checking mutual understanding among remote collaborators. When engaged in collaborative tasks, seeing each other's faces improves the negotiation of common ground, as shown by Veinott et al. [Vei+99]. Properly conveying gaze direction is challenging in such systems, due to the disparity between camera and viewer positions, as well as the fact that video is displayed on a flat screen. Previous work has explored solutions to support the correct interpretation of gaze direction. For instance, *Hydras* [SBA92] used multiple mobile devices combining a screen and a camera. These devices can be placed in a way that reflects remote collaborators' positions, thus preserving eye contact. *Multiview* [NC05] relies on a multi-view screen and multiple cameras to display an individual view with a correct perspective for each user. Pan and Steed [PS14] designed a cylindrical screen to preserve gaze direction by displaying perspective-correct images for multiple viewpoints around a conference table. Nevertheless, most of these systems consider static users sitting around a table.

Other work has focused on users moving in their physical space, using wall-sized displays to simulate a large glass window between the remote spaces. Willert et al. [Wil+10] connected two wall-sized displays by capturing video through a 2D grid of cameras and displaying video remotely with a motion parallax effect when the viewer moves. Consequently, this system supports only one user at each remote location. Dou et al. [Dou+12] created a similar setup by using a wide-angle camera to capture the background and multiple video+depth cameras to capture users in the foreground. Users are segmented using the depth data and overlaid on the wide-angle video, preserving eye contact with remote collaborators.

When collaborating remotely, users often need to work on shared digital content. Early work on telepresence investigated how to provide users with the same interaction space. *VideoDraw* [TM90] was a shared drawing tool that overlaid the

Figure 2.3: Telepresence across large interactive spaces: (a) *MirrorBlender* blends video feeds and shared screens from remote collaborators (Figure from Grønbæk et al. [Grø+21]), and (b) *t-Room* uses screens arranged in a circle to overlay remote collaborators' video feeds onto shared digital content (Figure from Luff et al. [Luf+15]).

drawing with a shadow of a remote collaborator's arm. It thus allows users to perceive the collaborator's hand gestures. It also includes a screen displaying the remote collaborator's face on top of the drawing area. *VideoWhiteboard* [TM91] improved on this system by overlaying the shared drawing with a shadow of the collaborator's full body. *Clearboard* [IK92] extended this idea by blending the shared drawing with the video of the remote collaborator using transparency. The system could thus convey facial expressions and gaze awareness. More recently, *MirrorBlender* [Grø+21] designed a video-conferencing system that enables users to reposition, resize and blend with transparency multiple video feeds and shared screens from remote collaborators (Figure 2.3-a). Finally, *t-Room* [Luf+15] is a telepresence system connecting two remote locations with large screens arranged in a circle around a digital tabletop (Figure 2.3-b). Cameras are attached on top of each screen and capture users inside the interactive space. Shared digital content on the screens is overlaid with the remote video in a way that preserves the physical relations between the users and digital objects. However, this system has the drawback of requiring exactly the same complex setup at both locations.

2.3.2 *Collaborative virtual environments*

Collaborative virtual environments (CVEs) enable remote users to share virtual content, whether they use AR or VR devices. Early work primarily focused on distributing and synchronizing the virtual environment data across remote locations, as studied during my PhD [**Fle+10c**; **Fle+10b**]. Practical solutions such as Ubiq [Fri+21] are now available for creating CVEs. Recent work mainly concentrates on improving awareness and collaborative interaction among remote collaborators.

In such environments, providing mutual awareness is mandatory to support effective collaboration and social presence among remote users. A large number of studies have investigated the use of avatars to provide embodied representations of users in the virtual environment. Current systems can now create highly realistic avatars with simple technical setups, such as those proposed by Bartl et al. [Bar+21]. However, other authors prefer using low realism or cartoon-like avatars [FM21; KMI19], as realistic avatars can induce an Uncanny Valley effect [MMK12]. There is not a clear consensus regarding the impact of these two approaches on social presence, and it could highly depend on the collaborative context, as highlighted by Yoon et al. [Yoo+19]. In addition, a few systems have used live video [Ben+95] or real-time 3D reconstruction [Bec+13] to represent users in virtual environments (Figure 2.4-a). Congdon et al. [Con+23] compared the effects of video and 3D avatar representations on user trust in a CVE. The results are not clear-cut, but it appears that animated 3D avatars can perform as well as full-body video and better than head-and-shoulder video. Once again, the collaborative context and the environment surrounding the users' representations are likely to have significant effects on these results, and further studies would be required.

While remote users can interact in CVEs using the individual interaction techniques presented in Section 2.2.2, additional techniques can be useful to support collaborative practices. In particular, it is crucial to manage conflicts that may arise when users manipulate the same virtual objects and to allow cooperative manipulation of these objects. To prevent conflicts, *Spacetime* [Xia+18] creates a parallel version of objects whenever a conflict occurs. Users can thus manipulate their own

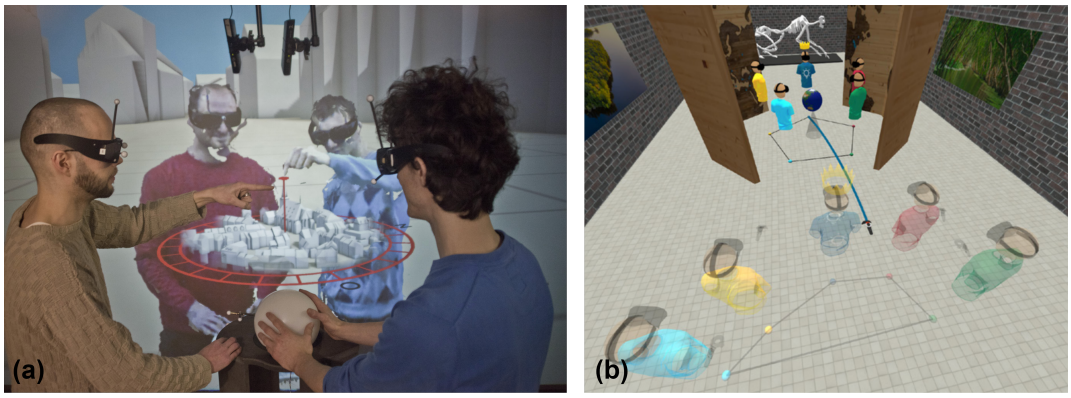

Figure 2.4: Collaborative interaction in virtual environments: (a) a World-In-Miniature is used to move collaborators in the environment (Figure from Beck et al. [Bec+13]), and (b) a specific teleportation technique enables a group to travel together while adjusting its spatial formation (Figure from Weissker et al. [WF21]).

version of an object and review all the created versions before choosing the final one. Pinho et al. [PBF02] propose to combine users' actions by separating the degrees of freedom they control. To enable simultaneously manipulation of the same degrees of freedom, Noma and Miyasato [NM97] used a physical simulation to compute the force applied by each user to the virtual object. *SkeweR* [DLT06] enables two users to simultaneously grab a virtual object using two control points. The position and orientation of the object are determined by the positions of these control points. Aguerreche et al. [ADL09] suggest adding a third control point to more precisely manipulate the object orientation along the three axes. One user thus manipulates two control points, while the other manipulates the third one. During my PhD, I used this 3-hand manipulation technique to assess the co-manipulation of a shared virtual object by two remote users in CAVE-like systems [**Fle+12**].

Collaboration navigation in CVEs can be challenging because it can be difficult for users to meet and travel together in the virtual environment. It is even more complex when teleportation is involved, as collaborators may disappear when they "jump" to another location in the environment. To overcome this issue, *Spacetime* [Xia+18] proposes to use parallel versions of the collaborators' avatars. When a collaborator teleports, a parallel avatar remains at the original location, allowing any user to select it and follow the collaborator. Additionally, *Spacetime* enables users to create a parallel avatar by grabbing a collaborator's avatar. Users can move this parallel avatar to a new location where they want to show something to the collaborator. The collaborator then receives a notification and can decide to travel automatically to this new location. A World-In-Miniature (WIM), originally proposed by Stoakley et al. [SCP95], is also a convenient way of moving collaborators in the virtual environment by manipulating their 3D representation in the WIM, as shown by Beck et al. [Bec+13] (Figure 2.4-a). To support group navigation, Weissker et al. [WBF20] designed a teleportation technique that allows two users to travel together while adjusting their spatial formation. Later, they extended this technique to support groups of up to ten users [WF21] (Figure 2.4-b).

2.4 SUMMARY

This related work section defined large interactive spaces by providing a selection of relevant work on wall-sized displays, immersive virtual reality systems and augmented reality spaces. It also described the main characteristics of the interactive systems that I used in my research.

Moreover, I presented examples that illustrate interaction and collaboration aspects in such large interactive spaces. Some previous work introduced interaction techniques for managing a large amount of 2D content on wall-sized displays, while others proposed techniques for navigating and interacting with 3D content in mixed reality environments. Although the content differs between these two contexts, there are similarities in terms of interaction, especially regarding pointing techniques or other techniques to interact at a distance, such as portals or tangible props. These similarities arise from the fact that users interact within a 3D space in both cases. Nevertheless, it is worth noting that no standardized interaction exists in these systems and that techniques usually need to be customized to suit specific application contexts. It is also interesting to observe that only very few techniques offer true collaborative interaction among co-located users.

Finally, I surveyed previous work on remote collaboration across large interactive spaces, including telepresence systems and collaborative virtual environments (CVEs). For telepresence, some systems focus on accurately conveying gaze direction, while others provide users with the ability to share and interact with digital content. However, none of these systems deal with very large interactive spaces, where multiple users can both move freely and interact with shared content from various physical locations. For CVEs, some previous work has explored how avatars can enhance mutual awareness among remote users. While video can be valuable in certain collaboration contexts, only a few systems integrate video into mixed reality environment. It is also worth noting that a large body of work on CVEs has focused on interaction in collaborative virtual environments, allowing users to collaboratively manipulate virtual objects and navigate in the virtual environment. Telepresence systems and CVEs offer different advantages, depending on the collaboration contexts and application domains. Nevertheless, almost no previous work has attempted to combine the two approaches by connecting users of heterogeneous devices, including both immersive and non-immersive systems.

3

FROM INTERACTION TO COLLABORATION IN A SHARED INTERACTIVE SPACE

Large interactive spaces are powerful tools for fostering collaboration on both digital and physical content, as they can accommodate multiple users within the same system. In these spaces, users can easily perceive each other's activities, share information and distribute tasks. Nevertheless, providing all users with appropriate interaction capabilities is a fundamental prerequisite for an effective collaboration.

Although large high-resolution displays and mixed reality technologies are becoming more mature and widespread, interaction in such systems remains a challenge. On the one hand, novel interaction paradigms are needed to enable users to fully exploit the specific features of these technologies. In particular, the paradigms must handle large visualization spaces and 3D interaction, while allowing users to move freely within the system and interact from multiple locations. On the other hand, these paradigms need to be designed considering the collaborative nature of the interaction from the outset. They must enable all users to act together with similar capabilities, while managing conflicts that may arise when users interact with the same content and occupy a shared space.

In this chapter, I present my research on interaction and co-located collaboration within a shared interactive space. This work addresses three fundamental aspects of large interactive spaces by investigating (i) how users can handle a large visualization space, (ii) how they can interact in a 3D space, and (iii) how they can take advantage of the large physical space surrounding them. This chapter is divided into three sections, each covering a different aspect. For each aspect, I first propose interaction paradigms that leverage the specific features of the system and support multiple users. I then study how these paradigms impact collaboration and how they can be extended to further enhance collaborative interaction. This chapter emphasizes the application of the proposed interaction paradigms to various collaborative design scenarios, including 3D sketching, computer-aided design (CAD) and industrial assembly tasks. Although it provides domain-specific solutions for managing complex data, facilitating 3D interaction and enhancing collaboration, these paradigms can be adapted to a wide range of contexts and data types.

3.1 HANDLING THE LARGE VISUALIZATION SPACE

Ultra-high-resolution wall-sized displays can present large amounts of visual information with a high level of detail, as presented in Section 2.1.1. However, interacting with such large visualization spaces requires specific techniques that provide a wide range of actions along with a high degree of precision. Touch interaction is a relevant option for interaction with wall-sized displays, as it satisfies these requirements while also enabling multiple users to interact simultaneously from different locations. Nevertheless, touch interaction techniques need to be redesigned to suit wall-sized displays, as standard touch-based techniques are mainly designed for single users with small handheld devices or horizontal screens.

While most previous work about interaction with wall-sized displays concentrates on 2D content (Section 2.2.1), the benefits of wall-sized displays can be extended to 3D data. This section targets a collaborative scenario in which a design team wants to create and explore numerous alternatives of 3D computer-aided design (CAD) objects on a wall-sized display. The first subsection focuses on the design of touch-based interaction to manipulate 3D objects on a vertical display. This work was published at VRST 2016 [**LF16**]. The second subsection investigates how touch interaction can be used to generate and distribute multiple design alternatives of a CAD object on a large display. Building on these outcomes, the last subsection presents a collaborative system that supports collaborative exploration of CAD data on a wall-sized display, and evaluates its benefits. This last study leads to more generic recommendations on how a large visualization space can be shared to enhance collaboration and empower users to perform complex tasks. The work described in these last two subsections was published at CHI 2020 [**Oku+20**].

3.1.1 *Multi-touch 3D interaction on a vertical display*

We explored the design of touch-based interaction for users interacting with 3D content while standing in front of a large wall-sized display. This scenario is motivated by the increasing availability of large multi-touch screens in meeting rooms, classrooms or public spaces. Multi-touch interaction provides a relevant solution for manipulating 3D objects in such contexts, as it can be easy to use and learn even for non-experts. It also does not require additional hardware beyond what is already embedded in the device.

Previous work has investigated touch-based 3D interaction on standard devices such as smartphones, tablets or tabletops. It proposes several solutions to perform 6-degree-of-freedom manipulation of 3D objects, such as mimicking direct 3D manipulation on the objects [RDH09; HCC09; JSK12] or combining different touch inputs to control separated degrees of freedom (DOF) [HCC07; MCG10a]. Other studies demonstrate that controlling separated DOF [HCC07] or separating the control of translation and rotation [MCG10b] improves the performance of the 3D manipulation. While these outcomes are still valid for wall-sized displays, the proposed techniques must be adapted to the new constraints introduced by such devices. In particular, touch input techniques that require the use of several fingers from the same hand may not be convenient when users need to perform actions at the top or bottom of a large wall-sized display. Additionally, long drags across the screen should be avoided to prevent user fatigue.

Other studies have explored touch interaction on large displays for 3D navigation [Yu+10] or 3D data exploration [Lop+16; Cof+12]. However, most of the proposed techniques consider that users stay static in front of the display or require control devices such as tablets or tabletops. In the context of meeting rooms, classrooms or public spaces, we want to design solutions that do not require these additional devices and instead rely solely on direct interaction with the display.

We designed In(SITE) [**LF16**], an Interface for Spatial Interaction in Tactile Environments, to explore touch-based 3D interaction on wall-sized displays. This technique combines bimanual touch interaction and object teleportation features to enable users to perform 6-DOF manipulation on a large vertical display. In(SITE)

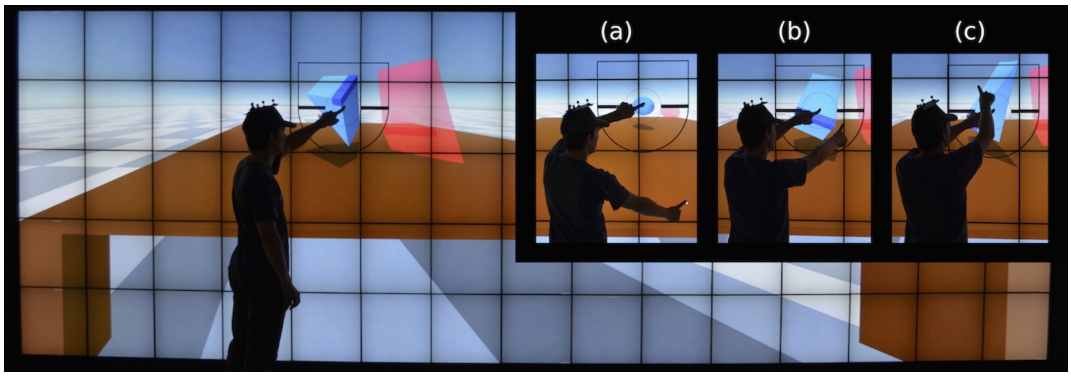

Figure 3.1: 6-degree-of-freedom manipulation of a 3D object on a multi-touch wall-sized display with the In(SITE) technique: the user performs (main picture) x and y translations, (a) z translation, (b) roll rotation, and (c) pitch and yaw rotations.

focuses on the selection, translation and rotation of 3D objects, but does not include scale modification in this first version.

In(SITE) provides a widget divided into several areas that enable separate manipulation of the different DOF (Figure 3.1). When users touch the screen, a raycast is performed in the 3D virtual environment starting from their head position and passing through the touch point on the screen. If the ray hits an object, the object is selected and the widget appears under the finger if it stays in contact with the screen for at least 1*s* (long touch). Users then control the x and y translations in a plane parallel to the screen by moving this primary finger. The z translation is controlled by touching outside the widget with any finger of the other hand. As this secondary finger moves closer to the primary finger, the object moves closer according to the ray axis, and vice versa. This interaction is inspired by the *Z-technique* [MCG10a]. For rotation, the lower area of the widget allows users to control the *roll* with the secondary finger by doing curved gestures, following the object rotation. The upper area allows users to manipulate the *yaw* and *pitch* with the secondary finger by doing respectively horizontal and vertical movements. The *yaw* and *pitch* rotations can be combined by performing diagonal movements.

To avoid long drags across the screen, In(SITE) includes teleportation features to achieve object translation over large distance. Users first need to select an object with a short touch (less than 1*s*), instead of the long touch used to display the widget. The object color is changed to provide feedback that it has been selected. Once users have selected an object, they can teleport it anywhere in the virtual environment using two methods:

- A short touch at the destination, either on the floor or on another object, makes the object fall from above the destination. This method is particularly useful for virtual environments with physical simulation, as the object can be stacked on top of other objects.

- A long touch at the destination makes the object appear under the finger, at the location defined by the intersection between the ray and the virtual environment. The interface then switches to manipulation mode and displays the widget, allowing users to perform final position adjustments. This method is especially useful when users want to reach a precise position for the object.

We conducted two controlled experiments to assess the usability and performance of In(SITE) in comparison to a standard virtual ray technique, also known as the ray-casting technique [JFH94; Min95]. We selected this technique as a baseline because it is widely used in many virtual reality applications and particularly relevant for moving objects over long distances in large visualization spaces. To overcome certain limitations of the virtual ray technique for 3-DOF rotation and ensure a fair comparison, we augmented the technique with a feature that enables rotation along a vertical axis. Both experiments were performed on a $5.90m \times 1.96m$ wall-sized display (see description of the WILDER system in Section 2.1.1). This wall-sized display does not support stereoscopic vision, but we implemented motion parallax to improve depth perception. Both experiments involved 16 participants each, and focused on a docking task with targets positioned on the floor or in mid-air. We hypothesized that the virtual ray would be faster for translation, whereas In(SITE) would be more accurate for fine adjustments, especially when rotation is involved. We also predicted that the teleportation can improve both techniques for translation.

The first experiment used spheres for the docking task, assessing only translation. It compared In(SITE) and the virtual ray technique, both with and without teleportation. The results did not show a significant effect of the techniques on the task completion time and the mean values were almost similar, suggesting that participants reached close levels of performance with both techniques. However, In(SITE) led to significantly fewer overshoots than the virtual ray while adjusting the final position of the object. This is confirmed by the subjective questionnaire which reported that participants found In(SITE) easier to use and more precise. In addition, this questionnaire showed that participants preferred both techniques with teleportation and considered them easier to use and less tiring than the ones without teleportation.

The second experiment used edges, including rotation in the task. It compared In(SITE) with the virtual ray technique, but did not include teleportation as the task involved only short translation. The results did not reveal a significant effect of the techniques on the task completion time, but In(SITE) also led to significantly fewer overshoots than the virtual ray for this task.

Overall, these experiments suggest that In(SITE) can be an alternative for interacting with 3D content on wall-sized displays, as participants reached close levels of performance and better precision for fine adjustments with In(SITE) compared to a standard virtual ray technique. According to participants' feedback, the teleportation feature improves translation tasks in terms of ease of use, fatigue, and user preference. However, In(SITE) is a first prototype, which can be further improved. In particular, additional work would be required to investigate other designs of the widget, adjust transfer functions for indirect interaction and adapt the technique to stereoscopic display by following the guidelines presented by Valkov et al. [Val+11].

3.1.2 *Interaction with numerous design alternatives*

In the context of industrial design, we focused on using touch interaction to explore a large number of design alternatives on a wall-sized display. The first objective was to enable non-experts in computer-aided design (CAD) to modify such parametric models and to generate many design alternatives, using simple touch interaction instead of specifying complex geometric parameters. The second

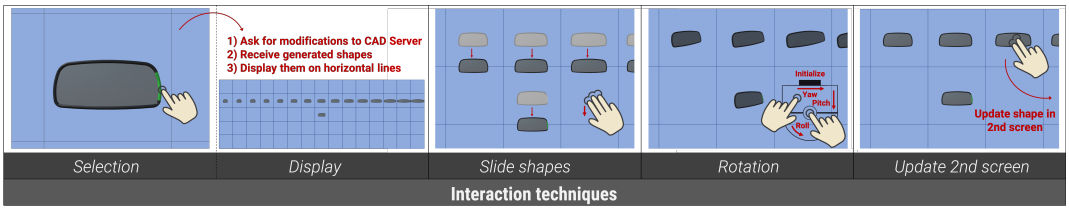

Figure 3.2: Interaction with *ShapeCompare*: when users select a part (Selection), the system displays a set of design alternatives on a row of the screen (Display). All alternatives can be scrolled up and down with a three-finger drag (Slide shapes) and rotated in 3D with the In(SITE) widget (Rotation). A specific alternative can be selected with a two-finger long press (Update 2nd screen) and displayed in context on another screen.

objective was to provide them with a solution for distributing and comparing these design alternatives by taking advantage of the large visualization space available. The target scenario was a co-located collaborative design situation, where a multidisciplinary design team, including designers, engineers, and ergonomists, wants to evaluate and adjust product designs using digital mock-ups, as described by Mujber et al. [MSH04].

Several interaction techniques have been proposed to assist designers with drawing and sketching during the early stages of the design process, including immersive drawing [Isr+09; SH16], surface modeling [Fio+02], digital tape-drawing [Bal+99; Gro+01; Fle+04; KZL07] and rapid prototyping with bimanual interaction [Ara+13]. However, only a few techniques target detailed design stages requiring the modification of parametric CAD models. A CAD model is a solid model defined by a set of mathematical operations (e.g., extrusion and boolean operations) applied to 2D sketches. Unlike drawing or surface modeling, modifying CAD models requires to manipulate parameters, which necessitates extensive training. Although some solutions enable non-experts to modify CAD data [Mar+17; Cof+13], they are limited to a single CAD model and do not support the generation of new alternatives.

We designed *ShapeCompare* [**Oku+20**] to meet the following criteria: (i) interaction in a large space, (ii) native CAD data modification and (iii) multiple-design comparison. We first implemented a service which generates multiple alternative shapes by varying parameter values of a native CAD model. This service can load native CAD files, modify parameters on request, and send back tessellated meshes using the CAA API of CATIA V5[1]. It also maintains a direct link between the 3D mesh parts and the CAD parameters using a *labeling* concept [CB04].

We created a first prototype to generate and visualize new design alternatives on a wall-sized display. For shape generation, users touch the part they want to change on a displayed shape (Figure 3.2). If the part can be modified, it turns green and the system requests a set of new alternatives by varying the CAD parameter related to this part. In this first version, we defined a minimum and maximum parameter value for each part, and chose a predefined number of values equally distributed within the range. For visualization, new shapes are distributed on an entire row of the screen, above other versions of the CAD model. Each row represents a set of design alternatives for a specific modification. Users can scroll up or down the alternatives using a three-finger interaction, and thus see the full

1 https://www.3ds.com/products-services/catia/

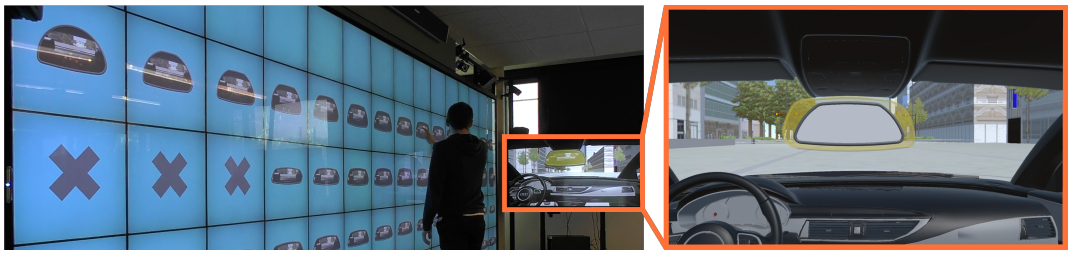

Figure 3.3: User study setup: the target shape is displayed with a transparent yellow color in a realistic environment on an external screen next to the wall-sized display.

design history. Users can also select a part of any shape in the design history and restart modification from that shape. To handle 3D objects, users can rotate the alternatives by using the widget provided by In(SITE) (Section 3.1.1). The rotation of all alternatives is synchronized to maintain a similar same viewing angle.

To assess this first prototype, we conducted a user study and brainstorming session with five students from the civil engineering department of our university. Although they were not CAD experts, they had knowledge of parametric modeling and design process. They had to modify a car rear-view mirror with *ShapeCompare* to reach a given target shape within a 5-minute time limit. They could select a specific alternative with a two-finger long press on the wall-sized display and visualize it in an automotive cockpit on an external screen (Figure 3.3). The target shape was overlaid with a transparent yellow color on this external screen, simulating the design skills of experts assessing alternatives in a realistic environment.

We evaluated our design through the observations of participants' behaviors and interviews. Overall, participants appreciated the interaction techniques and found the system to be beneficial for novice users as it does not require understanding or manipulating parameter values. They also found the shape visualization nice and helpful in generating new ideas. All participants agreed that, although *ShapeCompare* has limited functionalities and cannot replace traditional CAD software, it is valuable for adjustments that do not require changing the entire design intent.

The study outcomes helped us identify the main issues that the participants faced. Firstly, all of them found it difficult and frustrating to understand how part selection affects shape deformation. Secondly, they often needed time to find out how the generated shapes on the new row are different from the one they selected. Because the parameter values used to generate shapes are always distributed between a fixed minimum and maximum, the initial shape on the new row is not displayed above the previously selected one, but at a random position. Aside from these issues, participants often had difficulty distinguishing differences between neighboring shapes, especially for the radii of corners, top, and bottom parts.

Based on these results, we then redesigned *ShapeCompare* to improve: (i) understanding of shape modification and (ii) visualization of design history. For the first aspect, we drew inspiration from the *Suggestive Interface* [IIH07] and created a widget that shows small thumbnails presenting the minimum and maximum shape modification for all parameters of each part (Figure 3.4-left). This widget becomes visible when users select a shape, and allows them to choose a thumbnail for generating the corresponding set of design alternatives. For the second aspect, we changed the way the system generates design alternatives to ensure that the selected shape always appears in the middle of the new row, with equal numbers

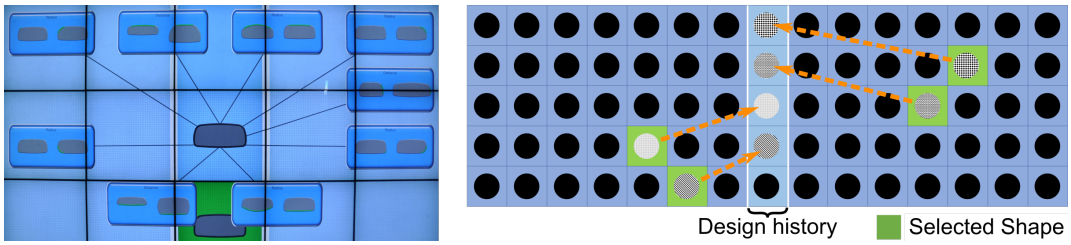

Figure 3.4: Updated version of *ShapeCompare*: (left) the selection widget shows small thumbnails with minimum and maximum shape modifications for all parameters and (right) all selected shapes are displayed in the middle of the next row to improve the visibility of the design history.

of alternatives displayed on both its left and right sides (Figure 3.4-right). Instead of defining a fixed minimum and maximum for the parameter values, we defined a specific offset for each parameter. The system thus generates the shapes by incrementally increasing and decreasing the parameter by the offset. Consequently, the middle column gathers all the previously selected shapes, allowing users to easily track the progression of modifications.

In summary, we employed an iterative design process involving potential users to create a custom interaction technique that facilitates CAD model exploration for non-experts. Our objective was to challenge the traditional design methodology in which users iterate on a single model. Instead, we proposed a solution that allows users to generate many design alternatives and explore them on a wall-sized display. Although we studied this approach in a specific context, visualizing "small multiples" on a wall-sized display could be extended to other contexts as long as parameter variations are involved. For instance, it can be applied to generative design [Che+18; Kaz+17] in which users can specify preferred designs to an Artificial Intelligence, or physical simulations such as weather predictions in which users can run several simulations with varying parameter settings. Furthermore, our approach is valuable for collaborative design, as it allows multidisciplinary teams, including non-CAD experts, to explore, compare, and reflect on design alternatives. I detail such a co-design scenario in the following subsection.

3.1.3 *Collaborative data exploration on a large display*

We aimed to investigate the potential benefits of using a wall-sized display to enhance collaboration within design teams during review meetings. Most of the time, the design process is iterative and relies mainly on two steps that involve many stakeholders: design discussion and CAD data adjustment. We aimed to create a collaborative system using a wall-sized display that could merge these two steps. It must enable multidisciplinary teams to collectively share and organize the large visualization space for generating and comparing numerous design alternatives.

Review meetings are a crucial aspect of the industrial design process. They typically take place in interactive systems that provide a full-scale design visualization, a large interactive space and a collaborative environment. For example, *Portfolio Wall* [Bux+00] displays different designs as tiled thumbnails on a large screen, mimicking traditional wall-mounted corkboards. Khan et al. [Kha+05] propose a tool that highlights the area where users need to pay attention on a projected display,

Figure 3.5: *ShapeCompare* enables multidisciplinary experts to generate and explore a large number of design alternatives on a wall-sized display.

thereby facilitating group meetings. Additionally, several virtual reality systems can display CAD data [Ber99; Ran+01; Rap+09]. However, all these systems limit users to comparing just a few static design alternatives during each review meeting, and these alternatives usually need to be prepared beforehand. Consequently, designers are unable to explore new ideas by generating and modifying design alternatives of CAD data in real time during the meeting. This limitation hinders their creativity and forces them to rely on a time-consuming iterative process. A few systems enable users to modify native CAD data [Mar+17; **Oku+18a**], but they focus on deforming a specific CAD model and do not consider the generation of new alternatives.

We used *ShapeCompare* [**Oku+20**] to create a collaborative system enabling multiple users to generate a large number of design alternatives, distribute them on a large wall-sized display and collaboratively compare them (Figure 3.5). This system relies on the interaction techniques described in Section 3.1.2. In such context, the wall-sized display is an efficient tool to show multiple variations of a same object, foster design discussions among multidisciplinary experts and enable them to explore more alternatives without using a conventional CAD system.

We conducted a controlled experiment comparing *ShapeCompare* with another visualization technique suitable for standard screens, called *ShapeSlide*. *ShapeSlide* displays only one shape at a time and enables users to change the shape displayed at the center of the screen with a sliding gesture. We used the same wall-sized display for both conditions (see description of the WILDER system in Section 2.1.1) in order to reduce bias that could be introduced by different devices, participants' positions, or interaction techniques. However, only a small part of the wall-sized display was used for *ShapeSlide*, simulating the use of a smaller screen. 12 pairs of participants performed a constraint solving task with both conditions. This task was based on actual industrial practices and involved modifying a car rear-view mirror, simulating expert negotiation on various design criteria. Due to difficulties in accessing actual industrial designers, we controlled participants' expertise by giving them individual design criteria based on simple numerical values computed by the system. In each pair, the first participant focused on the general properties of the mirror shape, such as aspect ratio and asymmetric balance, while the second concentrated on the mirror reflection, including visibility and size of the reflective area. The task ended when the design satisfied each participant's criteria, and when

they both agreed on it. We hypothesized that participants would find the right design faster and with fewer iterations with *ShapeCompare* than with *ShapeSlide*.

The main results show that pairs of participants reached the right design significantly faster with *ShapeCompare* than with *ShapeSlide*. The questionnaires also highlight that *ShapeCompare* was perceived as more helpful for communicating with the partner, and generally preferred by participants. The smaller task completion time with *ShapeCompare* could be explained by this better communication between participants. This is supported by the significantly larger number of deictic instructions used by participants with *ShapeCompare* than with *ShapeSlide*. The alternatives of *ShapeCompare* were often used as references for communication and help participants convey their ideas. On the contrary, more words related to *Magnitude* were used with *ShapeSlide* (e.g. *"much more"* or *"a bit less"*), as they needed to describe their requirements verbally or with their hand gestures. It demonstrates that the alternatives of *ShapeCompare* help collaborators build a common ground, as defined by Clark [CB91], and thus minimize communication costs. Finally, the results from the NASA TLX questionnaire did not show significant differences between conditions, which suggests that displaying lots of alternatives with *ShapeCompare* does not substantially increase the cognitive load of participants.

In summary, this work investigates co-located collaboration on a wall-sized display in the context of industrial design. We used *ShapeCompare* to create a collaborative system that enables multiple users to generate numerous alternatives of a CAD model and distribute them on the wall-sized display. In a controlled experiment, we demonstrated that visualizing many alternatives on a wall-sized display enhances design exploration and negotiation by increasing the common ground among collaborators. These findings can be extended to more generic contexts involving the comparison of multiple alternatives. In particular, the concept of "small multiples" holds promise for facilitating multidisciplinary teams in collaboratively exploring, discussing, and reflecting on their ideas using a wall-sized display. The current system remains a research prototype, which leaves plenty of space for exploration and improvement in terms of visualization and interaction. For instance, investigating additional methods for classifying and merging relevant design alternatives is a potential direction for future research.

3.2 INTERACTING IN A 3D SPACE

Large interactive spaces enable users to interact in a 3D space, as most of them can detect user positions and gestures with advanced tracking systems, such as infrared or "Inside-Out" tracking systems. 3D interaction has long been employed in mixed reality for interacting with 3D content in virtual environments (Section 2.2.2). However, it can also be valuable even if the content is visualized on 2D displays. For example, 3D interaction with a physical prop representing a brain was used to control brain scans displayed in 2D on a wall-sized display [Bea11]. Moreover, 3D interaction offers many opportunities in collaborative contexts, supporting multiple users interacting in the same space and providing them with their own interaction area. It can also allow collaborative interaction as users can manipulate together virtual objects in the 3D space [ADL10b; ADL10a]. Despite its advantages, 3D interaction needs to be adapted to its application contexts, as no standards yet exist. Designers should also consider that it may be imprecise and tiring for users if no

precautions are taken. When multiple users interact in the same 3D space, sharing the space may not be obvious, potentially leading to conflicts.

In this section, I explore 3D interaction in various design scenarios. The first subsection focuses on 3D interaction for modifying parametric CAD objects in the context of industrial design. The objective is to enable non-CAD experts to perform direct physical actions on the 3D shape of CAD objects in immersive virtual reality systems. This work was published in *Frontiers in Robotics and AI* [**Oku+18a**] and in a book chapter [**Oku+21**]. The second subsection investigates collaborative sketching in augmented reality. It presents a system that allows several users to interact with multiple versions of 3D content in the same physical space, managing conflicts and fostering creativity. This system was published at IHM 2023 [**FFT23**].

3.2.1 *CAD object deformation with physical actions*

We focused on techniques to modify parametric CAD data in large immersive VR systems. Our goal was to allow users to move around 3D CAD objects, feel them through haptic feedback, and modify them by performing physical actions on their surface. We targeted a scenario where non-CAD experts, such as stylists or designers, want to make simple modifications to CAD objects during a product review session in an immersive system.

While it is possible to create and modify primitives and meshes using shape-based interaction [Fio+02; De +13], applying these interaction techniques to CAD data is challenging due to the unpredictable object deformation resulting from parameter changes. A few VR-CAD applications enable users to modify native CAD data in an immersive system [Bou+10; Mar+17], but they do not support direct interaction with the CAD object shape. For instance, Martin et al. [Mar+17] use a one-dimensional horizontal gesture to increase or decrease a parameter value.

We developed *ShapeGuide* [**Oku+18a**; **Oku+21**], which enables users to deform CAD objects by directly pushing or pulling object surfaces in the virtual environment (Figure 3.6-right). It can include haptic feedback to enable users to feel the CAD object shapes and to increase the precision of deformation actions in the 3D space. To begin the modification process, users must first select the specific part of the CAD object that they want to modify. To handle the "unpredictability" of the shape deformation when modifying CAD parameters, a dedicated service computes a large number of possible shapes from a set of discrete parameter values associated with the selected part (Figure 3.6-left). This mesh pre-computation introduces a loading time after the selection to ensure real-time interaction later. In our prototype, this operation takes a few seconds, but this time can be significantly reduced with more powerful hardware and parallel mesh generation. Once the system has generated the set of shapes, users can explore them using a 3D hand motion. The system computes the distance between users' hand position and the nearest point on each generated mesh. It thus displays the closest mesh to the hand. If provided, haptic feedback is computed as an attractive force to the nearest point of each generated mesh using a magnetic force inspired by [Yam+02]. This haptic feedback attracts the user's hand to the surface of the closest mesh, keeping the hand steady on one mesh or guiding it toward neighboring meshes.

In a controlled experiment, we compared *ShapeGuide* to the one-dimensional horizontal scroll technique previously used by Martin et al. [Mar+17]. The direction

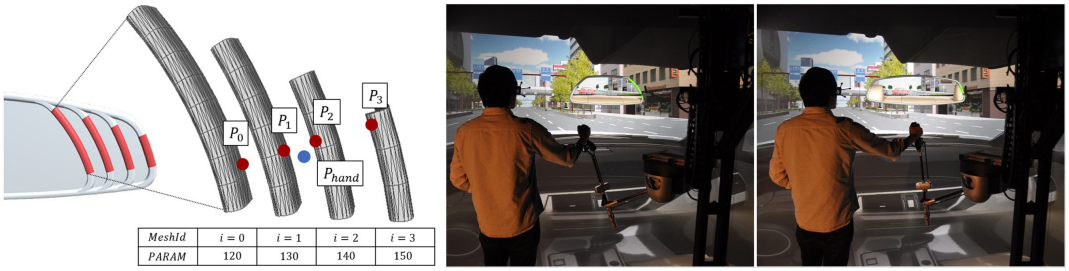

Figure 3.6: (Left) *ShapeGuide* precomputes several meshes of a rear-view mirror and selects the displayed shape according to the user's hand position P_{hand}. (Right) *ShapeGuide* allows users to modify the rear-view mirror shape using physical actions with haptic feedback, while immersed in a virtual car cockpit.

of the scroll was static and may not be consistent with the shape deformation in most cases. We also evaluated the effect of haptic feedback assistance on both techniques. 16 participants had to deform a car rear-view mirror to reach a target shape displayed in transparent yellow. The experiment was conducted in a CAVE system (see description of the Eve system in Section 2.1.2). Participants could interact everywhere in the CAVE with a *Virtuose* haptic arm mounted on a *Scale1* carrier. When no haptic feedback was provided, the haptic device let participants interact with zero force and resistance. We hypothesized that participants would perform the task faster and be more likely to start the deformation in the correct direction with *ShapeGuide* than with the scroll technique. We also predicted that the haptic feedback would improve the accuracy of both techniques.

Results demonstrate that *ShapeGuide* is 42% faster than the scroll technique for the rear-view mirror deformation. This improvement can be explained by a better consistency between shape deformation and user hand motion. In particular, we observed that *ShapeGuide* reduces by 80% the chance that participants move their hands in the wrong direction at the start of their gesture. Additionally, participants perceived *ShapeGuide* as less mentally demanding, less frustrating, less difficult to use, and preferred it over the scroll technique.

The main limitation of *ShapeGuide* is that it tends to produce more overshoots than the scroll technique, especially for parts where the shape variations are close to each other in 3D space. An overshoot occurs when participants reach the target shape but continue their gesture beyond it, causing them to move to the next shape and then come back. However, the results also show that haptic feedback reduces the number of overshoots for both techniques. Therefore, it can be an effective solution to improve the precision of *ShapeGuide*.

Overall, *ShapeGuide* provides an effective solution for deforming CAD objects within an immersive VR system, enabling physical actions to be performed directly on the object 3D shape. It can enhance the current industrial design process by allowing non-CAD experts to modify CAD objects without requiring an in-depth understanding of the CAD data internal organization. This will help avoid time-consuming iterations and potential misunderstandings that can occur when designers have to request modifications from CAD engineers. However, further work is still needed to improve mesh generation. It would be important to reduce generation time and allow users to select the number of generated meshes and the scale level of CAD parameter changes. Additionally, further evaluation of *ShapeGuide* with other industrial CAD models would be necessary.

3.2.2 *Collaborative sketching in augmented reality*

Large augmented reality spaces are valuable solutions for collaborative design, enabling co-located users to create virtual content that overlays their shared physical space. However, conflicts arise when several collaborators want to add or modify virtual content around the same physical objects. Although sharing content among collaborators is crucial in the creative process [Wal+20], others' content can sometimes distract users and hinder their creativity [GBR12]. Our objective was to create a system that helps multiple users share an augmented reality space, allowing them to independently develop their own virtual content while remaining aware of each other's activities and productions.

A wide range of research work focuses on co-located collaboration in mixed reality. Typically, users see and interact with identical virtual content, but a few systems introduce the ability to access or switch between simultaneous versions of this content. In *Slice of Light* [Wan+20a], multiple learners are immersed in distinct virtual environments while being co-located in the same physical space. The system allows the teacher to switch between all the environments by moving in the physical space. *Photoportals* [Kun+14] propose creating portals to access different locations in time or space within the virtual environment. *Spacetime* [Xia+18] uses containers to store and manipulate multiple versions of virtual objects, avoiding conflicts during concurrent manipulation in virtual reality. *VRGit* [Zha+23] provides a tool similar to a version control system to manage various versions of a virtual environment, thereby facilitating collaborative editing. However, these systems only address virtual reality, and do not take into account the relationship between virtual content and physical space, which is a crucial aspect of augmented reality.

Among previous work related to augmented reality, Looser et al. [LBC04] introduce magic lenses that display different layers of virtual content. However, they focus on the technical aspects and do not explore how these layers could be useful for collaboration in a creative process. The concept of *Duplicated Reality* [Yu+22] proposes to duplicate a portion of the physical world into an interactive virtual copy located elsewhere in the augmented reality space. By annotating this virtual copy, a user can guide another user who is performing actions in the physical world without disturbing them. While this system prevents conflicts during interaction, it does not handle multiple versions of the virtual content.

We developed a conceptual framework [**FFT23**] for co-located collaboration in augmented reality. This framework targets design scenarios where collaborators use physical objects as context, landmarks or guides to create 3D virtual content. It allows multiple versions of the virtual content to be associated with a single physical object, potentially representing multiple design alternatives. Users have the freedom to independently control which versions they perceive, and create their own versions without being constrained by those of others. They can also decide to share or not their versions, depending on the stage of the design process.

We reify each version as a *Version Object*, following the concept of reification proposed by Beaudouin-Lafon and Mackay [BM00]. *Version objects* are interactive representations that take the form of semi-transparent spheres containing a preview of the related virtual augmentations (Figure 3.7-c). Users have the ability to grab *Version Objects* and move them into space. *Version Objects* can thus be grouped

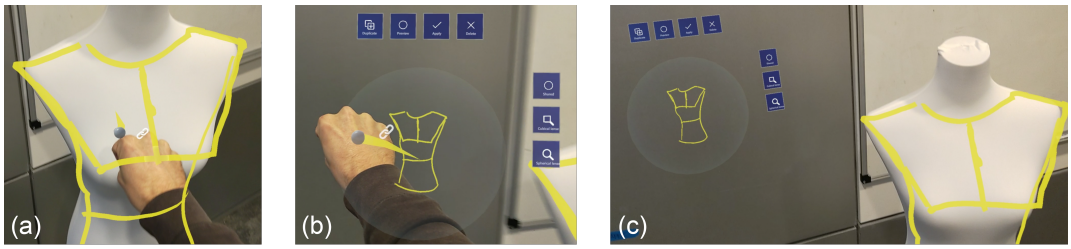

Figure 3.7: Creation of a *Version Object*: (a) a user grabs the 3D augmentations of a physical object. (b) This action creates a new *Version Object* with a preview of the augmentations. The *Version Object* can be moved and (c) stored in the AR space.

together in space, compared with each other, shared between collaborators and applied to the appropriate physical objects.

Users can create a *Version Object* by performing a grabbing gesture on a specific physical object at any time (Figure 3.7-a,b). This new *Version Object* represents the state of the object virtual augmentations at the time of its creation, similar to a photograph taken by a camera. By creating multiple *Version Objects*, users have the ability to capture various stages of their design process or save a version before making edits. When a *Version Object* is created, only its creator can see and access it. However, the creator can choose to share it with others.

Users can switch among different versions associated with a physical object by grabbing a *Version Object* and dropping it onto the corresponding physical object. This allows them to easily return to a previous version, explore various design options that they have created, or review versions shared by others. Our framework also provides users with the ability to simultaneously view multiple versions for comparing design alternatives. They can use either a preview that superposes two versions using transparency and color coding (Figure 3.8-a), or a 3D portal that renders one version inside the portal and another one outside (Figure 3.8-b). This portal can be freely moved in space by users.

To support collaborative design with *Version Objects*, the framework allows users to synchronize or desynchronize the virtual augmentations they see on a physical object. Desynchronization occurs when a user applies a specific *Version Object* to a physical object, thereby switching to a different version of virtual content compared to their collaborators. This feature can be useful to explore different design ideas or to benefit from a private space. During desynchronization, modifications made by collaborators are visible for a very brief period before fading out. This serves as feedback of collaborators' actions and indicates that the augmented reality space is temporarily out of sync. Users can then re-synchronize the virtual augmentations they see to work on the same content or share design ideas. Synchronization occurs when a user requests to synchronize with a specific collaborator. Modifications made by each user become visible to all. A preview mode also allows users to glance at a collaborator's version before deciding to switch to it.

We introduced a use-case scenario to illustrate the functionalities of our framework on a concrete example. This scenario involves two fashion designers who aim to create a new female jacket. They use augmented reality to sketch the virtual outlines of the jacket in 3D on a physical sewing mannequin, which serves as a support and guide for their creation (Figure 3.9). The two designers are co-located in the same room and use AR headsets. Our framework gives them the opportunity

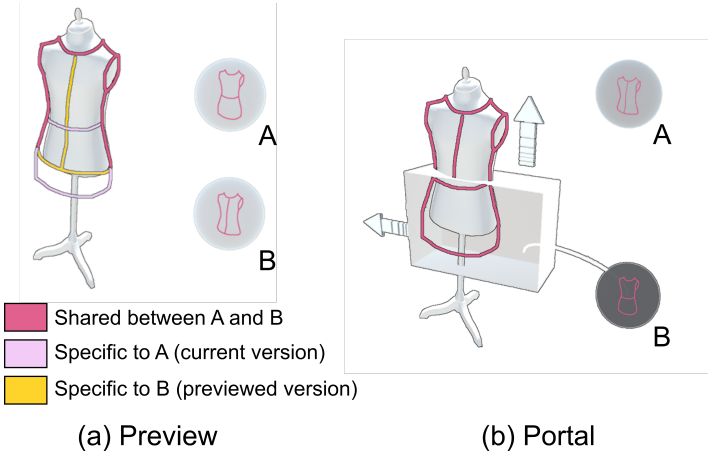

Figure 3.8: Comparison of two *Version Objects*: (a) a preview highlights differences between the current and previewed versions, or (b) a 3D portal allows users to explore differences between the versions inside and outside the portal.

Figure 3.9: Co-located users sketching in AR the virtual outlines of a jacket on a physical sewing mannequin.

to explore their own design ideas, to share them with each other and to collaboratively review them. The scenario consists of two phases: an initial divergence phase followed by a convergence phase, similar to what can be encountered in various creative or engineering processes. During the divergence phase, the designers use the desynchronized mode to individually create different design alternatives, but they can still share *Version Objects* to draw inspiration from each other. During the convergence phase, they switch to the synchronized mode to collaboratively review existing alternatives and make use of preview and portal tools to compare them. We implemented this scenario with two *Hololens 2* headsets from Microsoft.

In conclusion, we propose a framework that supports the various phases of collaborative design in augmented reality. It enables co-located users to perceive distinct versions of the virtual content associated with a physical object. These versions are reified into *Version Objects*, allowing users to control the virtual augmentations they see, explore their own design ideas, or share multiple design alternatives with collaborators. We illustrated the framework capabilities through a fashion design scenario, but its application can be extended to many design processes using augmented reality. Future work should focus on a formal evaluation of our framework, including its impact on the design process. In particular, the experience of co-located users viewing different content considerably differs from regular practices and can be disturbing. Therefore, further study is needed to understand how it influences collaboration and whether it increases users' cognitive load. Moreover, future work should consider more advanced solutions to display and organize the *Version Objects* in space. Drawing inspiration from the representations used by *VRGit* [Zha+23] in a virtual reality context could be valuable. Finally, we could consider creating virtual versions of physical objects using 3D reconstruction techniques, thus extending the concept of *Duplicated Reality* [Yu+22].

3.3 LEVERAGING THE LARGE PHYSICAL SPACE

By definition, large interactive spaces provide users with a vast physical space in which to move and interact. This physical space corresponds to the room or area accessible to users in mixed reality systems, but also to the space available in front of wall-sized displays or other 2D visualization systems. It offers valuable opportunities to explore virtual content through physical navigation. Previous work showed that physical navigation can improve spatial memory [JSH19] and performance in visual search tasks [BNB07] on 2D wall-sized displays. Additionally, physical navigation in mixed reality systems can provide users with vestibular cues that enhance spatial understanding [LaV+17] and immersion [Uso+99], while reducing cybersickness. When designing interaction in such systems, it is crucial to consider the physical space surrounding users and maximize physical displacements.

In collaborative contexts, a spatial relationship naturally exists among users who are co-located in the same physical space. This relationship needs to be considered during collaboration interaction or preserved in virtual environments. For example, when users equipped with VR headsets share the same room, preserving a consistent mapping between their physical and virtual positions allows them to have direct physical contact [Min+20] or co-manipulate shared physical props [SJF09] while immersed in the virtual environment.

This section mainly focuses on immersive virtual reality systems. The first subsection investigates several techniques that enable users to be aware of their physical space when navigating in a virtual environment. These techniques aim to optimize the mapping between the physical and virtual spaces, thereby maximizing users' physical displacements and enabling tangible interaction. These different results were published at EuroVR 2019 [**Zha+19**], VRST 2020 [**Zha+20**] and at the *Workshop on Everyday Virtual Reality* at IEEE VR 2021 [**Zha+21**]. The second subsection addresses a collaborative scenario in which multiple users independently navigate in a virtual environment while remaining in the same physical space. It proposes two collaborative navigation techniques that help users recover a consistent spatial mapping between their physical and virtual positions when they need to interact together. These techniques were published at IEEE VR 2022 [**Zha+22**].

3.3.1 *Virtual navigation with physical space awareness*

In the context of virtual reality systems, we aim to enhance immersion and interaction by taking advantage of the large physical space surrounding users. In all VR applications, this physical space, referred to as the users' physical workspace in this section, is mapped onto a specific region of the virtual environment. This spatial mapping allows users to physically walk in the virtual environment and to perform tangible interaction. Tangible interaction involves associating physical objects with virtual counterparts and using them as substitutes to manipulate the virtual content with passive haptic feedback [SVG15]. Such tangible interaction increases the sense of presence in the virtual environment [Hof98].

The mapping between the real and virtual world is a fundamental issue in every VR applications, and previous work has explored solutions to manage this relationship. Some applications [Che+19; Sra+16] opt to have a fixed one-to-one mapping between the real and virtual environments to avoid user collisions with

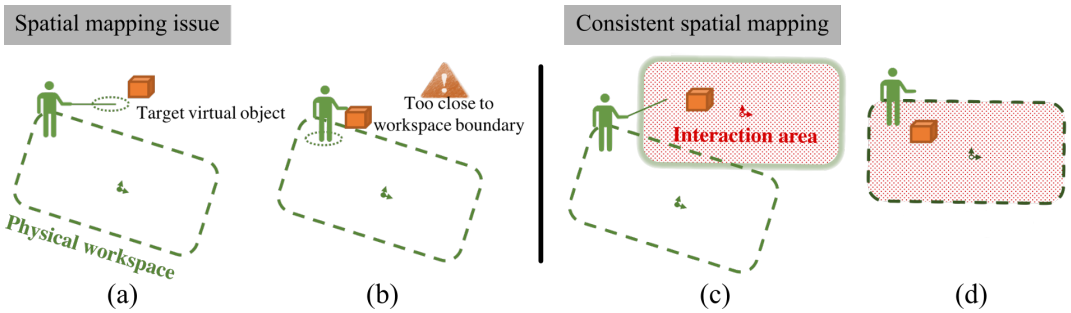

Figure 3.10: (Left) suboptimal spatial mapping that can occur after teleportation: (a) a user teleports themselves to interact with a virtual object without any knowledge of their position in the physical workspace and (b) the object is still out of the workspace boundaries after teleportation and cannot be reached. (Right) the application establishes a consistent mapping between the physical workspace and interaction areas: (c) a user teleports themselves to interact with a virtual object included in an interaction area and (d) the mapping is established during teleportation, allowing the user to walk and reach objects in the entire area.

the real world and grant direct access to all virtual objects. However, this approach constrains the size and shape of the virtual environment. To address this limitation, redirected walking [RKW01] or other view distortion techniques [SWK16] can be used to map a large virtual environment to a smaller physical workspace while allowing users to walk freely. Another approach is to use a self-overlapping architectural layout which allows users to walk through multiple virtual rooms while staying in the same physical room, as proposed by *Impossible Space* [Sum+12]. Nonetheless, these solutions may not be suitable for all applications since they require a reasonably large physical space, and physically walking could also be tiresome when users have to travel long distances.

Virtual navigation, such as teleportation, allows users to travel in a virtual environment beyond their physical workspace boundaries. Some previous studies propose taking into account the users' physical workspace as they navigate in the virtual environment, modeling it as a virtual vehicle [BT02] or a virtual cabin [**Fle+10a**]. However, virtual navigation alters the spatial mapping between the physical workspace and the virtual environment. This mismatch can result in a suboptimal mapping, causing users to unexpectedly reach the physical workspace limits, restricting physical walking and reducing direct access to virtual objects (Figure 3.10-left). Consequently, users may need to repeatedly rely on virtual navigation over short distances instead of using physical movements, which reduces immersion. Additionally, virtual navigation can break the relationship between tangible objects and their virtual counterparts. *Redirected Teleportation* [Liu+18] proposes to combine teleportation and physical walking by maximizing the space available for walking after each teleportation. To activate teleportation, users step into a portal to reposition and reorient themselves away from the physical space limits. However, this technique does not take into account the virtual objects that users need to access and cannot handle tangible objects.

In this section, we aim to recreate a consistent mapping between physical and virtual spaces in specific areas of the virtual environment after a virtual navigation. The main idea is that users can travel long distances freely by using virtual navigation without any distortion or need for additional actions, such as entering a portal.

However, when they need to interact with multiple virtual objects in the same area, the application can assist them in establishing a consistent spatial mapping (Figure 3.10-right). This enables users to directly access the objects by physically walking in the consistent area. Furthermore, the application can help users recover the spatial relationship between tangible objects and their virtual counterparts to perform tangible interaction. We focused on teleportation since it is widely used in VR applications, but this work can be extended to other navigation techniques. We considered standard teleportation with instantaneous transition and no viewpoint animation. Although it can lead to disorientation, this method is the most widely used in VR applications to avoid motion sickness.

We considered three solutions for defining the spatial mapping. First, application designers can specify the mapping in advance for specific areas when tasks are predefined, such as for virtual escape games or VR training with assembly tasks. For more generic applications, users can either choose the mapping manually or select automatically generated areas based on the layout of virtual objects. Finally, the mapping can be defined by the physical object positions for tangible interaction.

3.3.1.1 *Mapping defined by application designers*

As a first step, we investigated VR applications that involve predefined virtual object manipulations taking place in specific areas of the virtual environment. Application designers can thus position the interaction areas where manipulations will occur in advance. These interaction areas have the same dimensions as the users' physical workspace and will be used to create a one-to-one mapping between these physical and virtual spaces.

We introduced two switch techniques [**Zha+19**] based on teleportation to help users recover the mapping between their physical workspace and the interaction areas. In both techniques, users teleport themselves in the virtual environment by using a virtual ray to point towards the destination. However, when users point towards an interaction area, a specific representation is displayed to notify them that a special teleportation technique will be triggered. This teleportation technique adjusts the users' position and the orientation at the destination, ensuring their physical workspace matches the interaction area. Once users are teleported in this area, they can physically walk to access all virtual objects of the area.

The two switch techniques use different representations to display the interaction areas (Figure 3.11). The *Simple switch* shows a transparent cube with a green border indicating the boundaries of the area. The *Improved switch* uses the same cube representation, but it adds a semitransparent cylinder with a 3D arrow showing the users' future position and orientation in the area. This simplified avatar aims to help users anticipate their future location in the area and avoid disorientation. The avatar position and orientation are updated in real time, which means that users can see their avatar moving in the interaction area if they physically walk in their physical workspace before the teleportation.

We conducted a controlled experiment with 18 participants to compare the two switch techniques with a standard teleportation technique used as a baseline. The experiment was carried out using a CAVE system (see description of the Eve system in Section 2.1.2). Participants completed a box-opening task in four separate rooms connected by corridors. They traveled long distances between rooms using the teleportation technique. In each room, participants followed instructions to open

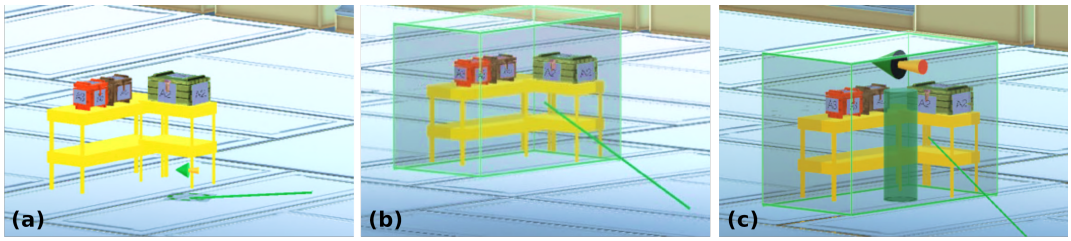

Figure 3.11: (a) standard teleportation technique: users' future position is displayed on the ground when they point towards the destination with a virtual ray. (b) *Simple switch*: a predefined interaction area is highlighted when users point towards it. (c) *Improved switch*: the representation included a semitransparent cylinder with a 3D arrow showing users' future position and orientation in the area.

three out of four boxes. The boxes were positioned in a U-shape in the room and were included in the same interaction area. Participants could teleport themselves either anywhere in the room with the baseline or within the interaction area with the switch techniques. We hypothesized that both switch techniques would improve task performance compared with the baseline. We also predicted that *Simple switch* would be faster, but would increase disorientation compared with *Improved switch*.

Results highlight that helping users recover a consistent spatial mapping improves performance and immersion. The two switch techniques significantly reduce task completion time, the number of teleportations required to achieve the task, and the collisions with physical workspace boundaries compared to the baseline. Participants also reported that both switch techniques were less mentally and physically demanding than the baseline. When comparing the two switch techniques, the *Simple switch* is faster than the *Improved switch* to perform the teleportation in the interaction area because users do not need to look at the avatar. However, the *Improved switch* seems to improve spatial understanding after teleportation as it reduces the time and head rotation required to find the first box, although we did not measure significant differences in the experiment.

In this work, we demonstrate the benefits of creating a consistent spatial mapping between users' physical workspace and specific regions of the virtual environment. This approach is particularly useful in complex scenarios that involve large-scale navigation and manipulation sub-tasks which require access to multiple objects in the same area. We also evaluated the effect of showing a simplified avatar to represent users' future location after teleportation in the interaction area. It seems that the avatar can be beneficial to reduce disorientation, even if it increases the time needed to trigger the teleportation by a few seconds. However, additional studies are required to fully assess the impact of the switch techniques on disorientation.

3.3.1.2 *Mapping defined by users*

Defining in advance the interaction areas where object manipulations will occur is not possible for all VR applications. In this second step, we explored more generic solutions that allow users to define by themselves the spatial mapping between their physical workspace and a designated area of the virtual environment. The goal is to make users aware that a virtual workspace related to their physical workspace exists and to enable them to choose the position of their future virtual workspace before each teleportation.

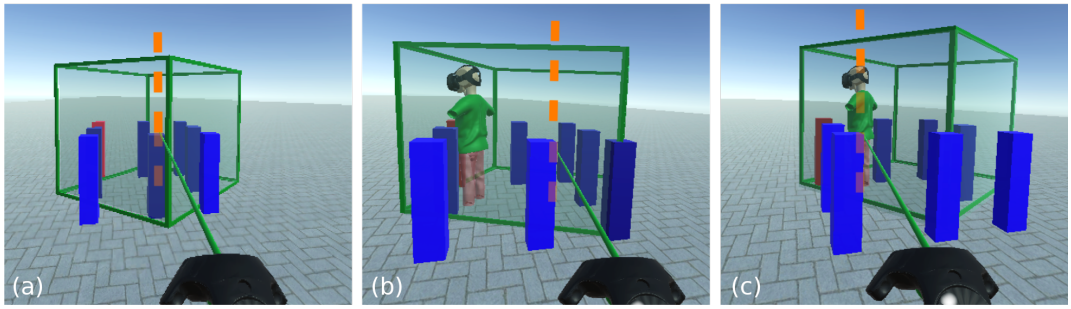

Figure 3.12: Three manual techniques allow users to position their future virtual workspace before the teleportation: (a) *Exo-without-avatar*, (b) *Exo-with-avatar* and (c) *Ego-with-avatar*. The orange dotted line represents the rotation axis of the 3D volume in this figure, but it is not visible to users in the virtual environment.

We designed both manual and automatic techniques to position this future virtual workspace [**Zha+20**]. In manual techniques, users directly adjust the position and orientation of a 3D volume representing their virtual workspace in the virtual environment by using a virtual ray attached to a VR controller (Figure 3.12). The intersection between the virtual ray and the virtual ground defines the 3D volume position, while a circular gesture on the controller touch pad controls its orientation. In automatic techniques, an algorithm computes a set of potential virtual workspaces according to the layout of the interactable objects in the virtual environment. Each virtual workspace alternative is represented by a 3D volume which becomes visible when users point towards it. Users can thus browse these alternatives and select the one they prefer with their virtual ray. In both types of techniques, the virtual objects inside the 3D volume are highlighted to help users understand what objects will be contained in their future virtual workspace. Once users trigger the teleportation, they are moved to the selected virtual workspace and can access all the virtual objects inside this workspace by physically walking.

In a first experiment, we compared three manual techniques (Figure 3.12):

- *Exo-without-avatar* implements an exocentric manipulation by allowing users to move and rotate the 3D volume around its central axis.
- *Exo-with-avatar* uses the same exocentric manipulation, but also includes a transparent avatar that shows users' future position in the virtual workspace after teleportation.
- *Ego-with-avatar* proposes an egocentric manipulation that also includes an avatar. It uses the users' future position (i.e., the avatar position) as the axis to move and rotate the 3D volume. This technique can thus be perceived by users in a different way: they move and rotate the avatar in the virtual environment, and the 3D volume just indicates the space that would be available after teleportation.

The purpose of this experiment was to evaluate the benefits of the avatar representation and to compare the exocentric and egocentric manipulations in terms of spatial awareness and performance. 12 participants performed a simple task in which they had to adjust their future virtual workspace position to enclose eight pillars, teleport themselves inside this new virtual workspace, and touch a specific pillar displayed in red. This last action was included to assess their spatial awareness. Participants

used an *HTC Vive Pro Eye* VR headset in a $3m \times 3m$ physical area. We hypothesized that both techniques with avatar would reduce the time required to touch the pillar compared with *Exo-without-avatar*. We also predicted that *Exo-with-avatar* would reduce the time required to position the virtual workspace, but would increase the time required to touch the pillar compared with *Ego-with-avatar*.

Results show that both conditions with the avatar decrease the time required to touch the pillar after teleportation by over 50%, compared to *Exo-without-avatar*. On the contrary, the time spent positioning the virtual workspace before teleportation appears to be shorter without the avatar, which is consistent with our previous study (Section 3.3.1.1). Although the avatar slightly increases the time and cognitive load required to position the virtual workspace, it can help users better understand the upcoming teleportation and reduce disorientation. This finding is supported by participants' qualitative feedback, which reported better anticipation of their location after teleportation and less disorientation. Regarding the manipulation technique, the results did not show significant differences between exocentric and egocentric techniques in terms of user performance. However, participants preferred the *Ego-with-avatar* condition over the *Exo-with-avatar* condition since it was perceived as "*easier for positioning themselves*" and "*easier for finding*" the target pillar after teleportation.

In a second experiment, we compared a manual technique, an automatic technique and a standard teleportation technique in a more realistic task. Based on the results and participants' preference from the first experiment, we chose the *Ego-with-avatar* technique for the manual technique. For the automatic technique, participants used a virtual ray to select the future virtual workspace from a set of alternatives computed by the system, as described previously. 12 participants performed a task similar to an escape room, where they had to travel through 8 virtual rooms. In each room, they needed to grab 10 objects one by one and bring them in one of 2 boxes available in the room. Half of the rooms had a *No-overlap* layout, which consisted of 2 disjointed areas, each containing five targets and one box. The automatic technique computed 2 virtual workspace positions for this layout. The other half of the rooms had an *Overlap* layout for which the 10 objects and the 2 boxes were placed randomly in a single area. The automatic technique computed 4 overlapping virtual workspace positions to cover all this area. This experiment used the same VR setup as the first one. We hypothesized that both manual and automatic techniques would improve performance and sense of presence compared with the baseline. We also predicted that the automatic technique would perform better for *No-overlap* layouts and worst for *Overlap* layouts compared with the manual technique.

Results show that both manual and automatic techniques outperform the standard teleportation technique in terms of efficiency and immersion. In particular, they significantly reduce the task completion time, the number of teleportations required to achieve the task, and the collisions with physical workspace boundaries. Participants also reported a higher sense of presence in the IPQ questionnaire [RS02; Sch03] with both manual and automatic techniques compared to the standard one. Regarding the comparison between the manual and automatic techniques, both achieve close performance, but each one has advantages depending on the virtual object layout. The automatic technique causes fewer collisions with the physical workspace boundaries in sparse environments (i.e., *No-overlap* layout), but induces

a higher cognitive load for crowded environments (i.e., *Overlap* layout) compared to the manual technique.

Overall, this work highlights the benefits of allowing users to choose the position of their virtual workspace before teleportation. Depending on the virtual object layout, both manual and automatic techniques can be valuable. For manual techniques, exocentric and egocentric approaches perform similarly, but users tend to prefer the egocentric approach. In addition, including an avatar to show the user's future position can decrease disorientation and minimize the time required to locate targeted objects after teleportation. Further studies would be mandatory to assess the proposed techniques in other scenarios including different shapes and sizes of physical workspaces, various virtual object densities and other types of tasks. Automatic techniques could be enhanced by adjusting the clustering algorithm based on the specificity of these scenarios. Finally, the visual representation of virtual workspace can be improved to prevent overloading the user's field of view.

3.3.1.3 *Mapping defined by tangible object positions*

Finally, we studied how to recover the spatial relationship between tangible objects and their virtual counterparts after a virtual navigation. Tangible interaction is a simple and inexpensive solution to provide haptic feedback by associating virtual objects with real objects that share similar physical properties. For instance, a real chair can allow users to sit in the virtual environment or holding a closed umbrella can simulate the sensation of holding a virtual sword [SVG15]. This passive haptic feedback can improve the sense of presence in virtual environments [Hof98]. However, when users perform virtual navigation to travel beyond what is possible according to their physical workspace boundaries, the spatial relationship between tangible objects and their virtual counterparts is disrupted, and tangible interaction is no longer possible.

We explored three advanced teleportation techniques to recover the spatial relationship with a specific tangible object [**Zha+21**] (Figure 3.13). We proposed to teleport (i) the user, (ii) the virtual object, or (iii) both to a new position, while recovering their relative positions. To demonstrate this, we developed a first prototype involving a tracked physical chair that can be used to sit in the virtual environment. The chair has virtual counterparts which the user can interact with in the virtual environment. This prototype used an *HTC Vive* VR headset to immerse the user in the virtual environment and a *Vive Tracker* to track the chair position. For all techniques, the interaction steps are the same: the user first selects the virtual object involved in the tangible interaction (i.e., one of the virtual chairs in our example), then the user adjusts the teleportation destination if necessary and, finally, the user triggers the teleportation.

In *user teleportation*, the selected virtual object serves as an anchor for the teleportation: the user's future position will be defined by the position relative to the object physical counterpart (Figure 3.13-a). For complex objects, such as the chair, the user has only one option as future destination. However, for more symmetric objects, the user may be able to choose several destinations around the object. For instance, in the extreme case of a ball, the user can have an infinite number of destinations all around the ball. According to the available options, sometimes the user does not have other choices than teleporting themselves inside other virtual objects, such as the table next to the chair, for example. In such cases, the system displays the user's

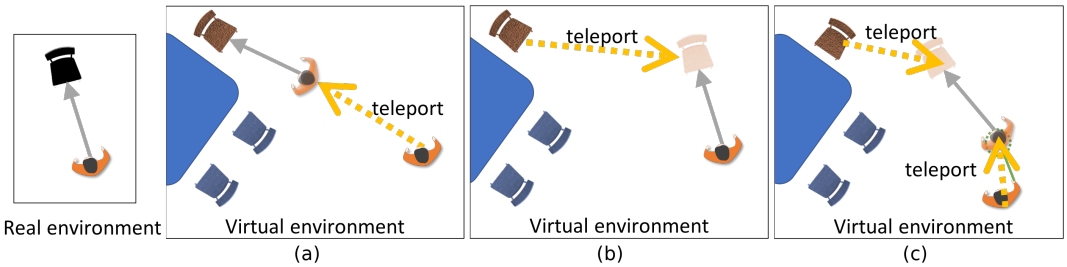

Figure 3.13: Three teleportation techniques allow tangible interaction by recovering the spatial mapping between a real chair and its virtual counterpart: they involve teleporting (a) the user, (b) the object or (c) both to a new position.

future position with a colored avatar, highlights the collisions with virtual objects and asks the user to physically move to a new position that avoids collisions before triggering the teleportation. Colored feedback shows both virtual objects colliding with the user's future position and available areas where the user should move to avoid these collisions.

In *object teleportation*, the user's current position serves as an anchor for teleporting the selected virtual object (Figure 3.13-b). This virtual object will be moved close to the user at the same relative position as the corresponding physical counterpart, enabling tangible interaction. As a consequence, there is a unique position possible for the selected virtual object. However, this position may already be occupied by other virtual objects surrounding the user. In such cases, the system detects the collisions in advance and computes a new position to which the user must teleport themselves before completing the object-based teleportation operation.

In *hybrid teleportation*, the user specifies where both themselves and the selected virtual object will be teleported (Figure 3.13-c). The user can use a virtual ray to define this future position, as in standard teleportation techniques. The virtual object position is computed based on the relative position of its corresponding physical counterpart. A specific representation at the intersection between the virtual ray and the virtual floor shows the future positions of both the user and the virtual object. The user can rotate this representation to adjust the future virtual object location all around their future position. This allows the user to avoid potential collisions with other virtual objects and to choose an appropriate position to prepare for the upcoming tangible interaction.

These three approaches have their respective advantages and disadvantages, and can be applied in different scenarios based on the tangible object characteristics. On the one hand, *user teleportation* may be more appropriate to allow users to access virtual objects that are considered immovable. On the other hand, *object teleportation* may be more suitable for interacting with small objects or tools at specific locations of the virtual environment. Finally, *hybrid teleportation* does not require additional strategies to prevent users or tangible objects from being teleported inside or behind other virtual objects. However, it can be time-consuming and mentally demanding for users. In future work, we need to conduct user studies to evaluate these three techniques in various contexts, including different shapes, sizes, and potential mobility of the tangible objects. Safety issues must also be considered in scenarios where tangible interaction alternates with free navigation. Indeed, physical objects can become invisible obstacles in the users' physical workspace when they no longer have a consistent spatial mapping with their virtual counterpart.

3.3.2 *Collaborative navigation to restore spatial consistency*

Multiple users can be co-located in a virtual reality system, for example, when they are all in the same room wearing VR headsets. In such cases, their relative positions in the physical space usually match those in the virtual environment. This spatial consistency enables users to have direct physical contact with each other [Min+20] or co-manipulate shared tangible props [SJF09]. However, this often excludes individual virtual navigation capabilities, such as teleportation, to preserve the one-to-one mapping between users' relative positions in the physical and virtual spaces. As a consequence, users can only explore virtual environments with approximately the same size and shape as their physical space. Some previous techniques allow co-located users to achieve virtual navigation, but restrict them to traveling together in the virtual environment. For example, *C1x6* [Kul+11] investigated group navigation in a projection-based system and *Multi-ray jumping* [WKF19] introduced collaborative teleportation techniques for co-located users wearing VR headsets [WKF19]. However, group navigation limits users' freedom during a continuous VR experience and is not suitable for many collaborative scenarios. To address these limitations, our objective was to design a system that enhances users' navigation freedom while preserving the capability of sharing the same physical space. In particular, this solution should enable users to independently navigate in a virtual environment, but also help them recover a consistent spatial mapping between their physical and virtual positions when they need to interact together. By doing so, this system would effectively support various phases of the collaboration, including individual exploration and tightly coupled manipulation.

Very few systems have explored how to restore the spatial mapping among co-located users after individual virtual navigation. The system proposed by Min et al. [Min+20] allows co-located users to individually explore a virtual environment larger than their physical workspace using redirected walking. When users need to perform direct physical interaction, such as shaking hands, they use a recovery algorithm that adjusts redirected walking parameters and recovers a consistent spatial mapping. However, this solution requires a large physical space and is not compatible with other navigation techniques, such as teleportation. In a single-user context, we proposed several techniques to recover spatial consistency between the user's physical space and a specific area of the virtual environment, as detailed in the previous subsection. The technique presented in Section 3.3.1.2 enables users to define the future position of their physical workspace in the virtual environment before a teleportation. We extended this technique to a collaborative context.

We proposed two techniques [**Zha+22**] that assist co-located users in recovering spatial consistency after individual teleportation when necessary for the subsequent collaborative interaction. Both techniques use a virtual representation of the users' shared physical workspace, which enables them to adjust the mapping between their physical and virtual spaces. We refer to this virtual representation as the "virtual workspace", as it corresponds to the area of the virtual environment that will be physically accessible to both users after teleportation. In addition, the future group configuration in the virtual workspace is represented by preview avatars, showing where users will be positioned after teleportation with a transparent color.

In the *Leader-and-Follower* technique, only one user, referred to as the *leader*, defines the future position of the virtual workspace before teleportation (Figure 3.14-left).

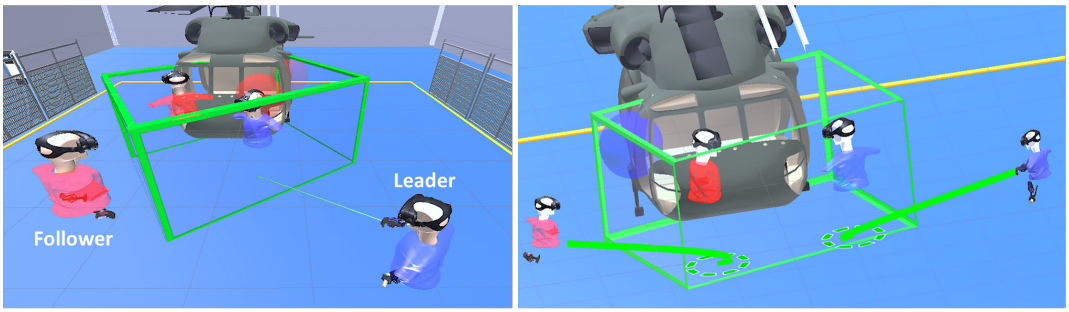

Figure 3.14: Two techniques for recovering a consistent spatial mapping after individual teleportation. The 3D volume framed in green represents the future position of users' virtual workspace mapped to their physical space. The preview avatars show their future positions inside this virtual workspace. (Left) the *Leader-and-Follower* technique allows one user to manipulate the position of the virtual workspace, while the second one can only communicate verbally regarding position requirements. (Right) the *Co-manipulation* technique integrates the inputs from both users, allowing collaborative positioning.

This user manipulates the virtual workspace with a virtual ray attached to a VR controller by using the *Ego-with-avatar* technique. We selected this technique based on the findings of a previous experiment described in Section 3.3.1.2. The virtual workspace position is defined by the intersection of the virtual ray with the virtual ground, while its orientation is controlled by performing a circular gesture on the controller touch pad. The rotation axis is determined by the future position of the *leader* within the virtual workspace. The second user, referred to as the *follower*, can only see the virtual workspace and verbally communicate with the *leader* regarding its position. Once the position is deemed satisfactory, the *leader* ends the manipulation and is automatically teleported to the newly positioned virtual workspace. Subsequently, the *follower* can use a virtual ray to select the virtual workspace and teleport themselves inside it, thereby recovering the spatial consistency between the users. We have divided the teleportation process into two steps, rather than simultaneously teleporting both users, to prevent unwanted teleportation of the *follower* which can lead to frustration and disorientation.

In the *Co-manipulation* technique, the users manipulate the virtual workspace together to define its future position before teleportation (Figure 3.14-right). Both users use a virtual ray attached to their VR controller to indicate their targeted future positions in the virtual environment. The technique equally incorporates inputs from both users using a physically-based approach. The user-defined targeted positions and the users' future positions in the virtual workspace (represented by their preview avatars) are connected by a mass-spring-damper system (Figure 3.15). This system computes the position and orientation of the virtual workspace, enabling users to manipulate it concurrently. Bending rays [Rie+06] are used to provide continuous feedback on users' mutual actions. The navigation technique switches from individual teleportation to the co-manipulation of the virtual workspace as soon as the two users' targeted future positions are close to each other. Once the users agree on the virtual workspace position, one of them can end the co-manipulation. Both users are then teleported into the newly defined virtual workspace, recovering the spatial consistency between them.

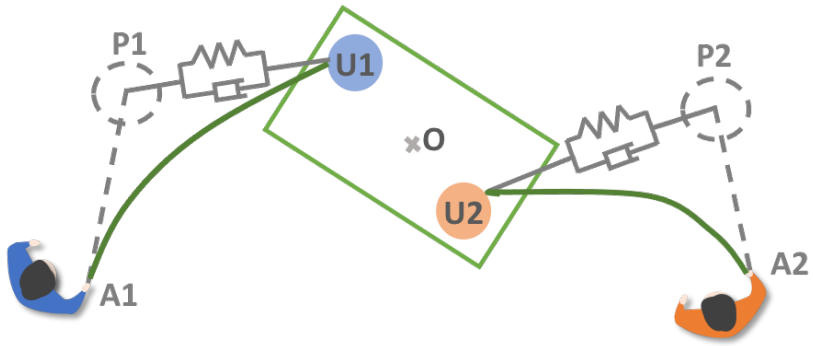

Figure 3.15: *Co-manipulation* technique integrating users' inputs using a physically based system: both a spring and a damper connect the user-defined targeted positions ($P1$&$P2$) with the users' future positions in the virtual workspace ($U1$&$U2$). Bending rays between the users' VR controller ($A1$&$A2$) and their future positions in the workspace ($U1$&$U2$) provide feedback on their mutual actions.

We conducted a controlled experiment to compare these two techniques in a virtual riveting task that consisted of individual navigation and collaborative assembly phases. 24 participants were grouped into pairs and equipped with *HTC Vive Pro Eye* VR headsets, while being co-located in a $3m \times 4m$ physical area. Participants first performed individual tasks at separate locations within a virtual factory: one participant prepared the hammer, while the other collected rivets. Next, they regrouped in a designated area to rivet a helicopter shell together (Figure 3.16). To accomplish this, they needed to recover a consistent spatial mapping between themselves. This spatial consistency allowed for direct physical contact between the participants' VR controllers, providing passive haptic feedback as they hammered the rivet. When recovering spatial consistency, the virtual workspace had to enclose three riveting locations: two locations were only known by one participant, while the third one was only known by the other. As a result, they had to negotiate the positioning of the virtual workspace. We hypothesized that *Co-manipulation* would reduce the time spent negotiating the future workspace position and induce better workspace positioning compared with *Leader-and-Follower*.

Results show that *Co-manipulation* significantly reduced participants' time spent positioning the virtual workspace compared to *Leader-and-Follower*, decreasing the overall task completion time. This can be explained by the fact that participants' intents can be communicated through the manipulation with *Co-manipulation*, eliminating the need for verbal descriptions of positioning requirements. Although no significant difference was found between the two conditions regarding riveting time, participants experienced more frequent collisions with their physical workspace boundaries and had to reposition the virtual workspace more often to perform the riveting task with *Leader-and-Follower* than with *Co-manipulation*. This suggests that participants achieve better positioning of the virtual workspace with *Co-manipulation*. However, *Co-manipulation* could introduce conflicts during the manipulation of the virtual workspace. In particular, some participants found it difficult to understand how they influenced the movement of their virtual workspace.

In summary, we have compared two interactive techniques that assist two co-located users in defining the area of the virtual environment where they want to restore a consistent spatial mapping between their physical and virtual positions. The *Leader-and-Follower* technique allows only the *leader* to position the future virtual

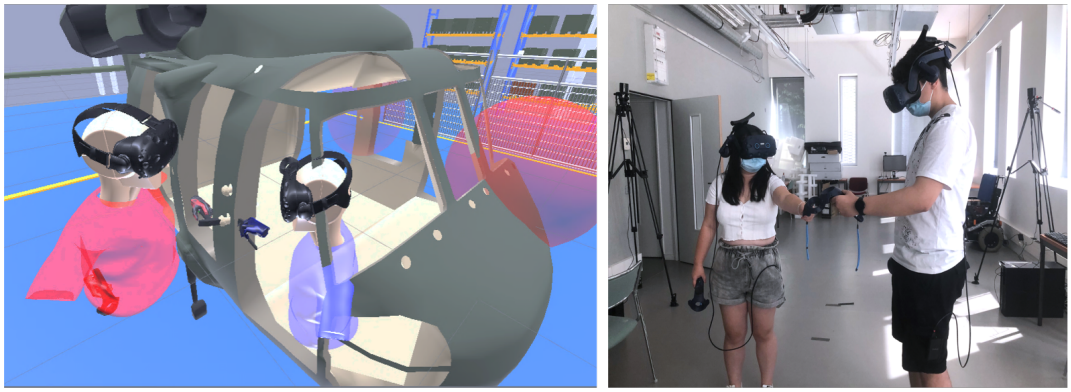

Figure 3.16: (Left) virtual view and (right) real view of the collaborative riveting task on a helicopter shell: users' VR controllers represent a hammer and riveting pliers in the virtual environment. They come into direct contact, providing passive haptic feedback as the rivet is hammered.

workspace, while the *Co-manipulation* technique enables collaborative positioning. Although the *Co-manipulation* technique may be difficult to handle for some users at first, it significantly reduces the time needed to negotiate for the position of the area and enables better placement. Further investigation is necessary to assess these two techniques in various other collaborative scenarios. In particular, the *Co-manipulation* technique can be extended to more than two users, but this may introduce more conflicts and difficulties during concurrent manipulation. In such cases, it could be appropriate to test different spring-damper values, giving users unbalanced control during manipulation and creating an alternative between the co-manipulation and leader-follower approaches.

3.4 CONCLUSION

Appropriate interaction techniques are mandatory to support collaboration in large interactive spaces. In this chapter, I first introduced several interaction paradigms designed to address the specific characteristics of these systems. I especially focused on three key aspects: (i) the large visualization space, (ii) the 3D space available for interaction and (iii) the large physical space surrounding users. All the proposed paradigms support multiple users interacting from different locations within the same interactive space. In a second step, I investigated the impact of these paradigms on co-located collaboration and how they can be extended to provide users with specific features enhancing collaboration. This illustrates examples of how we transition from multi-user interaction to truly collaborative interaction.

For the first aspect, we designed touch-based interaction techniques for creating, manipulating and organizing numerous design alternatives of a 3D object on a wall-sized display. We then studied how pairs of users collaborate during a collaborative design task using these techniques. This study demonstrates that distributing numerous design alternatives on the wall-sized display enhances design exploration and negotiation by increasing the common ground among collaborators.

For the second aspect, we proposed deforming industrial CAD objects in an immersive VR system by physically pulling or pushing on their surface. When multiple users perform such 3D interaction in a shared space, conflicts can arise

and they can disturb each other when interacting with the same data. To address this issue, we designed a collaborative AR system that enables users to interact in 3D with distinct versions of the virtual content. Such systems can provide users with the abilities to explore their own design ideas in the 3D space, while also facilitating the sharing of design alternatives with collaborators at a later stage.

For the third aspect, we investigated various techniques for navigating in a virtual environment, taking advantage of the large physical space surrounding users to maximize physical displacements and allow tangible interaction. We then extended the proposed concepts to collaborative navigation with two co-located users. We designed collaborative navigation techniques that enable users to restore a consistent spatial mapping between their physical and virtual positions, after it has been disrupted by individual navigation in the virtual environment. Our findings highlight the benefits of providing appropriate collaborative tools for such tasks rather than relying solely on verbal communication.

The work presented in this chapter mainly targets collaborative design scenarios, including 3D sketching, computer-aided design and industrial assembly tasks. Although the final interaction techniques are specific to each application context, we created more generic concepts which can be extended to various other domains. The ability to generate and distribute a large number of alternatives on a wall-sized display can be applied to many other ideation or data exploration scenarios involving parameter variations. Providing users with both individual and shared virtual content in a 3D space can be useful in many creative applications. Allowing users to navigate individually in a virtual environment while maintaining the capacity to restore spatial consistency between their relative positions can be valuable in any collaborative virtual reality application that involves physical or tangible interaction among users. Moreover, the design processes we employed can be applied in other contexts for customizing the interaction techniques. For example, the prototyping methods used to design interaction on the wall-sized display could be beneficial to adjust the interaction in a wide variety of applications.

This research contributes to the development of novel interaction paradigms for large interactive spaces. However, interacting with such systems is still new to users and not easy to learn and understand. Many challenges remain in standardizing the interaction techniques and making them easier to discover. Given the various types of devices available, ranging from virtual reality headsets to large wall-sized displays, it is crucial to design consistent interaction techniques that allow users to interact seamlessly across the mixed reality continuum [Mil+95; SSW21]. Standardized techniques should avoid users having to relearn the whole set of interaction mechanisms every time they switch devices.

This research also investigates the design of collaborative systems with dedicated collaboration features and demonstrates their benefits for co-located collaboration. However, further evaluations are needed to comprehensively assess these systems given the wide variety and complexity of collaborative scenarios. In addition, future work should consider users with different levels of expertise and explore solutions for adapting interaction to this expertise. Lastly, collaborative systems are now increasingly using hybrid configurations, including both co-located and remote users. Consequently, the proposed collaborative interaction techniques must be extended to support remote users.

4

COLLABORATION AND AWARENESS ACROSS REMOTE SPACES

Large interactive spaces provide new opportunities for remote collaboration as they can connect distant users and create a shared collaborative space. This allows collaborators to interact while being remote and leverage the benefits of each other's interactive systems. Moreover, the large visualization and physical spaces available in such systems offer a wide range of possibilities for enhancing awareness and communication among these users.

Nevertheless, these collaborative environments raise many challenges, including both technical aspects and issues related to interaction and awareness among remote users. Firstly, technical solutions are needed to allow data sharing and collaborative interaction at a distance. It is especially important to handle users with heterogeneous devices and asymmetric setups. Secondly, these systems must also facilitate understanding between users, as distance and technology can alter awareness and communication among them. In particular, large interactive spaces require suitable means of representing remote users, showing their actions and interaction capabilities, as well as transmitting non-verbal communication cues. These solutions should take advantage of these systems to go beyond reproducing the standard face-to-face collaboration that happens when no technology is involved, as proposed by Hollan & Stornetta in their article "Beyond Being There" [HS92].

In this chapter, I present my research on remote collaboration and telepresence systems. The first section focuses on different technical aspects involved when connecting heterogeneous interactive spaces, ranging from wall-sized displays to immersive virtual reality systems. For each aspect, I describe how the technology can be leveraged to support an effective collaboration. The second section explores how video-mediated communication can enhance awareness among remote collaborators. In this section, I detail the design of telepresence systems covering various forms of collaboration, including one-to-one collaboration, one-to-many collaboration or collaboration between users of immersive and non-immersive technologies.

4.1 CONNECTING HETEROGENEOUS SPACES

Remote collaboration across large interactive spaces cannot become widespread if it requires all users to have the exact same physical devices, especially given the wide range of devices currently available. My goal is to design collaborative systems that can accommodate users with heterogeneous devices and asymmetrical setups. In particular, I want to take advantage of the asymmetrical interaction capabilities to foster new collaboration strategies, as presented in our position paper [**Fle+15b**].

This section mainly focuses on the technical aspects of connecting remote users across heterogeneous platforms and enabling communication among them. The first subsection explores data sharing and demonstrates how immersive and non-immersive spaces can be interconnected to support collaboration on computer-aided design (CAD) data in the context of industrial design. This work was published at

the *3DCVE workshop* at IEEE VR 2018 [**Oku+18b**] and in a book chapter [**Oku+21**]. The second subsection concentrates on 3D audio for transmitting and rendering remote users' voices. It proposes various spatial audio mappings to connect remote spaces with different sizes and shapes. The related system was presented at the Web Audio Conference 2018 [**Fyf+18**]. Finally, the last section proposes a method for reconstructing live 3D models of users' heads and transmitting them to remote locations. Such models can be used to create and animate realistic avatars of remote users in immersive telepresence or virtual reality systems. This method appeared at Eurographics 2014 [**Fle+14**].

4.1.1 *CAD data synchronization for collaborative modification*

Connecting remote users across heterogeneous platforms can offer several advantages in an industrial design process. It allows multidisciplinary experts to work together despite being located in different branches of a company, but also provides them with specific systems tailored to their needs. For example, style designers may require a large, high-resolution screen to explore and compare multiple alternatives, while ergonomists may prefer an immersive VR system to view the product in context. However, sharing computer-aided design (CAD) data across heterogeneous platforms and modifying it in real time are challenging. While modifying CAD parameters from virtual environments is complex, managing collaborative modifications is even harder, as it requires additional synchronization mechanisms. Our goal was to create a distributed system allowing remote users to modify together native CAD data across heterogeneous platforms.

While distributed systems for collaborative virtual environments have been studied since the 1990s in academic research [**Fle+10c**], only a few studies have addressed collaborative product reviews [Lei+96; LD97] and collaborative VR-CAD applications [Mah+10; AGoo]. However, these systems do not support collaborative modification of CAD-part parameters. The *Multi-Agent System* [Mah+10] allows engineers and ergonomists to manipulate the position and orientation of CAD objects across a VR platform and workstations, but it does not allow the shape of the object to be modified through its parameters. *DVDS* [AGoo] enables users to create a 3D model with hand gestures in a virtual environment, but it relies on a dedicated CAD system, and does not implement a distributed architecture to share this model across remote platforms. In the previous chapter, I described the design of two interaction techniques that enable non-CAD experts to modify native CAD data in large interactive spaces without using conventional CAD software. *ShapeCompare* (Section 3.1.2) facilitates the generation and visualization of numerous design alternatives on a wall-sized display, while *ShapeGuide* (Section 3.2.1) allows users to deform CAD objects through physical actions in an immersive VR system.

Building upon this work, we created a distributed architecture that synchronizes CAD data across remote platforms and deals with collaborative modification [**Oku+18b**; **Oku+21**]. This architecture is based on an external server, named *VR-CAD server*, which provides centralized access to native CAD data by embedding the CAA API of CATIA V5[1] (Figure 4.1). This server loads and modifies CAD data according to users' requests, using a *labeling* concept [CB04]. The *labeling* maintains a direct link between the 3D geometries displayed in each platform and

1 https://www.3ds.com/products-services/catia/

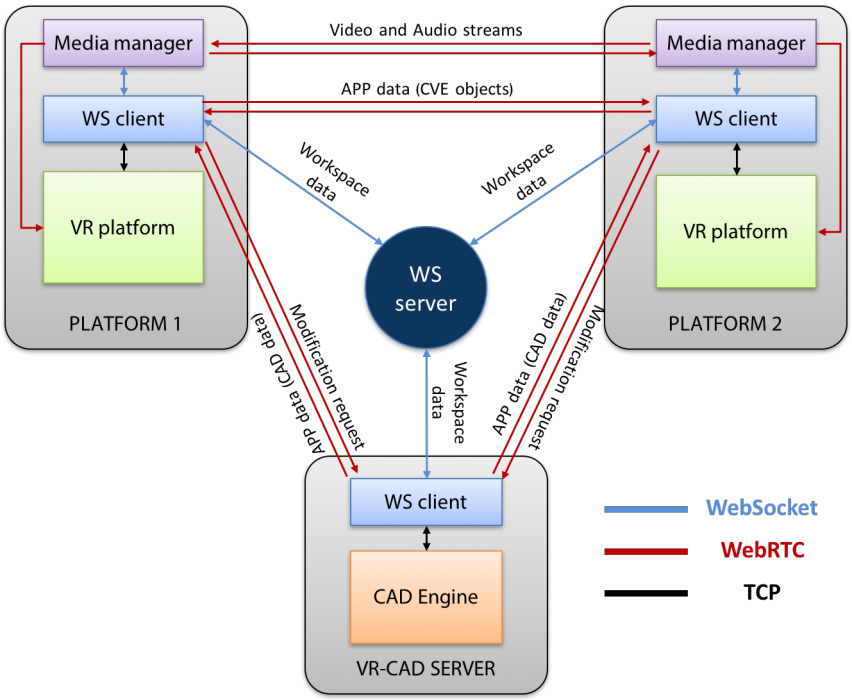

Figure 4.1: Distributed architecture for collaborative CAD data modification across remote platforms: the *VR-CAD server* is responsible for loading, modifying and synchronizing CAD data, while the *WS server* and *WS clients* manage connections between each platform and with the *VR-CAD server*.

the corresponding CAD parameters. Consequently, when users wish to modify a CAD object by interacting with its geometry, the *VR-CAD server* can retrieve the parameters to be modified and generate the desired shape. Once the modification request has been processed, the server sends back the tessellated meshes and a new *labeling* file to each platform, thus updating the visualization.

We used a hybrid network architecture to handle connections between each platform and with the *VR-CAD server*. A centralized architecture connects each platform to the *VR-CAD server* and manages CAD data synchronization. Additional peer-to-peer connections allow fast communication between platforms for all the other data types, including audio and video streams. To implement this architecture, a *Workspace (WS) client* deals with the network communication on each platform and on the *VR-CAD server*. A *WS server* handles authentication and initialization of the connections between all the *WS clients*, but direct peer-to-peer connections are used between *WS clients* to transmit data with the WebRTC[2] protocol. Since the communication layer is independent from the platform technical specifications, this architecture can connect heterogeneous platforms with various visualization systems and interaction devices.

The *VR-CAD server* supports both independent and cooperative modifications of the CAD data. Independent modifications enable several remote users to act on different CAD parameters simultaneously. When users modify multiple parameter values at the same time, the *VR-CAD server* processes the modification requests in the order they are received, and updates the CAD object on all platforms, regardless of the other ongoing modifications. Although this can be a little confusing for users,

2 https://webrtc.org/

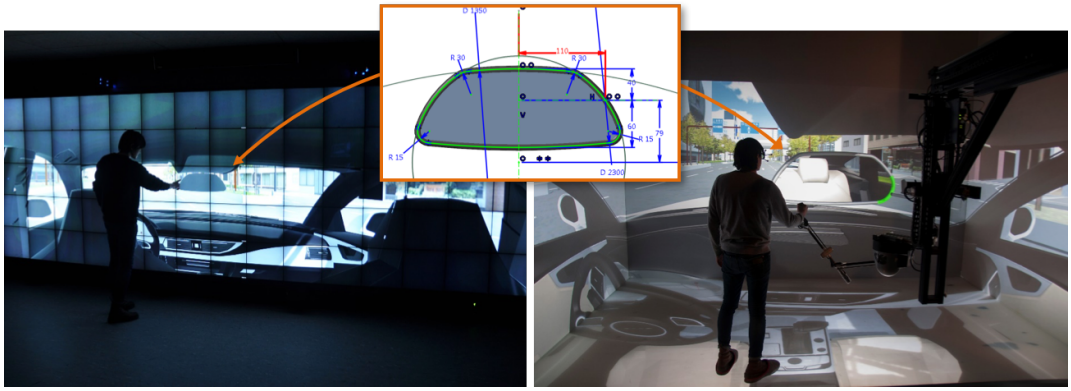

Figure 4.2: Collaborative modification of a car rear-view mirror between (left) a wall-sized display and (right) an immersive VR system. The *VR-CAD server* synchronizes the CAD data of the mirror between both platforms.

it can be effective if they coordinate well and modify complementary parameters. For example, one can modify the radius of a cylinder, while another can change its length. Cooperative modifications allow several remote users to modify the same CAD parameter simultaneously. In this case, the *VR-CAD server* manages concurrent modifications by using a dedicated concurrency control mechanism. In this first prototype, we simply used an averaging technique to combine the parameter values of each user, as proposed by Ruddle et al. [RSJ02]. However, future implementation could explore more sophisticated techniques, such as those studied in previous work [PBF08; ADL10a]. Such cooperative modification can be useful to help remote experts with divergent design constraints negotiate the shape of the CAD object through interaction.

As a proof of concept, we used this distributed architecture to implement collaborative modification of CAD data between a wall-sized display and an immersive VR system (Figure 4.2). We chose these two remote platforms because they have different visualization and interaction features. The wall-sized display has a high-resolution touch screen controlled by a cluster of 10 computers (see description of the WILDER system in Section 2.1.1). The VR system is composed of four large stereoscopic screens controlled by a cluster of 5 computers and a haptic device mounted on a carrier which enables users to interact everywhere in the system (see description of the EVE system in Section 2.1.2). The two platforms are located in remote buildings and connected to separate LAN networks. We set up the architecture by connecting a *WS client* to the master node of each platform and to the *VR-CAD server*. When the master nodes receive CAD data from the *VR-CAD server*, it still has to replicate this data on the slave nodes of the cluster.

We explored a collaborative design scenario where users could benefit from the asymmetric interaction capabilities to collaboratively modify the native CAD data of a car rear-view mirror. A team of style designers could use *ShapeCompare* to quickly generate various alternatives of the rear-view mirror on the wall-sized display, while an ergonomist could use the VR system to sit in the virtual car cockpit and review these alternatives. This enabled the ergonomist to assess live visibility in the rear-view mirror under realistic driving conditions. The ergonomist could also fine-tune the design of an alternative in VR by using the haptic device and the *ShapeGuide* technique. Additionally, we explored a second scenario in which two

users perform concurrent modifications of the rear-view mirror within the virtual car cockpit from both the wall-sized display and the VR system (Figure 4.2). Each user was able to modify the mirror shape by pushing or pulling on its surface using the *ShapeGuide* technique. The user in front of the wall-sized display used the finger on the touch screen, while the user in the VR system used the haptic device. When they modify the same CAD parameter, the *VR-CAD server* manages the concurrent modifications as described previously.

In summary, this work mainly focused on the technical aspects of connecting heterogeneous platforms and sharing native CAD among them. We proposed a hybrid distributed architecture that supports collaborative modifications of native CAD data from remote platforms. The CAD data is distributed and synchronized through a dedicated server, while other data and media streams are directly shared between platforms through peer-to-peer connections. We successfully implemented a proof of concept between a wall-sized display and an immersive VR system. Future work should further evaluate the benefits of such an asymmetric system in real collaborative design situations. We should also focus on improving collaborative interaction and providing appropriate feedback of the other users' actions.

4.1.2 *Spatial mapping for 3D audio communication*

When connecting remote users located in large interactive spaces, transmitting voice is a crucial aspect of communication. Using spatialized 3D audio for rendering voices can provide users with additional cues regarding the positions and activities of their remote collaborators. To achieve this, we can map the positions of voice sources to the actual 3D positions of the remote collaborators within their interactive system. However, mapping remote audio spaces with the local 3D space becomes challenging when the interactive systems differ in size and shape. It introduces additional complexities when systems are asymmetric or use different visual representations of the remote collaborators, such as avatars in a virtual environment as opposed to video feeds on 2D displays. Our goal was to create a technical system capable of transmitting audio along with users' 3D positions and rendering spatialized sound, in order to enable us to explore various spatial audio mappings across remote heterogeneous spaces.

Early work investigated spatial audio in telepresence systems [HRB97] and studied binaural audio in such a context [CAK93]. Binaural audio involves recording and rendering distinct sounds for each ear to replicate 3D audio as experienced by users in a real environment. To perceive this binaural audio accurately, users must use headphones. Similarly, Keyrouz and Diepold [KD07] employed binaural audio to allow a teleoperator to perceive the sound of a remote environment in 3D. However, these studies focused on the technical aspects of audio capture and rendering. They also assumed a one-to-one mapping between the recording environment and the rendering space, without exploring alternative mappings.

We created a telepresence system that records users' voices and 3D positions, transmits this data to remote platforms and renders spatialized sound using binaural audio feedback [**Fyf+18**]. In each platform, all users are equipped with wireless microphones and headphones, allowing them to move freely within the system while communicating. Voices are captured through an audio interface and sent to

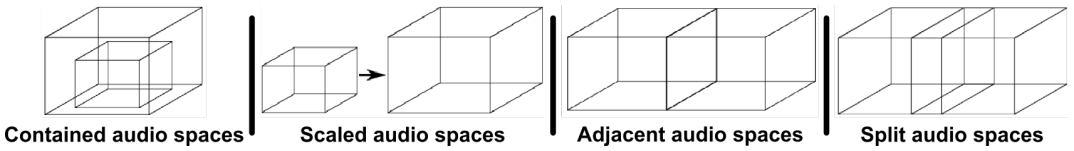

Figure 4.3: Multiple mappings between remote audio spaces of heterogeneous systems.

a media server based on *Kurento*[3]. Users' positions in the 3D space are captured by a *VICON* infrared tracking system. All users wear reflective markers, on their headphones for example. These markers identify each user individually. Audio and 3D positions are transmitted, along with the video, to remote platforms using the WebRTC[4] protocol. The media server receives remote audio streams with their associated 3D positions, and computes the binaural rendering using the *Audiostack*[5] software. It uses both the remote users' positions to compute the voice source locations and the local users' positions to compute the proper binaural audio feedback corresponding to their position and orientation. Finally, *Audiostack* provides users with appropriate audio feedback through their headphones. As a result, users perceive their remote collaborators' voices as coming from specific 3D locations. These locations remain consistent even if the users move or turn their head.

This spatialized audio feedback creates 3D audio spaces that are mapped with the local 3D space of the interactive system, and vice versa in the remote locations. This mapping can be modified or distorted by changing the audio space positions in the local reference frame or by altering the sound quality, with attenuation effects for example. We explored various mapping between audio spaces that are useful to connect heterogeneous spaces with different sizes and shapes (Figure 4.3):

- **Contained audio spaces:** when two remote spaces have different sizes, one obvious solution is that the smaller space is contained within the larger one. The smaller one can be positioned anywhere inside the larger one, allowing a specific placement of the remote collaborators inside the larger space. Although this solution offers real-scale mapping of the two spaces, it can be frustrating for the users in the smaller space to hear others from far away and not be able to join or follow them.

- **Scaled audio spaces:** a second solution when two remote spaces have different sizes is to scale the smaller space to the size of the larger one, and vice versa. This configuration can be useful when the two spaces have the same virtual content displayed on 2D screens of different sizes. The locations of user voice sources are thus consistent with positions relative to the virtual content. However, the speed at which voice sources move may not match the real displacements of the users in the remote location.

- **Adjacent audio spaces:** if we do not want the audio spaces to overlap, they can also be virtually placed adjacent to each other, creating a larger audio space. This configuration works well for telepresence systems with large screens showing the remote locations. It can thus give the feeling that the remote collaborators are "on the other side" of the screens.

3 https://kurento.openvidu.io/
4 https://webrtc.org/
5 https://www.aspictechnologies.com/audiostack/

- **Split audio spaces:** a mixed solution consists in overlapping only a subpart of the audio spaces, thus creating different areas within the audio space that may or may not be shared. This configuration is relevant to support different moments in collaboration, and allows users to move around depending on who they want to talk to. For example, users can stay in the shared area to talk with the remote collaborators, then move to the non-shared area when they want to have side discussions with their local collaborators.

- **Distorted audio spaces:** the previous configurations assume that the positions of the voice sources in the local space match the position of the corresponding remote collaborators in the remote space, modulo the possible translation, rotation or scaling required by the configuration. However, the mapping can be further distorted or changed entirely. For example, some systems can display video from remote users in small windows on larger screens, or on mobile devices such as tablets or telepresence robots. In this case, the position of voice sources can be made consistent with the position of the related video streams. Distorted audio spaces open up a wide range of possibilities.

To conclude, we have proposed a technical system that combines audio streaming, motion tracking and spatialized binaural audio in the context of remote collaboration across large interactive spaces. This system allows transmitting users' voices along with their respective 3D positions, and rendering voice sources spatialized within the 3D space of remote locations through binaural feedback. To handle heterogeneous remote platforms, we proposed various mappings between the remote audio spaces and the local 3D space of each platform. This is a preliminary work on how to enhance remote collaboration by customizing these audio mappings. This concept needs to be refined and evaluated in various technical and application contexts. Although the proposed mappings theoretically extend to more than two platforms, we have only tested them in this simple configuration. We must also investigate their impact on collaboration and, especially, how they can support different collaborative dynamics, such as interrupting others' activities, engaging in discussion, initiating side discussions, or transitioning from tightly-coupled to loosely-coupled collaboration.

4.1.3 *3D head reconstruction for immersive telepresence*

Appropriate visual representations of remote users are essential for collaboration across large interactive spaces. They can convey non-verbal cues that are essential for communication, such as eye gaze direction, facial expressions and gestures. However, not all visual representations are suitable for all types of interactive systems when connecting heterogeneous platforms. In particular, video is not well suited to immersive virtual reality systems and 3D displays due to its 2D nature. Avatars can be used in such systems, although facial expressions and eye gaze are usually poorly represented. In this work, we aimed to reconstruct a live 3D model of the users' head to improve avatars and better convey facial expressions and eye gaze to remote collaborators. To easily adapt the proposed system to a wide range of interactive spaces, we targeted a simple solution based on a single consumer level hybrid sensor capturing both color and depth.

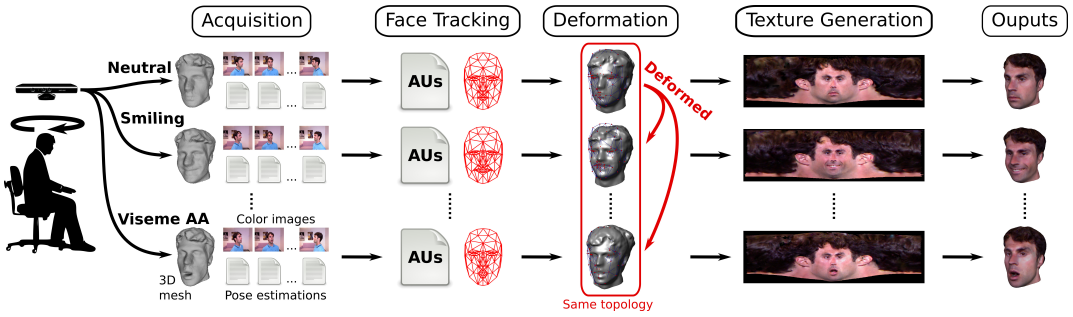

Figure 4.4: Acquisition step of the 3D head reconstruction: data are captured and processed to create a complete and fully textured 3D head model for each facial expression.

3D head reconstruction has been widely studied in the literature, as detailed by Pighin and Lewis [PL06]. Some techniques achieve very accurate 3D models using a large set of high-resolution cameras [Bee+10] or structured lights [Zha+04]. They can also provide an animated version of such head models [Bee+11]. Nevertheless, these techniques require an expensive and complex equipment. At the time we carried out this research, they did also not operate in real time which was not suitable for remote collaboration. Other techniques achieved real-time reconstruction by fitting a deformable face model to the depth data of the users' face captured by a depth sensor [Wei+11; Li+13]. This deformable face model is usually created from a large database of human face scans. As a consequence, it does not fit the specific appearance of the users' head because hair, eyes and interior of the mouth are missing. A colored texture of the face can be generated [Wei+11], but it is static and inconsistencies appear for small face features, including eyes, teeth, tongue or wrinkles. Consequently, these techniques are unable to properly convey facial expressions, which are crucial for non-verbal communication.

We proposed a 3D head reconstruction method that animates the model in real time and makes it suitable for remote collaboration [**Fle+14**]. It uses only a single consumer level hybrid sensor capturing color and depth, such as the Microsoft Kinect used in our implementation. This sensor has to be located in front of the users and does not require any calibration, which makes it easy to install in large interactive spaces. However, this type of sensor provides noisy and incomplete data due to poor sensing quality and occlusions. Our method fuses the noisy and incomplete real-time output of the sensor with a set of high-resolution static textured models captured offline in a preliminary step. The method is decomposed into two steps: an acquisition step that captures and pre-processes data, and a reconstruction step that reconstructs the head model in real time.

For the acquisition step (Figure 4.4), users must spin on a chair in front of the sensor and display different facial expressions during each turn. These expressions include visemes, as well as other variations such as a neutral expression, open mouth, smile and raising or lowering eyebrows. For each facial expression, we use the KinectFusion algorithm [Iza+11] to generate a 3D mesh of the head along with a set of color images captured from different angles around the head. Each image is accompanied by its camera pose estimation relative to the head position. We also use the face tracker from the Microsoft Kinect SDK [Cai+10] to track the head and to characterize the related facial expression in each data set. The tracker provides us with a set of descriptors that are stored with each 3D mesh. After the data

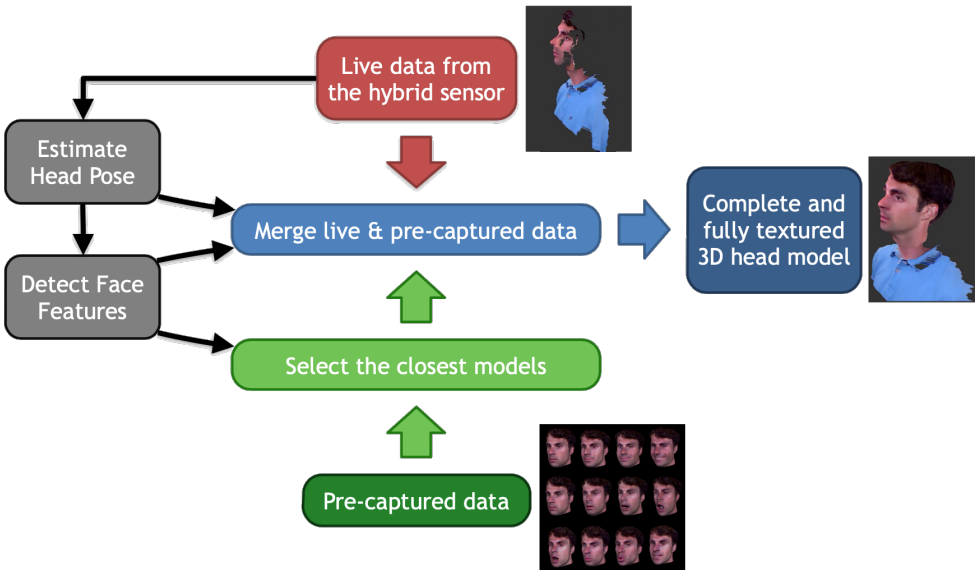

Figure 4.5: Live reconstruction pipeline of the users' 3D head: color and depth data from the sensor are combined with the pre-captured data to create a complete and fully textured 3D model.

capture, we deform the neutral face mesh to fit the other meshes using a method based on cross-parametrization [KS04]. This produces a set of deformed meshes, all with the same topology, which can later be smoothly interpolated to match the current facial expression during the reconstruction step. Finally, a cylindrical texture is generated for each facial expression by combining the color images using a cylindrical projection . We use alignment and smoothing processes to compensate for inaccuracies in camera pose estimations and disparities in color balance and lighting across images. The output of the acquisition step consists of a set of 3D meshes that share the same topology, along with their corresponding cylindrical texture and facial expression descriptors.

For the reconstruction step (Figure 4.5), users need to stand in front of the sensor during the remote collaboration session, as in any videoconferencing systems using a camera. We use again the face tracker from the Microsoft Kinect SDK to estimate head pose and detect users' current facial expression. The descriptors detected by the tracker are then compared with the ones stored with the pre-captured 3D meshes by computing the Euclidean distance. 3D meshes that correspond to the closest facial expressions are selected and interpolated to create a new 3D mesh that matches the current facial expression of the user (Figure 4.6-b). In our implementation, we chose to select the two closest meshes, but it is possible to select more if needed. The cylindrical textures associated with the selected meshes are also interpolated to create a new static cylindrical texture (Figure 4.6-c). This static texture allows us to have a 360° texture of the head with a better resolution than the images directly captured from the sensor. However, the dynamic facial features, such as the eyes, mouth or wrinkles, are not consistent with the users' current face. Therefore, we propose to use the static texture as a background, but to combine it with a dynamic texture extracted from the sensor video stream, which provides the salient features of the face. A gradient is used to extract these features from the dynamic texture (Figure 4.6-d) and, conversely, smooth such features in the static texture. The final texture is obtained by merging these two textures

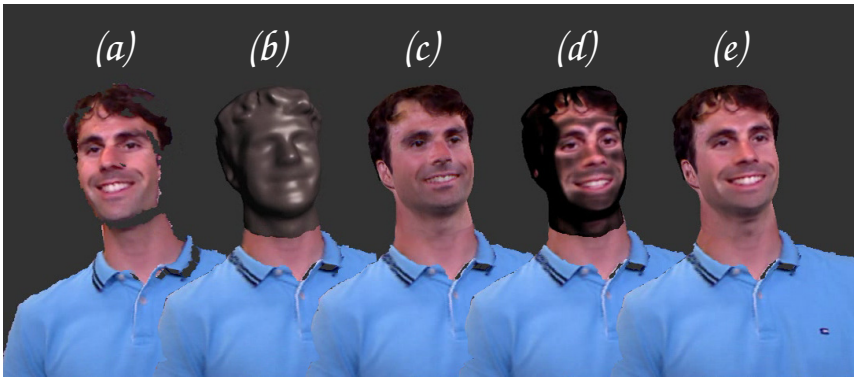

Figure 4.6: Live reconstruction steps: (a) raw data from the sensor and (b) mesh interpolated from the 3D meshes with the closest facial expressions. This new mesh is then textured with (c) the static cylindrical texture, (d) the dynamic texture showing only the salient features, and (d) both textures combined.

(Figure 4.6-e). As a result, a complete and fully textured 3D model of the users' head is reconstructed in real time, accurately preserving facial expressions.

We obtained promising preliminary results showing that our method is able to handle a wide range of subjects and different lighting conditions. Although the 3D mesh was not precise enough to capture small face features, we observed that the dynamic texture can compensate for this limitation. This suggests that integrating elements from live video can be valuable in the context of remote collaboration. Further studies would be required to evaluate our method on a larger scale and better understand which parameters are the most important for communication between remote users. In particular, it would be interesting to assess how the texture quality and 3D mesh accuracy impact the perception of facial expressions. Moreover, our method is well suited to remote collaboration, as the amount of data transmitted over the network is relatively low. Once the acquisition step and system initialization are complete, only the color and depth video streams need to be sent, and the 3D reconstruction can be performed remotely. Finally, our method is highly dependent on the results from the face tracker and most failures occur when it cannot accurately detect facial expressions. This mainly happens when users look away from the camera. However, this could be improved by using a more accurate tracker as our method does not rely solely on this tracker.

In summary, we proposed a method for reconstructing and animating 3D head models of remote users. It relies on a single consumer-level hybrid camera which captures both color and depth. The key features of this approach are an interpolation of pre-captured 3D meshes corresponding to different facial expressions, and a fusion of static and dynamic textures to respectively enhance resolution and incorporate dynamic features extracted from the live video. This work was conducted ten years ago from the writing of this manuscript. Hardware and software solutions have improved dramatically in recent years, especially with the use of machine leaning techniques and the creation of large data sets of human faces. It is now possible to achieve much higher quality results in real time, as proposed by *Pixel Codec Avatars* [Ma+21] for example. However, the closer avatars become to human aspect, the more they raise concerns related to the Uncanny Valley [MMK12]. We must be careful when choosing user representations depending on the application context and further explore the impact of these representations on collaboration.

4.2 ENHANCING AWARENESS WITH VIDEO-MEDIATED COMMUNICATION

Video-mediated communication has long since demonstrated its considerable strengths in remote collaboration systems, as detailed in Section 2.3.1. Video-mediated communication can also be valuable for enhancing users' awareness across remote large interactive spaces, as presented in our position paper [**Fle+15a**]. However, most previous systems are designed for meetings where users sit around a conference table, relying on video as a substitute for face-to-face conversation. These systems do not support large spaces where users move around and work on shared data. Previous work on Media Spaces [BHI93; Mac99] has created systems that support peripheral awareness, chance encounters, locating colleagues and other social activities. Nevertheless, Media Spaces have not explored setups where distributed groups work on shared data in large interactive spaces.

This section explores the design of telepresence systems for large interactive spaces, focusing mainly on non-immersive systems. The first subsection addresses collaboration across wall-sized displays and investigates how to capture and where to display video in such systems. A first perceptual study was published at CHI 2015 [**AFB15**], while the main part of this work was published at CHI 2017 [**Ave+17**]. The second subsection aims to integrate a remote user into a co-located group collaboration, properly conveying gaze direction. These results appeared at INTERACT 2019 [**Le+19**]. Finally, the last subsection details how a laptop or desktop user can collaborate with a remote collaborator wearing an augmented reality headset, taking advantage of multiple video viewpoints on this remote collaborator and the augmented content. This work was published as CSCW 2022 [**FFT22b**] and the related system was demonstrated at IHM 2022 [**FFT22a**].

4.2.1 *Telepresence across wall-sized displays*

Large wall-sized displays are powerful tools for supporting co-located group collaboration, but they can also accommodate remote users by connecting other wall-sized displays. Video-mediated communication is crucial in such remote collaborative scenario to enhance awareness and mutual understanding among users, as previously discussed. Some previous work investigated telepresence systems across two wall-sized displays. Most of these systems aim to display the remote video feed using all the available screen space, creating the illusion of having a glass between the two remote spaces [Wil+10; Dou+12]. However, this does not support collab-

Figure 4.7: Two remote wall-sized displays showing the remote collaborator's video, along with the same content.

oration on shared digital content. Luff et al. [Luf+15] introduced a telepresence system that supports remote collaboration on shared digital objects. This system preserves the physical relations between video and digital objects, allowing users to understand where collaborators are looking or pointing at. Nevertheless, this system relies on a circular configuration of the screens, requires the exact same setups at both locations, and exclusively mimics co-located collaboration without the flexibility to go beyond physical constraints.

Our goal was to design a telepresence system connecting two distant rooms equipped with wall-sized displays showing shared content (Figure 4.7). We explored how this system can combine the shared task space with the shared person space, as defined by Buxton [Bux92]. The former refers to the ongoing task, involving actions such as making changes, annotating and referencing objects. The latter refers to the collective sense of co-presence, involving facial expressions, voice, gaze and body language. Buxton [Bux09] defines the overlap between these two spaces as the reference space, where *"the remote party can use body language to reference the work"*. We first investigated this reference space on a wall-sized display and assessed how accurately users can interpret deictic gestures in a remote video feed. We then explored how to capture and where to display the video feeds on the wall-sized displays by creating a telepresence system based on camera arrays embedded in the display. Finally, we evaluated this system on two collaborative tasks.

4.2.1.1 *Study of deictic gestures*

Referencing shared objects is crucial to support mutual understanding and effective collaboration [Mac99]. Video-mediated communication can affect users' ability to correctly perceive deictic instructions due to technological limitations, including camera and video placements, lens distortion and latency. We focused on a scenario where two wall-sized displays share the same content and simultaneously display a remote user's video feed at the same relative position as the recording camera at the remote location (Figure 4.8). Our objective was to investigate users' ability to determine accurately which shared object the remote user is referencing, without the need for dedicated technology such as telepointers.

While some previous studies have assessed the accuracy of direct eye contact in video-mediated communication [Che02], none of them focused on the accuracy

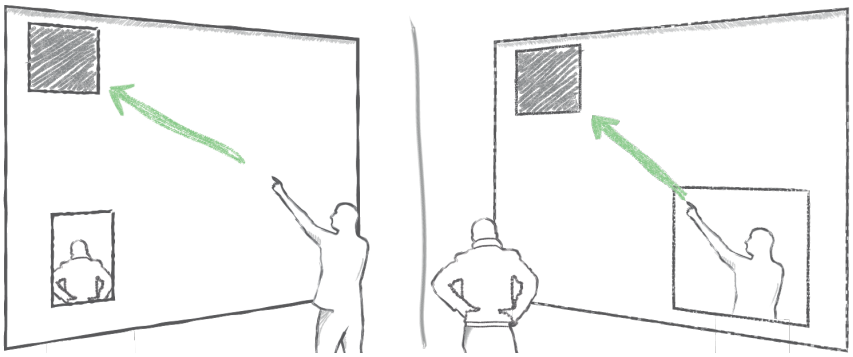

Figure 4.8: Users working on shared objects across two remote wall-sized displays: (left) a user shows a shared object by pointing at it and (right) the remote user tries to understand which object is being pointed through the video.

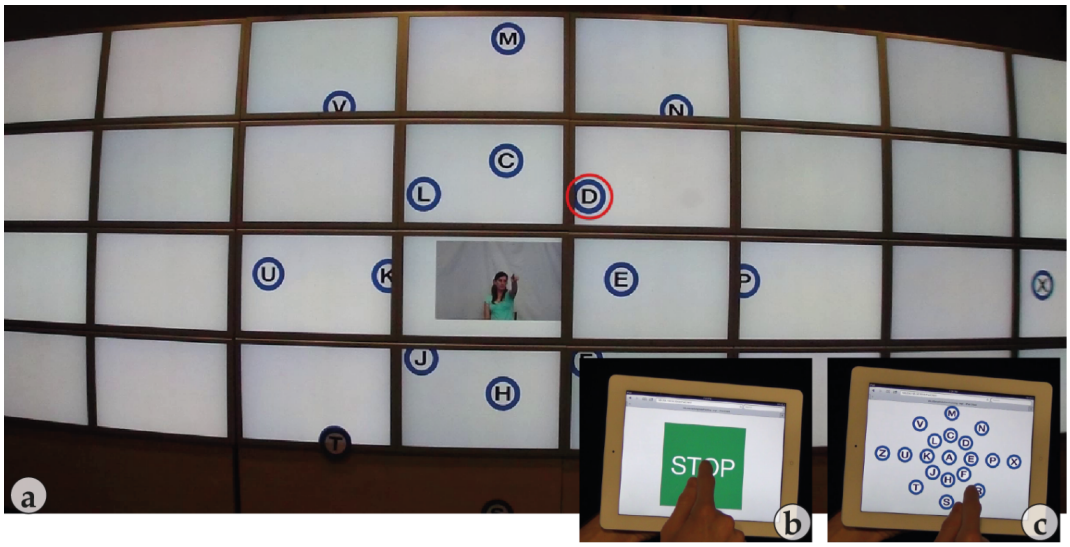

Figure 4.9: Experimental setup from the participants' point of view: (a) the remote user is pointing at target "D" (only highlighted in this figure). To answer, participants first (b) pressed the "STOP" button and then (c) selected the corresponding targets on the tablet.

of remote pointing. Wong & Gutwin [WG10] assessed pointing accuracy, but in a collaborative virtual environment. Users were represented by avatars, which is very different from live video feeds. However, they noted that determining how accurately viewers can interpret pointing direction is a fundamental question to explore before designing support for pointing in remote collaboration systems.

We conducted a controlled experiment to study (i) how accurately participants perceive a reference to a shared object performed by a remote user, either by looking at it or pointing at it with the hand, and (ii) whether the participants' position in front of the wall-sized display influences this accuracy [**AFB15**]. 12 participants looked at a large number of videos of the remote user referencing a specific target on the wall-sized display (Figure 4.9-a). To avoid bias in the experiment, we used pre-recorded videos of three actors playing the role of the remote user. For each video, participants had to indicate on a tablet which target was referenced by the remote user (Figure 4.9-b,c). We controlled three factors:

- 2 techniques used by the remote user to indicate the target: HEAD combines natural head rotation and gaze, while HEAD+ARM combines natural head rotation, gaze and pointing with the arm and finger.
- 5 positions of the participants in front of the display: CENTER located in front of the video, FARLEFT, LEFT, RIGHT and FARRIGHT respectively located at $2m$ on the left, $1m$ on the left, $1m$ on the right, and $2m$ on the right.
- 19 targets on the wall-sized display, arranged in 8 directions and 3 distances from the central target.

To analyze the results, we decomposed the errors into two measures since the targets were arranged in a circular pattern around the video: distance error and angle error. The unit of distance error is normalized, so that an error of one corresponds to one target closer or further from the center, relative to the designated

target. The unit of angle error is in degrees, so that an error of 45° corresponds to one target next to the designated target. First of all, the errors are relatively small, with an overall mean of 0.34 ± 0.52 for distance error and 3.90 ± 12.04 for angle error. This suggests that participants were generally able to accurately identify the referenced target. For the techniques, the distance error is not significantly different when using HEAD or HEAD+ARM, but the angle error is significantly larger when using HEAD+ARM compared to HEAD. Although the effect size is small (1.65°), this result was unexpected. After analyzing the videos, we noticed that the direction of the arm does not always indicate the correct target. In fact, people place the tip of their finger on the line of sight between their eyes and the target, as described in [HC02]. As the video is a 2D representation, it may be hard to perform this 3D geometrical interpretation for viewers, and could lead to errors. For the positions, we observed almost no significant effects of the relative position between the participants and the video on accuracy. This effect on deictic gesture perception is analogous to the *Mona Lisa effect* observed for gaze. The *Mona Lisa effect* describes the fact that the video of a subject looking at the camera is perceived by remote users as looking at them, regardless of their position. At the extreme positions FARLEFT and FARRIGHT, we still measured a slightly higher angle error, but this can be explained by the fact that the observers are looking at video with an angle of 49°, making the task harder.

In conclusion, we assessed how accurately users perceive deictic gestures through video when sharing digital content across remote wall-sized displays. This study shows that users can accurately identify the referenced object, that eye gaze alone can be more accurate than finger pointing, and that the relative position between the viewer and the video has minimal effect on accuracy. Based on these findings, we have derived the following implications for designing future telepresence systems suitable for remote collaboration across wall-sized displays:

1. Additional technical features are not always mandatory to indicate digital objects, as users can accurately interpret gaze and arm pointing. Telepointers and extendable arms [Hig+15] may not necessarily be required if the video is positioned consistently with the content.

2. The arm and gestures are not always needed to indicate digital objects, as users can rely solely on gaze. This allows users to perform deictic actions while holding other interaction tools in their hands.

3. Users can move in front of the wall-sized display, or the video feed can be moved along the display, as the relative position between users and video does not affect accuracy. We can thus consider manipulating the video position on the wall-sized display to meet the requirements of various collaborative tasks.

4.2.1.2 *Design of a telepresence system*

Based on the design recommendations from the previous study, we set out to create a telepresence system supporting video-mediated communication across wall-sized displays. The main challenge was to provide users with audio-video communication as they move in front of the display and interact with shared content.

To determine the optimal camera and video positions, we conducted preliminary observational studies using low-fidelity prototypes. We divided a wall-sized display

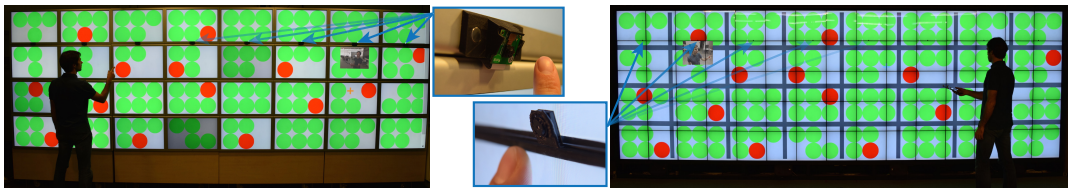

Figure 4.10: *CamRay* provides video-mediated communication between (left) WILD and (right) WILDER wall-sized displays. (Center) close-ups on the cameras embedded in the displays.

with a curtain to simulate two remote locations and simplify the technical setup. The first prototype included two tablets running videoconferencing software, each held by a helper to test multiple placements. Two participants had to create a slideshow presentation from their respective location. We noticed that they looked at the content much more than the video feed. In fact, they only looked at each other when they disagreed or needed to discuss a specific issue. After the debriefing with participants, we hypothesized that they might have looked at the video more often if it did not require switching screens and decided to place the video on the wall-sized display.

The second prototype used two cameras at each location: a front-facing camera attached to the screen and a back camera located at the back of the room, facing the screen. At each location, three video feeds were also displayed: on the left, a window displayed the remote front-facing video with a small thumbnail showing the local front-facing video, and on the right, another window displayed the remote back-facing video. Two participants had to sort research papers to prepare the related work section of a publication. Papers were laid out on the large screen at each location, with their position and current page synchronized. We observed that participants physically moved to a specific video window depending on the task at hand. They used the front-facing video to discuss paper content or how to cluster papers, while they used the back-facing video to understand which paper the other was refereeing or where the other was pointing. However, they had to interrupt their work to glance at video windows, which was perceived as annoying. We concluded that we should be able to capture users' faces even as they moved along the screen and to display the video feeds in a flexible manner. We identified two requirements for video placement, each corresponding to specific phases of the collaboration: one should support face-to-face conversations, while the other should support the use of deictic instructions.

To meet these requirements, we created *CamRay* [**Ave+17**], a telepresence system connecting remote wall-sized displays. We implemented a prototype of this system between the WILD and WILDER platforms, located in two different buildings (see descriptions of the two systems in Section 2.1.1). *CamRay* uses an array of eight cameras embedded in each display, capturing the users' faces (Figure 4.10). The cameras are equally spaced along the horizontal axis of both displays and located on the nearest bezel above users' eye level. We used *Raspberry Pi* camera modules, each one connected through a ribbon cable to a *Raspberry Pi*[6] located at the back of the display (Figure 4.11). Each *Raspberry Pi* captures video with a resolution of 800×600 pixels, encodes it in H.264 and streams it to a dedicated computer over UDP using

6 https://www.raspberrypi.org/

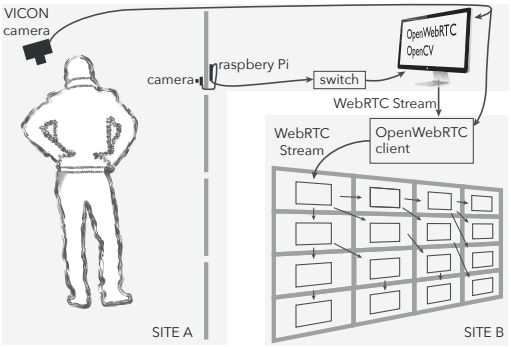

Figure 4.11: *CamRay* architecture transmitting video from site A to site B. A similar setup is used to transmit video from site B to site A.

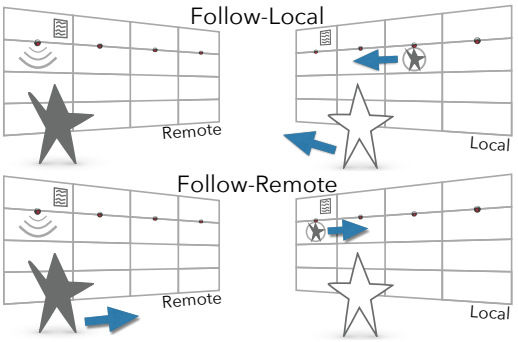

Figure 4.12: The two video modes of *CamRay*. The arrows show the participant whose position controls the video at the local site.

GStreamer[7]. Users' position is tracked by a *VICON* infrared tracking system and also sent to the same computer. This computer executes a C++ application based on *OpenCV*[8] to select the video feed of the camera in front of the user using tracking data. Finally, the application streams the selected video along with the tracking data to the remote wall-sized display using the WebRTC[9] protocol.

On the remote location, a server receives the WebRTC video stream along with the tracking data, and transmits it to the visualization cluster that controls the wall-sized display. Each node of the cluster runs a web application based on *NW.js*[10]. This application is able to display the WebRTC video stream and to forward it to other nodes. Only a specific node in the cluster receives the stream from the server, and forwards it to 2 or 3 other nodes, which in turn forward it to two or three other nodes, and so on (Figure 4.11). This tree pattern allows us to transmit the video stream to all cluster nodes with low latency and avoid overloading the server with multiple video streams. As a result, a video window showing a remote user can be displayed, spanning several screens if required, and moved all over the display. This window can appear on top of the content displayed on the wall-sized display. In addition, the server receives both the positions of the local and remote users, and can use this information to define the placement of the video window.

Based on the observational studies, we implemented two modes for positioning the video window on the wall-sized displays using *CamRay* (Figure 4.12):

- With *Follow-Local*, the video window follows the horizontal position of the local user, providing constant visual contact with the remote collaborator. This mode creates a virtual face-to-face, where both remote users are always visible to each other even when located at different positions in front of their respective displays.

- With *Follow-Remote*, the video window follows the horizontal position of the remote user, conveying his relative position to the shared content. This mode allows users to accurately interpret deictic instructions made by the remote

7 https://gstreamer.freedesktop.org/
8 https://opencv.org/
9 https://webrtc.org/
10 https://nwjs.io/

collaborator as the video window has a consistent position, according to the shared content.

In both modes, the video window does not move continuously, but is snapped under each camera of the array. The video is thus placed congruently with a camera, allowing direct eye contact between users. In addition, the video is horizontally mirrored to maintain a spatial consistency between the video and the shared content: a user looking to the left is therefore displayed as looking to the left in the video at the remote location. Consequently, the remote user is seen as standing behind the display, as in Clearboard [IK92]. Moreover, we do not display any feedback of the users' own video, as nobody used it during our observations. Some participants even reported that they trusted the system to capture them properly, since they were not responsible for adjusting the camera position.

The two proposed modes support different aspects of non-verbal communication, including eye gaze, facial expressions and gestures. In particular, Fussel et al. [Fus+04] distinguish two categories of gestures in video-mediated communication: *"pointing gestures, which are used to refer to task objects and locations, and representational gestures, which are used to represent the form of task objects and the nature of actions to be used with those objects"*. We hypothesized that each method is best suited to support different types of non-verbal cues: *Follow-Remote* consistently positions the video relative to the remote users' position and content, facilitating the accurate understanding of pointing gestures, while *Follow-Local* maintains constant visual contact, making it easier to perceive eye contact, facial expressions and representational gestures. To test this hypothesis, we ran two controlled experiments comparing both modes on two collaborative tasks. The first experiment studied *CamRay* during a data manipulating task that relies on pointing gestures. The second experiment assessed *CamRay* in a knowledge-sharing task that benefits from easy perception of eye contact, facial expressions and representational gestures.

4.2.1.3 *Evaluation on a data manipulation task*

In a first controlled experiment [**Ave+17**], we aimed to assess the ability of *CamRay* to properly convey pointing gestures between two remote wall-sized displays. To achieve this, we needed a data manipulation task that requires the production and interpretation of such gestures. We drew inspiration from the disc classification task designed by Liu et al. [Liu+14]. In a co-located collaborative situation [Liu+16], they explored a condition in which one participant instructed another on where to classify discs on a wall-sized display. They observed that this condition mainly relied on deictic instructions. We thus implemented a remote version of this condition. In this version, an *Instructor* had to determine how to classify discs and give instructions to a remote *Performer* who performed the manipulation (Figure 4.10).

Both wall-sized displays were divided into 32 containers, each capable of holding up to 6 discs. On the *Instructor*'s wall-sized display, discs were labeled with small letters which indicated how to group discs in containers. All the discs in the same container needed to have the same letter to be considered properly classified. Properly classified discs were highlighted in green, while misclassified discs were shown in red. The *Instructor* was not able to move discs. On the *Performer*'s wall-sized display, red and green discs were displayed without labels. The *Performer* was able to move discs with a pointing device.

12 pairs of participants performed the classification task with three video conditions: *Follow-Local*, *Follow-Remote* and a control condition, named *Side-by-side*. This control condition used a fixed video window on a separate screen on the left side, perpendicular to the wall-sized display. Each participant alternately assumed the roles of *Instructor* and *Performer* for each video condition, completing the task twice in each role. The experimental setup was composed of the WILD and WILDER wall-sized displays (see system descriptions in Section 2.1.1). We hypothesized that participants would perform the task faster and rely more deictic instructions with *Follow-Remote* than with *Follow-Local* and *Side-by-side*.

The main results demonstrate that participants classified discs significantly faster with *Follow-Remote* than with *Follow-Local* and *Side-by-side*. There are three reasons for this increase in performance. First, *Performers* followed the *Instructors*' position and gaze more closely, and were faster to drop the disc at the correct spot. In particular, the results revealed that the distances between the *Performers*' cursor and the *Instructors*' position, as well as between the *Performers*' cursor and the *Instructors*' estimated gaze point, were smaller with *Follow-Remote* than with the other conditions. Second, participants used more deictic instructions and fewer words with *Follow-Remote* than with the other conditions, reducing the time spent by the *Instructors* giving verbal instructions. Third, participants made fewer misunderstanding errors with *Follow-Remote* than with the other conditions, also reducing the time spent correcting errors. In addition to these results, qualitative feedback showed that a large majority of participants preferred *Follow-Remote* when playing the role of *Performers* (22/24), while *Side-by-side* was ranked first twice. Surprisingly, only half the participants preferred *Follow-Remote* when playing the role of *Instructors*, while *Follow-Local* was ranked first ten times and *Side-by-side* was ranked first twice. This preference might be due to the fact that *Instructors* liked seeing their remote collaborator's face as they gave instructions to check for understanding.

In summary, *Follow-Remote* proposes to display video consistently to the shared content, according to the remote user's position. The results demonstrate that participants were better able to understand deictic instructions with *Follow-Remote* than with the other conditions, reducing the overall cost of communication, as explained by Clark and Brennan [CB91]. As a consequence, *Follow-Remote* provides better performance on the data manipulation task. Nevertheless, some participants preferred the constant visual contact created by *Follow-Local* when checking for their collaborator's understanding. These potential benefits of *Follow-Local* should be further explored in tasks that involve more discussion and knowledge-sharing.

4.2.1.4 *Evaluation on a knowledge-sharing task*

While the first experiment focused on deictic instructions, this second experiment aimed to explore how *CamRay* could convey representational gestures, along with eye contact and facial expressions. We believed that the persistent face-to-face provided by *Follow-Local*, even when users move in front of the display, could be valuable for better perceiving these non-verbal communication cues. Our goal was to design a task involving discussion and knowledge-sharing that would benefit from these specific cues. We drew inspiration from a realistic scenario in which two experts have to combine their knowledge to resolve a problem.

We created a task in which an *Instructor* sees an image located at a random position on the wall-sized display and has to describe it to a remote *Performer*. At

Figure 4.13: (Left) sign language images and (right) Khmer characters used in the knowledge-sharing task to evaluate *CamRay*.

the remote location, the *Performer* has to search for this image among a set of 21 images spread all over the wall-sized display. In the task, the two wall-sized displays do not show the exact same content. This task mimics the scenario in which an expert retrieves and shares some knowledge with a collaborator, who has to find this information in a large data set. By using only images, the task eliminates the need for personal judgments or negotiation when choosing among possibilities. It also does not require participants to memorize or process information, thus mitigating potential biases in the experiment. To ensure that participants perform representational gestures, we conducted pilot tests with various types of images. We selected images of sign language and Khmer (Cambodian) characters (Figure 4.13). The sign language images are straightforward to describe because participants can replicate the hand poses. The Khmer characters are more difficult to describe and require a combination of gestures and speech.

6 pairs of participants performed this task with the same video conditions as in the first experiment: *Follow-Local*, *Follow-Remote* and *Side-by-side*. For each video condition, participants swapped roles and completed the task twice in each role: once with sign language images and once with Khmer characters. The experimental setup was also composed of the WILD and WILDER platforms. We hypothesized that participants would reach better performance and produce of more representational gestures with *Follow-Local* than with *Follow-Remote* and *Side-by-side*.

The overall results did not reveal strong effects of the video conditions in terms of performance, including task completion time and errors. We believe this is due to the fact that participants often decided to walk towards the video in *Follow-Remote* or *Side-by-side*, thus recreating the face-to-face condition. The results demonstrate that participants traveled longer distance with *Follow-Remote* and *Side-by-side* than with *Follow-Local*. They also synchronized more often their relative position with *Follow-Remote*. As a consequence, it is difficult to measure difference in terms of performance, as participants could potentially benefit from face-to-face conversation in all video conditions. We still noticed that participants made fewer errors with *Follow-Local* than with *Follow-Remote* for the sign language images. This difference could be explained by the fact that *Instructors* moved away from the described image in *Follow-Remote*, and sometimes forgot the exact hand gestures to perform, as describing this type of image relies heavily on representational gestures. Concerning the production of representational gestures, we did not observe significant differences overall. However, we were surprised to notice that *Instructors* used significantly more gestures with *Follow-Local* than with the other conditions when

replying to clarification requests. We hypothesized that participants spontaneously moved to create a face-to-face situation when they provided the initial explanation. But, when a clarification was required, participants were not always facing each other with *Follow-Remote* and *Side-by-side*. Finally, qualitative feedback shows that participants preferred *Follow-Local* for this task over the other two conditions, regardless of whether they were *Instructor* or *Performer*.

While the experiment does not reveal strong evidence that *Follow-Local* improves performance in a knowledge-sharing task, it does provide some hints that this condition encourages the production of representational gestures and makes their interpretation more accurate. The results also show that the technological hindrance is less pronounced with this condition. In particular, participants are not required to synchronize their position and need to travel less. As a result, participants preferred the *Follow-Local* condition for such a knowledge-sharing task and perceived a lower task load compared to the *Follow-Remote* condition.

4.2.1.5 *Summary*

In this work, we demonstrated in a first experiment that video feed displayed on a wall-sized display can properly convey deictic instructions, including pointing and gazing. Based on this observation, we designed *CamRay* to support video-mediated communicate across wall-sized displays. It embeds an array of cameras on each display to capture users as they move in the interactive space. We also proposed two methods for positioning the video feed on wall-sized displays according to local and remote users' positions: *Follow-Local* creates a virtual face-to-face, while *Follow-Remote* positions the video consistently relative to the shared content.

Two controlled experiments show that each method has its own advantages, making them suitable for different collaborative tasks. *Follow-Remote* supports deictic instructions for a data manipulation task, while *Follow-Local* supports representational gestures for a knowledge-sharing task. Nevertheless, these results are not clear-cut for *Follow-Local* and its potential benefits should be further studied, taking into consideration other non-verbal cues, such as eye contact and facial expressions. In particular, this method could be valuable for discussion and negotiation tasks. However, operationalizing such tasks in a controlled experiment is challenging, and the evaluation of collaboration should be extended beyond simple performance metrics. Given the advantages of both methods, future work should also explore how to seamlessly integrate them without hindering the collaboration process or overloading users.

Although the first prototype was implemented for two remote users in separate wall-sized displays, *CamRay* can accommodate more than one user per location, as the tracking system can individually identify multiple users. It can also scale to more than two locations, as the server can receive multiple WebRTC connections simultaneously. However, further developments would probably be necessary to support large groups at one location or numerous remote locations. In terms of collaboration, we need to explore further the collaborative behaviors that arise in such large groups, including coupling styles and territorial dynamics.

4.2.2 *Perception of a remote user's gaze direction*

While large interaction spaces can foster collaboration within a co-located group, integrating a remote user into such collaboration remains a challenge for current telepresence systems. In this work, we considered a simple scenario in which *co-located collaborators* sit around a table containing various physical artifacts, such as paper printouts or 3D mock-ups. To integrate a remote user in this scenario, most current telepresence systems use a screen and a camera situated at one edge of the table to support video-mediated communication. The wide perspective provided by this side camera makes it difficult for the remote user to see the physical artifacts on the table. In contrast, the co-located collaborators have a closer view of the remote user, with a much narrower perspective. The difference in perspective, along with the offset between the camera position and the video position at the remote location, does not allow co-located collaborators to properly interpret the gaze direction of the remote user. This hinders communication as co-located collaborators can easily understand each other's gaze direction, but struggle to do so for the remote user, potentially excluding this user from the collaboration. Our goal was to create a telepresence system that accurately convey the remote user's gaze direction to the group. The co-located collaborators should be able to understand if the remote user is gazing at one of them or at specific physical artifacts on the table.

Gaze is crucial for the collaboration, as it helps predict conversational attention [Ver99; Ver+01], perceive references to physical objects [ATI18], support remote instructions [HYS16; Yao+18] and enhance users' confidence [Akk+16]. Failing to properly convey gaze can lead to confused communication [VVV00], reduced effectiveness [MG02] and extra efforts to accomplish collaborative tasks [Akk+16; HYS16]. Previous work has explored gaze perception in remote collaboration, but has mainly focused on conveying either gaze awareness between distant users [SBA92; NC05; Gig+14] or gaze on shared digital content [IK92; KK06], leaving the problem of gaze awareness towards physical artifacts under studied. In addition, such systems often require specialized and complex hardware setups on the remote user's side [PS14; Ots16; Got+18], which might be unrealistic for traveling users.

We created *GazeLens* [**Le+19**], a video conferencing system designed to improve co-located collaborators' ability to interpret the remote user's gaze (Figure 4.14). At the group location, a 360° camera is located at the center of the table, and captures a panoramic video of the collaborators seated around it. A ceiling-mounted camera

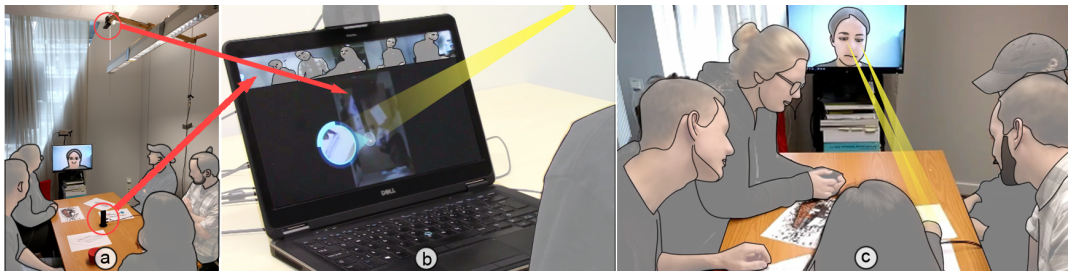

Figure 4.14: *GazeLens* system: (a) a 360° camera and a ceiling-mounted camera respectively capture the co-located collaborators and the physical artifacts; (b) the video from the two cameras are displayed on the remote user's screen, with a virtual lens guiding attention towards a specific screen area; (c) the remote user's gaze is properly aligned towards the observed artifact at the group location.

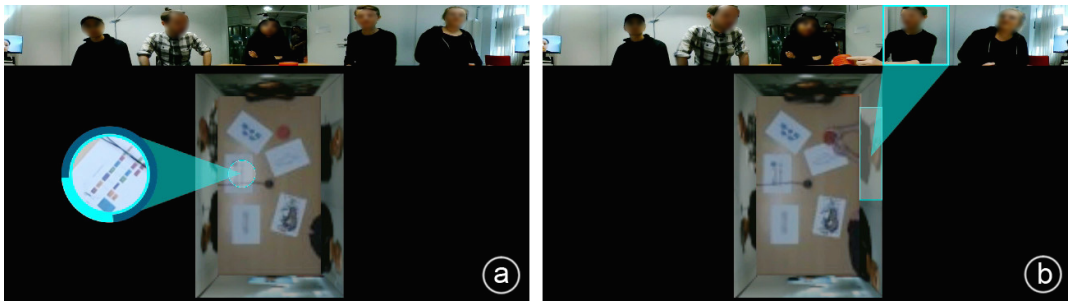

Figure 4.15: *GazeLens* interface: (a) a magnifying lens shows a close-up view of a physical artifact, and (b) a lens indicates a collaborator's position around the table.

captures the physical artifacts on the table, minimizing occlusions. At the remote location, the user has a standard computer equipped with a webcam on top of the screen. *GazeLens* combines the video feeds of the two cameras in a unified interface (Figure 4.15). We designed this interface to strategically direct the remote user's attention toward specific screen areas, allowing co-located users to more accurately interpret the remote user's gaze direction. The 360° video is displayed at the top of the screen, just under the webcam, reproducing eye contact for co-located users when the remote user looks at them in the video feed, as suggested by Chen [Che02]. The top-view video is displayed in the middle of the screen with the correct orientation and aspect ratio relative to the actual table. We used a focus-based approach that mimics foveal and peripheral vision to maximize variation of the remote user's gaze. The top-view video of the table appears slightly blurred, and the remote user can use a magnifying lens to obtain a sharper and closer view of the physical artifacts. This lens is positioned at a specific location on the screen, guiding the remote user's gaze in the right direction with respect to the webcam position (Figure 4.15-a). As physical artifacts can have various orientations on the table, we provide a rotation tool on the lens to rotate its content if needed. Finally, to keep the remote user aware of the co-located collaborators' position around the table, a second view of the 360° video is wrapped around the table top-view video in the interface. This additional view is slightly blurred, and a square lens connects it to the 360° video displayed at the top of the screen, helping the user in correlating these two views (Figure 4.15-b).

We conducted a first controlled experiment to evaluate the effectiveness of *GazeLens* in conveying the remote user's gaze in comparison to a conventional videoconferencing system (Figure 4.16). This baseline used a wide-angle camera to capture the entire room at the group location and displayed the corresponding video in full-screen mode on the remote computer, instead of the *GazeLens* interface. To minimize experiment bias, three actors assumed the role of the remote user. We recorded multiple videos of these actors looking at 14 targets under the two video conditions. 9 targets were arranged in a 3×3 grid on the table, while 5 targets were located at the co-located collaborators' position around the table (Figure 4.16-a). 12 participants took the role of a group member and looked at the pre-recorded videos while sitting at two different locations around the table: in front and on the side of the screen where the remote user's video is displayed (positions C and A in the figure). After viewing each video, they were asked to indicate which target the actor was looking at. We hypothesized that *GazeLens* would improve accuracy of gaze interpretation for both sitting positions compared with the baseline.

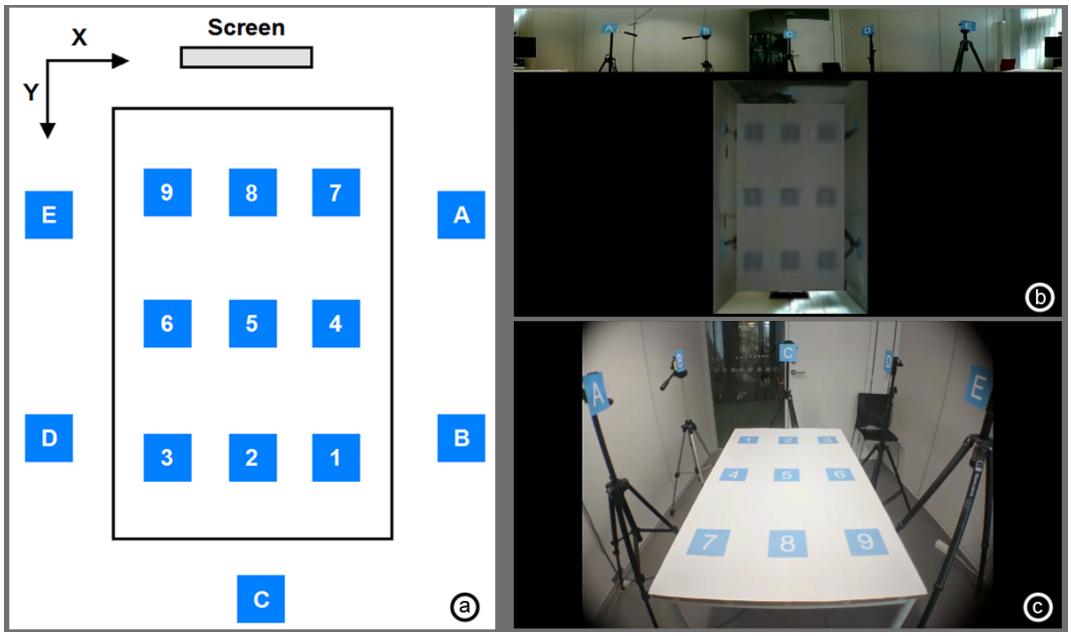

Figure 4.16: Experimental setup of the first study. (a) Multiple targets were laid out at the group location. These targets were viewed at the remote location through the (b) *GazeLens* interface or (c) a conventional videoconferencing system using a wide-angle camera.

Results show that *GazeLens* significantly increases the overall gaze interpretation accuracy of the participants compared to the conventional videoconferencing system. The position of the participants with respect to the screen does not influence these findings. In addition, participants can easily distinguish whether the remote user is looking at a co-located collaborator or at a physical artifact on the table, with over 85% accuracy. However, the results are less good when it comes to distinguishing artifacts on the table, although they were still better than with the conventional videoconferencing system. As a consequence, we decided to conduct a second experiment focusing on the physical artifacts.

A second controlled experiment used the same experimental setup. The only distinction was that the targets were arranged on the table exclusively, and with two densities: 9 targets in a 3×3 grid or 25 targets in a 5×5 grid. 12 participants took part in this experiment. We hypothesized that *GazeLens* would improve accuracy of gaze interpretation for both target densities compared with the baseline.

Results show that *GazeLens* significantly improves gaze interpretation accuracy for table artifacts for sparse, but also dense arrangements, compared with the conventional videoconferencing system. When using *GazeLens*, the accuracy reaches approximately 54% with the sparse layout versus 26% with the baseline. With the dense layout, it drops to around 26% for *GazeLens* and 12% for the baseline. We also analyzed lateral and depth errors. Although *GazeLens* outperforms the conventional system for both types of errors, it results in more depth errors compared to lateral errors. This could be explained by the fact that vertical screen space is limited in our interface, but also by the fact that vertical gaze direction is harder to interpret compared to the horizontal one, as studied by Chen [Cheo2].

As the two previous experiments focused on the co-located collaborators' perception, we also conducted a preliminary study to gather feedback from the remote

user's perspective. Five pairs of participants performed a puzzle-solving task under the two video conditions previously described. In this task, the remote user instructed a group member on how to arrange physical puzzle pieces on the table to match a predefined pattern. Participant swapped roles, taking turns as the remote user and the group member. We gathered qualitative feedback through interviews at the end of the experiment. Only one of the ten participants reported difficulties when using the *GazeLens* interface. He would have preferred another solution to activate the lens than a mouse click. Apart from that, all participants mentioned that it was easier to see the puzzle pieces with the *GazeLens* interface and preferred this condition over the conventional videoconferencing system.

Overall, *GazeLens* provides a telepresence interface that guides a remote user's gaze, enhancing gaze interpretation in video-mediated communication with multiple collaborators located in the same meeting room. Controlled experiments demonstrate that *GazeLens* improves the ability of these co-located collaborators to distinguish whether the remote user is looking at them or at physical artifacts on the meeting room table. *GazeLens* also enhances their accuracy in determining which specific physical artifacts on the table are referenced by the remote user, although there is still room for improvement. In particular, we designed *GazeLens* to be simple enough to be deployed in any meeting room with any camera, but we did not consider camera position, camera focal length, screen size or distance between the screen and the table when defining the lens position. Although achieving geometrically corrected gaze in video-mediated communication is almost impossible, our system could be improved by integrating these parameters. However, this could require some configuration and calibration steps, which can be time-consuming. It would be interesting to explore various trade-offs and find out which parameters have the biggest impact on the accuracy of gaze direction interpretation. Finally, future work should also explore how *GazeLens* can be extended to support multiple remote users. Each remote user can be represented by a dedicated screen around the table, but it could be valuable to investigate solutions using a single large screen, as most meeting rooms are usually equipped with only one screen dedicated to video-meditated communication.

4.2.3 *Exploration of a remote augmented reality workspace*

Augmented reality (AR) makes it possible to create large interactive spaces by integrating virtual content in any physical space. Nevertheless, sharing this virtual content with a remote collaborator is a challenge, especially when this user does not have access to AR or VR equipment. Yet, such asymmetrical collaboration configurations are common today in many circumstances, as more and more collaborators travel or work from home. While video-mediated communication plays a crucial role in enhancing remote collaboration in such contexts, it cannot provide a comprehensive understanding of the AR workspace including both physical and virtual content. Our objective was to establish an effective collaboration between augmented reality and remote desktop users, leveraging the benefits of AR for the remote user.

Several AR technologies propose to video-stream the AR user's perspective. However, these solutions do not provide a view of the user, thus failing to convey non-verbal cues such as gestures, body postures, or facial expressions. Moreover,

Figure 4.17: Remote-view configurations for the first study: (left) HEADSET VIEW, (middle) EXTERNAL VIEW and (right) VIRTUAL VIEW. Remote participants give instructions to the AR user on how to arrange 3D shapes on a virtual support.

the remote user's viewpoint is limited to the AR user's perspective, which hinders the remote user's ability to adequately perceive and explore the AR workspace. Nevertheless, ensuring view independence enhances collaboration performance, as reported by Tait and Billinghurst [TB15]. Some systems create a 3D reconstruction of the AR workspace, allowing the remote user to navigate independently in the 3D scene. However, this 3D reconstruction requites heavyweight and complex hardware setups [AAT13; Bai+20], consumes large communication bandwidth, is affected by network outages [Ahs+21] or imposes significant constraints on the view possibilities of the remote user [Gau+14; Moh+20]. In addition, the AR user is often reconstructed in 3D along with the environment, resulting in a poor user representation that reduces expressiveness and creates an "*uncanny valley of XR [extended reality] telepresence*" [Jon+21].

In this work, we targeted an asymmetric collaboration scenario between a local *AR user* wearing an optical see-through headset (Microsoft Hololens 2) and a *remote user* who participates from a distance through a desktop application. We restricted our design space to lightweight setups using only a single external depth camera on the AR user's side, in addition to the camera of the headset. This external camera could be easily replaced by a webcam and a smartphone as many devices are now equipped with depth sensors. The remote user simply uses a standard laptop or desktop computer with a webcam. We aimed to enhance video-mediated communication between these two users with new visual and interaction modalities.

As a first step, we conducted a user study [**FFT22b**] to investigate the trade-offs associated with different AR workspace representations and scene viewpoints. 24 participants took the role of the remote user and instructed an experimenter, who acted as a confederate, to accomplish a puzzle-solving task in AR. Participants were presented with a randomly generated pattern composed of 8 puzzle pieces among the 18 available in the AR workspace. The experimenter's task was to replicate this pattern on a virtual plane placed on his table. Participants provided instructions under three conditions (Figure 4.17):

- HEADSET VIEW: participants viewed an augmented video from a first-person viewpoint provided by the AR headset camera.
- EXTERNAL VIEW: participants viewed an augmented video from a third-person viewpoint provided by the external camera.
- VIRTUAL VIEW: participants viewed a fully virtual representation of the 3D scene. They could freely navigate in the scene and choose their own viewpoint. No information regarding the physical environment was visible, but a simplified avatar represented the AR user.

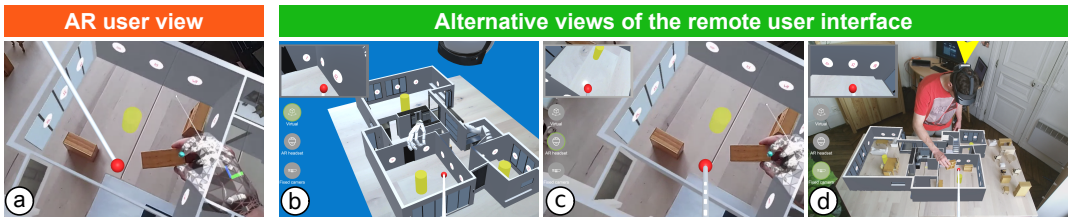

Figure 4.18: A remote user guides an AR user achieving a physical furniture arrangement task in a virtual 3D house model with *ARgus*. (a) The AR user view displayed in the headset and the three AR workspace representations combined in the desktop interface of the remote user: (b) a fully virtual view, (c) an augmented first-person view, and (d) an augmented third-person view.

After completing the task, participants rated the perceived difficulty for different components of the task and their overall preferences. The results indicate that each view configuration has its own qualities that are difficult to substitute using the other views. The External View provides a global perception of the AR workspace and helps participants search for puzzle pieces. The Virtual View supports independent navigation, helping participants give instructions from a convenient and stable viewpoint. Finally, the Headset View is effective for perceiving the AR user's actions and communicating egocentric instructions.

Building on these findings, we focused on combining these multiple representations and providing remote users with direct control over their use. We designed *ARgus* [**FFT22b**; **FFT22a**], a multi-view video-mediated communication system that combines the three representations through interactive tools for navigation, previewing, pointing, and annotation (Figure 4.18). *ARgus* receives the augmented video from both the AR headset and the external depth camera. It also maintains a synchronized version of the virtual scene, and can generate virtual views from any location. This enables the remote user to seamlessly switch between the Headset View, the External View and any viewpoint of the Virtual View. Additionally, *ARgus* offers the ability to display live previews of each view in a thumbnail at the top of the current view, allowing users to quickly glance at a view or decide whether it is worth switching to another view.

The *ARgus* interface provides three buttons for transitioning between views. Hovering the mouse over a button displays the preview thumbnail of the corresponding view. Clicking on the button activates the view. We used a trajectory and field-of-view interpolation of the camera in the virtual scene when switching views to avoid abrupt transition and disorientation. 3D navigation in the virtual scene is possible by using the mouse. The virtual scene also offers an alternative to preview and switch between views by hovering over and clicking on dedicated 3D widgets: the AR user head for the Headset View and the 3D model of the external camera for the External View. When viewing one of the two augmented video views, the remote user can still use the mouse to navigate in 3D, but this immediately switches the representation to the Virtual View. We also provided a pointing stick and annotation features to enhance communication. The pointing stick can be activated in any view, but it temporarily freezes the Headset View to allow for accurate pointing. Annotations are represented by colored spheres visible in any view.

We conducted a second user study that observed how 12 participants used *ARgus* to provide remote instructions for an AR furniture arrangement task. We

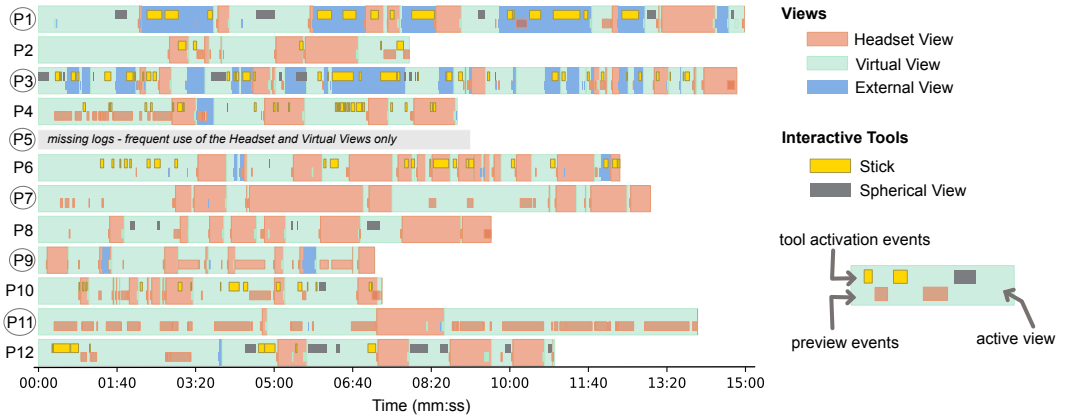

Figure 4.19: Timelines showing the use of the three views, the previews and the pointing stick for each participant when performing the experimental task with *ARgus*. Circled participants were exposed to *ARgus* first.

compared *ARgus* to a controlled condition using only the Headset View without any interaction functionalities. We designed a furniture arrangement task, which involves both physical and virtual objects in the 3D scene. Participants had to give instructions to an experimenter, who acted as a confederate, for positioning physical miniature furniture in a virtual 3D model of a house. Participants were given a set of constraints to fulfill for the furniture arrangement. These constraints rely on random parameters to create various arrangement tasks unknown to the experimenter. Participants performed a distinct task for each video condition, and then answered a questionnaire to provide feedback about the two conditions.

To analyze the strategies used by participants to complete the task with *ARgus*, we created timelines showing the time spent in each view, as well as the use of previews and communication tools (Figure 4.19). Since we did not encourage participants to be fast, the time range does not always reflect active collaboration, as some participants spent initial time thinking about the task or exploring the 3D house model. Overall, participants frequently transitioned between views or used previews, which demonstrates that *ARgus* is especially useful to perform the task. This is confirmed by the fact that *ARgus* reduced participants' reliance on verbal instructions and was generally preferred compared to the condition using only the Headset View. Nevertheless, we observed that participants employed different strategies when using *ARgus*. While 3 participants (P1, P3, P9) found the External View very useful, others judged that the Virtual View and Headset View were enough to complete the task. A few participants, such as P11, relied extensively on the preview feature, whereas others used it temporarily, mainly before switching views. We hypothesized that mastering all combinations of views and previews, as well as developing strategies to use them effectively in various collaboration steps, may require a long learning process that was not assessed in this study.

In summary, we explored how different views can enable a remote desktop user to collaborate with an AR user by perceiving both the physical and virtual content surrounding this AR user. We first compared three representations of the AR workspace and showed that each of them presents different benefits, targeting different collaboration aspects. Based on these findings, we developed *ARgus*, a multi-view collaboration system that provides tools for effectively switching between views and navigating in the AR workspace. A second user study suggests

that the flexibility of *ARgus* allows remote users to verify spatial constraints more efficiently and reduces their reliance on verbal instructions. Future work needs to examine the multi-view collaboration strategies from the perspective of the AR user, and how to provide awareness about the visual perception and interaction capabilities of the remote user. Moreover, *ARgus* could be extended to multiple remote users, but this will pose significant challenges in terms of awareness since users will now have distinct representations of the 3D scene and various viewpoints.

4.3 CONCLUSION

Remote collaboration across large interactive spaces is becoming crucial in many situations, allowing remote experts to combine their expertise and offering the flexibility to work from home or reduce travel. As a consequence, new collaborative systems need to handle a wide range of collaboration scenarios and support users with asymmetric device configurations. In this chapter, I first studied the technical aspects of connecting remote users across heterogeneous interactive platforms. I then explored various telepresence systems for enhancing awareness among such remote users with video-mediated communication.

In the first section, I presented several systems for synchronizing CAD data across remote locations, transmitting spatialized 3D audio and reconstructing live 3D head models of remote users. However, these systems are still preliminary proofs of concept at this stage, and would benefit from further development and evaluation on real-life collaboration scenarios. For example, our architecture for synchronizing CAD data could be tested in a large-scale collaboration setting involving multiple design team members interacting with a wide range of immersive and non-immersive devices. The spatialized audio system could be integrated and tested with the telepresence systems proposed in the second section of this chapter.

In the second section, I investigated the potential of video-mediated communication to enhance remote collaboration in different configurations. I first focused on a one-to-one collaboration between two remote collaborators using similar interactive platforms. Next, I explored what happens when a user is away from the work team, involving a one-to-many collaboration. Finally, I addressed the situation in which users do not have access to the same equipment, leading to a one-to-one collaboration between users with immersive and non-immersive devices. For each collaboration configuration, our work is grounded in experimental findings that provide fundamental insights on how users can collaborate through video. In particular, we assessed users' ability to interpret deictic gestures through video and the impact of different representations of augmented reality content on collaboration. Based on these findings, we designed several telepresence systems following the requirements of each collaboration scenarios. I want to emphasize that these systems also represent technical achievements in themselves, involving complex features such as streaming multiple video feeds, augmenting video with virtual content, or transmitting video along with users' positions and actions. As the final step, we evaluated these systems on various collaborative tasks. In multiple cases, we observed that effectively conveying appropriate non-verbal cues or providing useful communication tools can enhance collaboration and reduce the reliance on verbal communication. Nevertheless, conducting an exhaustive evaluation of such collaborative systems is challenging due to the multitude of situations, distinct user

roles, and diverse phases in the collaboration process. Therefore, further evaluation of our systems in different tasks and contexts would be welcome.

While most of the work presented in this chapter can be extended to multiple users and multiple remote locations, they have only been tested with two users in one-to-one collaboration. This is mainly due to the complexity of conducting controlled experiments with more than two users, as a larger number of users considerably increases the potential biases of the experiments. However, we need to find new ways to assess collaboration in such multi-user scenarios. Relying on observational studies, as we did to evaluate *ARgus*, may be a solution. It could also be useful to compute real-time indicators of collaboration quality by automatically analyzing users' speech, gaze direction and relative movements, instead of doing it manually after the experiments, as we did in most of our work. In future work, I want to assess how the proposed systems can handle multiple users in each large interactive space. A first step will be to assess *CamRay* with two users in front of each wall-sized display. I also plan to extend these systems to multiple remote locations. For example, *GazeLens* and *ARgus* could be easily extended to integrate multiple remote users. A long-term objective will be to target true hybrid collaboration situations involving both co-located and remote users interacting with heterogeneous devices.

In its current state, this research treats collaboration as a single, simple activity between users. However, collaboration is much more complex, involving different collaboration styles that evolve over time. These styles can include tightly coupled and loose collaboration, subgroup collaboration, as well as spontaneous or side discussions. In future work, I want to assess how the proposed systems can support these different collaboration styles. I also want to extend these systems to better allow transitions among the different phases of a collaboration. For example, *CamRay* can probably handle different collaboration phases with its two video modes, but a solution to transition between the two modes would be required to support various collaboration dynamics.

5 FUTURE PERSPECTIVES AND CLOSING REMARKS

The previous chapters have presented my past work, which focused on investigating individual and collaborative interaction in shared interactive spaces (Chapter 3) and connecting remote users across large interactive spaces through appropriate communication and awareness cues (Chapter 4). This final chapter describes my future research directions and concludes this manuscript with more general remarks.

5.1 INTERACTION ALL ALONG THE MIXED REALITY CONTINUUM

Chapter 3 presented interaction and collaboration techniques targeting different levels of the mixed reality continuum, defined by Milgram et al. [Mil+95] and recently revisited by Skarbez et al. [SSW21]. These techniques include touch interaction on a 2D wall-sized display, 3D gestures in an augmented reality space and haptic interaction in an immersive virtual reality system. However, these interaction techniques remain designed for a specific device and cannot be applied at other levels of the mixed reality continuum. In this section, the term "level" designates a specific position of the continuum, without any notion of discrete tiers or hierarchy.

In general, with the wide diversification of computing devices, solutions exist for visualizing content and interacting at each level of the mixed reality continuum. Additionally, new devices start to offer the ability to transition along this continuum. For example, video see-through headsets now enable users to switch from real vision to various AR views and fully immersive VR views. Despite this, applications and interaction techniques stay siloed at specific levels of the continuum.

As each level has its own benefits and they all complement each other, I argue that new interactive systems should provide users with the ability to interact at multiple levels of the continuum and transition among them. In particular, I believe that such transitions are mandatory to integrate mixed reality into the everyday work pipeline. To be usable, such interactive systems must provide users with consistent interaction techniques to avoid confusing them by changing techniques every time they change level. My future work will concentrate on two main challenges for this research axis: (i) designing large interactive spaces that support transitions along the mixed reality continuum, and (ii) providing consistent interaction techniques that enable users to seamlessly interact across multiple levels.

Supporting transitions along the mixed reality continuum. To achieve these transitions, I envision that users could use either multiple devices or a single device allowing such transitions. I illustrate this concept by presenting a realistic scenario that highlights the benefits of each level of the continuum. This scenario was inspired by observations of real designers in the automotive industry who use a desktop computer along with a VR headset: they use CAD software on the desktop computer to design products and review them in 3D using the VR headset.

This scenario involves engineers who need to analyze a large number of 3D numerical simulation results (Figure 5.1). For tasks such as parameterizing the sim-

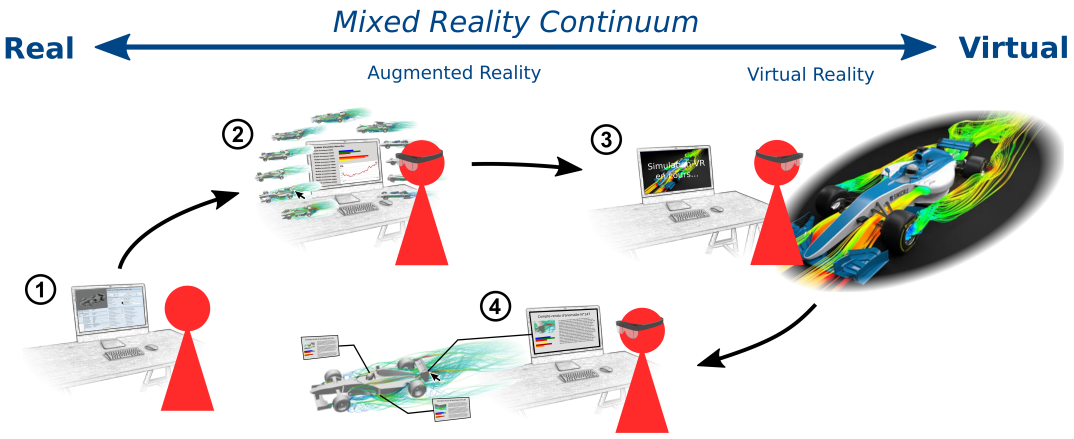

Figure 5.1: Scenario involving interaction along the mixed reality continuum: (1) interaction with the 2D desktop interface, (2) augmentation of the computer screen with 3D models, (3) immersion in a 3D model with virtual reality and (4) interaction with a hybrid interface combining augmented reality and desktop interface.

ulation software or sorting the results, a standard computer allows them to perform these actions efficiently by using the mouse and keyboard (Step 1 in the figure). To compare several simulation results, they use a mixed reality headset that displays these results in 3D around their computer screen (Step 2). In this configuration, they can interact with the data using the mouse both on the screen and in the 3D space, as proposed by Plasson et al. [PBN22]. Later, to better understand an unexpected detail in one of the results, they choose to immerse themselves in a 3D virtual environment, enabling them to view the simulation in context (Step 3). They can thus physically move around and interact in 3D with the displayed data. Finally, once they identify the problem, they return to a hybrid view that combines the computer screen with a 3D view of this specific simulation result. This configuration makes it easy to annotate the data with the keyboard (Step 4).

This scenario illustrates a global vision involving both multi-device interactions and transitions between different levels of the mixed reality continuum. Roo and Hachet [RH17] proposed *One Reality*, a conceptual framework enabling a comparable scenario. *One Reality* allows users to interact with a physical object and its virtual counterpart at 6 levels of the continuum, ranging from the physical object alone to an immersive view of the virtual counterpart in VR. Although all levels are synchronized, users still need to switch devices to transition between certain levels. Such systems require a distributed software architecture to share data across multiple devices. This architecture could be inspired by the work achieved during my PhD to distribute data across VR devices in a collaborative context [**Fle+10b**].

As a first step, I plan to explore simpler configurations that combine desktop interfaces, mobile devices or wall-sized displays with mixed reality technologies. For example, in the context of co-located collaboration, James et al. [JBC23] propose to extend a wall-sized display with shared and personal surfaces displayed using AR headsets. I have also initiated a project on 3D editing, in which we want to augment a standard computer screen by incorporating 3D views positioned around the screen with augmented reality. The goal of this future work will be to gain insights on how users can interact at various levels of the continuum and assess the need for transitions among these levels.

Providing consistent interaction across the mixed reality continuum. To seamlessly interact across different levels of the mixed reality continuum, users need consistent techniques that prevent them from having to switch to a whole new set of interaction techniques every time they change level. A first solution is to enable users to interact at multiple levels using the same input device and interaction technique. For example, Plasson et al. [PBN22] propose to use a mouse for interacting on a 2D screen, as well as with 3D views displayed next to the screen through an AR headset. James et al. [JBC23] extend the pointing and grabbing techniques from an AR headset to also grab 2D content on a wall-sized display. I plan to further explore multi-level interaction techniques for manipulating both 2D and 3D content with various devices and different interactive setups.

However, we cannot expect users to interact with the same technique across all levels of the continuum, given the wide range of devices available. We must therefore enable them to change techniques without being lost or having to relearn all the interaction mechanisms for each new application. I think that efforts should be made to propose standardized interaction techniques, especially when it comes to 3D interaction, which is fairly new to users and remains very specific from one application to another. Interaction discoverability should also be improved in mixed reality systems, as it is the case for 2D interfaces. Finally, we have to keep in mind that the new techniques must allow multiple users to interact in the same interactive space and potentially support collaborative activities.

5.2 HYBRID COLLABORATION ACROSS LARGE INTERACTIVE SPACES

Chapter 3 explored various co-located collaboration scenarios, while Chapter 4 focused on remote collaboration. However, none of this work investigates real hybrid collaborative situations, including both co-located and remote users. While *GazeLens* (Section 4.2.2) aimed to integrate a remote user into a co-located collaboration, this work did not explore the co-located aspect of the collaboration.

Hybrid collaboration has become a necessity due to major changes in our society and the new organization of work. We are experiencing more and more diverse collaborative situations, such as meetings with a colleague working from home, or work sessions between two distant groups. The COVID-19 pandemic significantly accentuated this trend [Yan+22]. However, current computer-mediated collaboration systems often lack flexibility to adequately support hybrid collaboration. This can lead to awkward situations in which colleagues within the same building opt to stay in their individual office for a videoconference meeting instead of attending together, or are forced to have side conversations via chat during such meetings.

My future research aims to investigate how large interactive spaces can foster real-time collaboration in hybrid situations. For example, these situations may include remote collaboration among co-located subgroups or collaboration between a co-located group and multiple remote users (Figure 5.2). As suggested in the previous research axis, users may interact at different levels of the mixed reality continuum using heterogeneous devices, ranging from simple smartphones to immersive VR rooms. I think that this diversity of devices has the potential to enhance hybrid collaboration. Nevertheless, this raises new challenges regarding (i) how to integrate users with heterogeneous devices in the collaboration, (ii) how to provide appropriate awareness among users, regardless of whether they

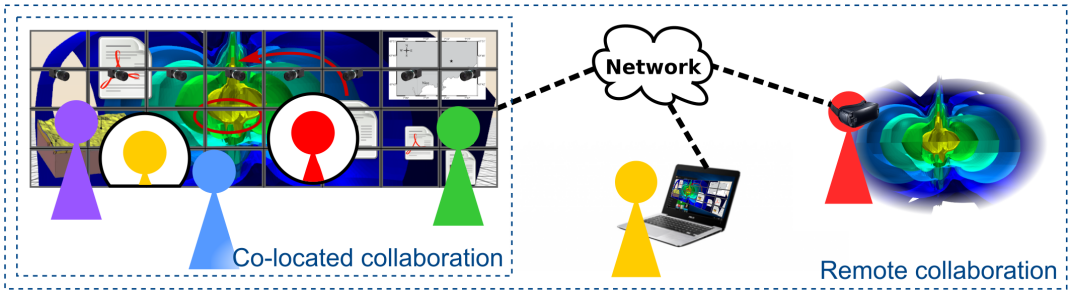

Figure 5.2: Hybrid collaboration involving co-located and remote users interacting with heterogeneous devices.

are co-located or remote, and (iii) how to support different dynamics during the collaboration process. Without trying to mimic collaboration in the real world, I believe we need new ways of collaborating that take advantage of the specific capabilities of each interaction device, in a similar spirit as the "Beyond being there" concept described by Hollan & Stornetta [HS92].

Integrating users with heterogeneous devices. As presented in the previous research axis, I envision that users will be able to interact at multiple levels of the mixed reality continuum and transition among these levels. Consequently, users will collaborate with both co-located and remote collaborators across various levels of the continuum. These heterogeneous situations provide many opportunities for exploring new collaboration scenarios. For instance, given that using immersive VR for extended periods of time can be strenuous, users could use non-immersive devices to remotely monitor collaborators immersed in VR and interact with them. This approach allows users to reserve VR for specific tasks during long work sessions, taking turns in VR as needed. I also plan to study another scenario in which a co-located group collaborates in front of a wall-sized display, while remote collaborators, who do not have access to such equipment, join them using VR headsets that display a virtual version of the content. A few studies have explored asymmetric collaboration across heterogeneous devices, but they mainly focused on co-located collaboration. For example, *ShareVR* [Gug+17] and *TransceiVR* [Tho+20] use a smartphone or a tablet to interact with a user immersed in VR, while *ShARe* [Jan+20] and *HMD Light* [Wan+20b] use a projector mounted on an AR or VR headset to share the view of the headset user. I plan to extend this previous work to broader hybrid collaborative situations including remote users.

The main challenge is to provide all users with appropriate interaction techniques to act on shared content and communicate their ideas, regardless of their geographical location or their device. These techniques should leverage the potential of every device to allow users to have complementary interaction capabilities. Additionally, it is necessary to find appropriate ways to represent users' activities and interaction capabilities to improve understanding among them. Given that users may have varying interaction capabilities, it is crucial that they can understand what others are currently doing and what they can do to enable effective collaboration. As a first step, I plan to extend *ARgus* (Section 4.2.3) to support multiple co-located AR users, as well as multiple remote users. This will require finding solutions to allow AR users to accurately understand what is the viewpoint of each remote user on the AR workspace, and give all remote users the ability to participate in 3D interaction.

Providing appropriate awareness between co-located and remote users. Allowing effective collaboration in hybrid situations requires to provide appropriate awareness among all users. This awareness is crucial for enhancing their mutual understanding and helping them build a common ground [CB91]. Providing such awareness is challenging between co-located and remote users, as they do not share the same interactive space and cannot see each other directly. I plan to explore multiple solutions for representing the remote users, but also the space surrounding them, in a way that ensures seamless interaction between co-located and remote users. Representing the space surrounding users is especially important to facilitate the establishment of a common ground, as we studied in *ARgus* (Section 4.2.3).

I first want to explore different visual representations of the remote users in collaborative situations involving mixed reality technologies. There is not clear consensus regarding the impact of user representations on collaboration. Yoon et al. [Yoo+19] compared the effects of realistic and cartoon-like avatars on social presence, while Congdon et al. [Con+23] compared the effects of video and 3D avatar representations on trust. Both studies concluded that the results could highly depend on the collaborative context and the environment surrounding the users. Although solutions exist to create high-quality realistic avatars, such as *Pixel Codec Avatars* [Ma+21], I believe that there are some collaborative situations where avatars may not be the most appropriate representation. It is especially the case for collaborative situations including users with both immersive and non-immersive devices, as using avatars for non-immersive users may not be meaningful. In such situations, I want to experiment with solutions that integrate real video streams into virtual environments in a way that goes beyond real-world collaboration. For example, we could attach virtual windows displaying remote collaborators' video on the side of the users' field of view in a VR environment or on the users' wrist in an AR environment. This will create a virtual face-to-face with the remote collaborators, as we explored with *CamRay* on wall-sized displays (Section 4.2.1). We also need to find solutions to represent users of immersive devices for the collaborators using non-immersive devices.

Although showing the spaces surrounding remote users is straightforward in non-immersive contexts with cameras, it becomes challenging for mixed reality environments that overlap multiple remote spaces with both physical and virtual content. Most previous work on remote collaboration in mixed reality focused on host-guest situations, where the guest is immersed in the augmented environment of the host [Teo+19; Piu+19; Bai+20]. Other research proposed sharing only a few physical objects [Ort+16] or virtual content [Mah+19], but not the entire spaces surrounding users. However, some collaborative situations require a shared space that combines the spaces of all remote users with their corresponding physical constraints. A few studies have explored this aspect, mainly focusing on the technical aspects of reconstructing and blending users' physical spaces [LMR14]. I plan to approach the problem from a different angle by studying how users perceive the remote spaces of their collaborators and identifying which cues are mandatory to build a mental representation of the shared space. I will then investigate representations that mix symbolic and realistic elements to reveal this shared space. These representations should prevent users from perceiving the shared space as a superposition of individual spaces. Instead, they should facilitate the establishment of a common ground between users, enhancing their mutual understanding.

Supporting various collaboration dynamics. As the number of users involved in hybrid collaboration increases, not all users will be collaborating together at all times. Collaborative systems will thus have to support different moments in collaboration, such as tightly coupled and loose collaboration, subgroup collaboration, and spontaneous or side discussions. However, current telepresence systems, including those presented in Section 4.2, do not adequately support these dynamic collaboration scenarios.

As an initial step, I plan to study collaboration dynamics in co-located situations without technology mediation. In particular, I want to observe groups of co-located users interacting in large interactive spaces, such as rooms equipped with large screens, tabletops, or AR devices. Building upon these observations, the goal is to extend our work on telepresence systems for wall-sized displays (Section 4.2.1) to support these collaboration dynamics. An obvious first goal is to explore manual and automatic solutions for switching between the *Follow-Local* and *Follow-Remote* modes of *CamRay*. Moreover, I can imagine various other collaborative feature based on video. For example, some video windows could appear or disappear as appropriate to manage tightly coupled versus loose collaboration. Other video windows could be split and displayed at different positions to encourage users to move to specific areas of the screens, thus fostering subgroup collaboration across remote platforms. Additional devices, such as smartphones, could also be used on the fly to make side conversations possible. In addition to video, I believe it would be valuable to incorporate spatialized 3D audio by using the system we developed (Section 4.1.2). Spatialized 3D audio could enable users to determine remote collaborator positions, manage subgroup discussions without disturbing others or ensure privacy for side conversations.

5.3 CLOSING REMARKS

The demand for computer-supported cooperative work has never been more critical, given the substantial growth of digital data and significant societal changes, including new work organization and the green transition. An increasing number of individuals are required to work from home or collaborate with colleagues worldwide while limiting their travel to mitigate their environmental footprint. Although computer-supported cooperative work has been studied for several decades, the vast majority of previous work considered simplistic collaborative situations, such as one-to-one or group collaboration among individuals with similar roles. However, real-world collaboration is considerably more complex, involving multiple roles, users entering and leaving the collaboration, and different collaboration dynamics, such as spontaneous or side discussions. I argue in this manuscript that large interactive spaces provide a unique opportunity to support such complex collaborative situations across time and space. Nevertheless, further research is still needed to handle hybrid collaboration and transitions along the mixed reality continuum.

While my contributions and future perspectives mainly concentrate on synchronous collaboration, large interactive spaces hold significant potential for fostering collaboration in asynchronous situations. This is a typical case where computer systems can provide users with collaborative interaction that goes far beyond what is possible without technology mediation. Fender and Holz [FH22] illustrate the benefits of mixed reality technology for co-located asynchronous collaboration.

However, only a few studies have addressed asynchronous collaboration, as observed by Irlitti et al. [Irl+16], leaving plenty of space for design exploration, as suggested by Chow et al. [Cho+19]. I believe that asynchronous collaboration can be a promising long-term perspective for future work.

Targeting complex collaborative situations also raises the question of how to evaluate collaboration, as it cannot be solely assessed through performance measures in lab experiments. The success of collaboration is determined by many underlying indicators that are challenging to quantify. These indicators include social presence, mutual understanding, active participation, and feeling of closeness or friendliness among collaborators. A few studies have attempted to measure some of these indicators through questionnaires or post-experiment conversational analysis, as we have done in some of our work [**Ave+17**; **Oku+20**; **FFT22b**]. However, these analyses are difficult and time-consuming, thus limiting the number of indicators that can be measured. With current advances in sensing technologies and artificial intelligence, we just started to create a system that evaluates collaboration quality in real time [**Léc+23**], as part of the PhD work of A. Léchappé, co-supervised with M. Cholet and C. Dumas, and in collaboration with A. Milliat. At the current stage, this system can collect gaze and speech signals, as well as compute speaking time distribution, turn-taking, speech overlaps, joint visual attention, and mutual gaze. A first iteration used these indicators to differentiate situations with active collaboration from those without collaboration. Future steps will consist of detecting more complex collaborative situations, and providing users with real-time feedback to prevent critical situations arising from poor collaboration.

To conclude, I believe that large interactive spaces hold huge potential for fostering collaboration in various real-world situations. Nevertheless, many challenges persist in providing collaborators with rich social interaction and appropriate collaborative features. Close collaboration with researchers in social sciences will be crucial to better understand how individuals collaborate through technology and to adequately evaluate such collaboration.

BIBLIOGRAPHY

[AAT13] Matt Adcock, Stuart Anderson, and Bruce Thomas. "RemoteFusion: Real Time Depth Camera Fusion for Remote Collaboration on Physical Tasks." In: *Proceedings of the SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '13. Hong Kong, Hong Kong: ACM, 2013, pp. 235–242. DOI: 10.1145/2534329.2534331.

[ADL09] Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. "3-Hand Manipulation of Virtual Objects." In: *Proceedings of the Joint Virtual Reality Conference of EGVE - ICAT - EuroVR*. The Eurographics Association, 2009. DOI: 10.2312/EGVE/JVRC09/153-156.

[ADL10a] Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. "Comparison of Three Interactive Techniques for Collaborative Manipulation of Objects in Virtual Reality." In: *Computer Graphics International*. CGI' 10. 2010. URL: https://inria.hal.science/inria-00534087.

[ADL10b] Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. "Reconfigurable Tangible Devices for 3D Virtual Object Manipulation by Single or Multiple Users." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST '10. Hong Kong: ACM, 2010, pp. 227–230. DOI: 10.1145/1889863.1889913.

[Ahs+21] Tooba Ahsen, Zi Yi Lim, Aaron L. Gardony, Holly A. Taylor, Jan P de Ruiter, and Fahad Dogar. "The Effects of Network Outages on User Experience in Augmented Reality Based Remote Collaboration - An Empirical Study." In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW2 (2021). DOI: 10.1145/3476054.

[Akk+16] Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. "GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks." In: *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. CHI EA '16. San Jose, California, USA: ACM, 2016, pp. 1151–1158. DOI: 10.1145/2851581.2892459.

[ATI18] Deepak Akkil, Biju Thankachan, and Poika Isokoski. "I See What You See: Gaze Awareness in Mobile Video Collaboration." In: *Proceedings of the Symposium on Eye Tracking Research & Applications*. ETRA '18. Warsaw, Poland: ACM, 2018. DOI: 10.1145/3204493.3204542.

[AEN10] Christopher Andrews, Alex Endert, and Chris North. "Space to Think: Large High-Resolution Displays for Sensemaking." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: ACM, 2010, pp. 55–64. DOI: 10.1145/1753326.1753336.

[And+11] Christopher Andrews, Alex Endert, Beth Yost, and Chris North. "Information Visualization on Large, High-Resolution Displays: Issues, Challenges, and Opportunities." In: *Information Visualization* 10.4 (2011), pp. 341–355. DOI: 10.1177/1473871611415997.

[AGoo] Rajarathinam Arangarasan and Rajit Gadh. "Geometric Modeling and Collaborative Design in a Multi-Modal Multi-Sensory Virtual Environment." In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference Volume 1: 20th Computers and Information in Engineering Conference (2000), pp. 757–765. DOI: 10.1115/DETC2000/CIE-14592.

[Ara+13] Bruno R. De Araújo, Géry Casiez, Joaquim A. Jorge, and Martin Hachet. "Mockup Builder: 3D modeling on and above the surface." In: *Computers & Graphics* 37.3 (2013), pp. 165–178. DOI: 10.1016/j.cag.2012.12.005.

[AA13] Ferran Argelaguet and Carlos Andujar. "A survey of 3D object selection techniques for virtual environments." In: *Computers & Graphics* 37.3 (2013), pp. 121–136. DOI: 10.1016/j.cag.2012.12.003.

[AFB15] Ignacio Avellino, Cédric Fleury, and Michel Beaudouin-Lafon. "Accuracy of Deictic Gestures to Support Telepresence on Wall-Sized Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'15. Seoul, Republic of Korea: ACM, 2015, pp. 2393–2396. DOI: 10.1145/2702123.2702448.

[Ave+17] Ignacio Avellino, Cédric Fleury, Wendy E. Mackay, and Michel Beaudouin-Lafon. "CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'17. Denver, Colorado, USA: ACM, 2017, pp. 6718–6729. DOI: 10.1145/3025453.3025604.

[Bai+20] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. "A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: ACM, 2020, pp. 1–13. DOI: 10.1145/3313831.3376550.

[Bal+99] Ravin Balakrishnan, George Fitzmaurice, Gordon Kurtenbach, and William Buxton. "Digital Tape Drawing." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '99. Asheville, North Carolina, USA: ACM, 1999, pp. 161–169. DOI: 10.1145/320719.322598.

[BNB07] Robert Ball, Chris North, and Doug A. Bowman. "Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. San Jose, California, USA: ACM, 2007, pp. 191–200. DOI: 10.1145/1240624.1240656.

[Bar+21] Andrea Bartl, Stephan Wenninger, Erik Wolf, Mario Botsch, and Marc Erich Latoschik. "Affordable But Not Cheap: A Case Study of the Effects of Two 3D-Reconstruction Methods of Virtual Humans." In: *Frontiers in Virtual Reality* 2 (2021). DOI: 10.3389/frvir.2021.694617.

[Bea11] Michel Beaudouin-Lafon. "Lessons Learned from the WILD Room, a Multisurface Interactive Environment." In: *Proceedings of the Conference on l'Interaction Homme-Machine*. IHM '11. Sophia Antipolis, France: ACM, 2011. DOI: 10.1145/2044354.2044376.

[BMoo] Michel Beaudouin-Lafon and Wendy E. Mackay. "Reification, Polymorphism and Reuse: Three Principles for Designing Visual Interfaces." In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '00. Palermo, Italy: ACM, 2000, pp. 102–109. DOI: 10.1145/345513.345267.

[Bea+12] Michel Beaudouin-Lafon et al. "Multisurface Interaction in the WILD Room." In: *Computer* 45.4 (2012), pp. 48–56. DOI: 10.1109/MC.2012.110.

[Bec+13] Stephan Beck, André Kunert, Alexander Kulik, and Bernd Froehlich. "Immersive Group-to-Group Telepresence." In: *IEEE Transactions on Visualization and Computer Graphics* 19.4 (2013), pp. 616–625. DOI: 10.1109/TVCG.2013.33.

[Bee+10] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. "High-Quality Single-Shot Capture of Facial Geometry." In: *ACM ACM Transactions on Graphics (siggraph'10)* 29.3 (2010), 40:1–9. DOI: 10.1145/1778765.1778777.

[Bee+11] Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. "High-quality passive facial performance capture using anchor frames." In: *ACM Transactions on Graphics (siggraph'11)* 30.4 (2011), 75:1–10. DOI: 10.1145/2010324.1964970.

[Ben+95] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. "User Embodiment in Collaborative Virtual Environments." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '95. Denver, Colorado, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 242–249. DOI: 10.1145/223904.223935.

[Ber99] Julien Berta. "Integrating VR and CAD." In: *IEEE Computer Graphics and Applications* 19.5 (1999), pp. 14–19. DOI: 10.1109/38.788793.

[BB05a] Anastasia Bezerianos and Ravin Balakrishnan. "The Vacuum: Facilitating the Manipulation of Distant Objects." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: ACM, 2005, pp. 361–370. DOI: 10.1145/1054972.1055023.

[BB05b] Anastasia Bezerianos and Ravin Balakrishnan. "View and space management on large displays." In: *IEEE Computer Graphics and Applications* 25.4 (2005), pp. 34–43. DOI: 10.1109/MCG.2005.92.

[BB09] Xiaojun Bi and Ravin Balakrishnan. "Comparing Usage of a Large High-Resolution Display to Single or Dual Desktop Displays for Daily Work." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: ACM, 2009, pp. 1005–1014. DOI: 10.1145/1518701.1518855.

[BCL15] Mark Billinghurst, Adrian Clark, and Gun Lee. "A Survey of Augmented Reality." In: *Foundations and Trends® in Human–Computer Interaction* 8.2-3 (2015), pp. 73–272. DOI: 10.1561/1100000049.

[BK02] Mark Billinghurst and Hirokazu Kato. "Collaborative Augmented Reality." In: *Communication of the ACM* 45.7 (2002), pp. 64–70. DOI: 10.1145/514236.514265.

[Bil+02] Mark Billinghurst, Hirokazu Kato, Kiyoshi Kiyokawa, Daniel Belcher, and Ivan Poupyrev. "Experiments with Face-To-Face Collaborative AR Interfaces." In: *Virtual Reality* 6.3 (2002), pp. 107–121. DOI: 10.1007/s100550200012.

[BWF98] Mark Billinghurst, Suzanne Weghorst, and Thomas Furness. "Shared Space: An Augmented Reality Approach for Computer Supported Collaborative Work." In: *Virtual Reality* 3.1 (1998), pp. 25–36. DOI: 10.1007/BF01409795.

[Blo22] Nick Bloom. *Working from Home and the Future of U.S. Economic Growth under COVID*. https://youtu.be/jtdFIZx3hyk, accessed: 2022-03-23. 2022.

[BHI93] Sara A. Bly, Steve R. Harrison, and Susan Irwin. "Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment." In: *Commun. ACM* 36.1 (1993), pp. 28–46. DOI: 10.1145/151233.151235.

[Bou+10] Patrick Bourdot, Thomas Convard, Flavien Picon, Mehdi Ammi, Damien Touraine, and Jean-Marc Vézien. "VR–CAD integration: Multimodal immersive interaction and advanced haptic paradigms for implicit edition of CAD models." In: *Computer-Aided Design* 42.5 (2010). Advanced and Emerging Virtual and Augmented Reality Technologies in Product Design, pp. 445–461. DOI: 10.1016/j.cad.2008.10.014.

[BT02] Patrick Bourdot and Damien Touraine. "Polyvalent display framework to control virtual navigations by 6DOF tracking." In: *Proceedings of the IEEE Virtual Reality (VR)*. 2002, pp. 277–278. DOI: 10.1109/VR.2002.996537.

[Bux09] Bill Buxton. "Mediaspace – Meaningspace – Meetingspace." In: *Media Space 20 + Years of Mediated Life*. Ed. by Steve Harrison. London: Springer, 2009, pp. 217–231. DOI: 10.1007/978-1-84882-483-6_13.

[Bux+00] William Buxton, George Fitzmaurice, Ravin Balakrishnan, and Gordon Kurtenbach. "Large displays in automotive design." In: *IEEE Computer Graphics and Applications* 20.4 (2000), pp. 68–75. DOI: 10.1109/38.851753.

[Bux92] William A. S. Buxton. "Telepresence: Integrating Shared Task and Person Spaces." In: *Proceedings of the Conference on Graphics Interface, GI'92*. Vancouver, British Columbia, Canada: Morgan Kaufmann Publishers Inc., 1992, pp. 123–129. DOI: 10.5555/155294.155309.

[Cai+10] Qin Cai, David Gallup, Cha Zhang, and Zhengyou Zhang. "3D Deformable Face Tracking with a Commodity Depth Camera." In: *Proceedings of the European Conference on Computer Vision Conference on Computer Vision: Part III*. ECCV'10. Heraklion, Crete, Greece: Springer-Verlag, 2010, pp. 229–242. DOI: 10.1007/978-3-642-15558-1_17.

[CM92] Thomas Caudell and David Mizell. "Augmented reality: an application of heads-up display technology to manual manufacturing processes." In: *Proceedings of the Hawaii International Conference on System Sciences*. Vol. ii. 1992, 659–669 vol.2. DOI: 10.1109/HICSS.1992.183317.

[CBF14] Olivier Chapuis, Anastasia Bezerianos, and Stelios Frantzeskakis. "Smarties: An Input System for Wall Display Development." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: ACM, 2014, pp. 2763–2772. DOI: 10.1145/2556288.2556956.

[CF17] Haiwei Chen and Henry Fuchs. "Supporting Free Walking in a Large Virtual Environment: Imperceptible Redirected Walking with an Immersive Distractor." In: *Proceedings of the Computer Graphics International Conference*. CGI '17. Yokohama, Japan: ACM, 2017. DOI: 10.1145/3095140.3095162.

[Che02] Milton Chen. "Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconference." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '02. Minneapolis, Minnesota, USA: ACM, 2002, pp. 49–56. DOI: 10.1145/503376.503386.

[Che+13] Weiya Chen, Anthony Plancoulaine, Nicolas Férey, Damien Touraine, Julien Nelson, and Patrick Bourdot. "6DoF Navigation in Virtual Worlds: Comparison of Joystick-Based and Head-Controlled Paradigms." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST '13. Singapore: ACM, 2013, pp. 111–114. DOI: 10.1145/2503713.2503754.

[Che+18] Xiang Chen, Ye Tao, Guanyun Wang, Runchang Kang, Tovi Grossman, Stelian Coros, and Scott E. Hudson. "Forte: User-Driven Generative Design." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, 2018, pp. 1–12. DOI: 10.1145/3173574.3174070.

[Che+19] Lung-Pan Cheng, Eyal Ofek, Christian Holz, and Andrew D Wilson. "VRoamer: Generating On-The-Fly VR Experiences While Walking inside Large, Unknown Real-World Building Environments." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 359–366. DOI: 10.1109/VR.2019.8798074.

[Cho+19] Kevin Chow, Caitlin Coyiuto, Cuong Nguyen, and Dongwook Yoon. "Challenges and Design Considerations for Multimodal Asynchronous Collaboration in VR." In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019). DOI: 10.1145/3359142.

[CB91] Herbert H. Clark and Susan E. Brennan. "Grounding in Communication." In: *Perspectives on Socially Shared Cognition*. Ed. by Lauren Resnick, Levine B., M. John, Stephanie Teasley, and D. American Psychological Association, 1991, pp. 259–292. DOI: 10.1037/10096-006.

[CM81] Herbert H. Clark and Catherine R. Marshall. "Definite Knowledge and Mutual Knowledge." In: *Elements of Discourse Understanding*. Ed. by Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag. Cambridge University Press, 1981, pp. 10–63.

[CSB83] Herbert H. Clark, Robert Schreuder, and Samuel Buttrick. "Common ground at the understanding of demonstrative reference." In: *Journal of Verbal Learning and Verbal Behavior* 22.2 (1983), pp. 245–258. DOI: 10.1016/S0022-5371(83)90189-5.

[Cof+13] Dane Coffey, Chi-Lun Lin, Arthur G Erdman, and Daniel F Keefe. "Design by Dragging: An Interface for Creative Forward and Inverse Design with Simulation Ensembles." In: *IEEE Trans. on Visualization and Computer Graphics* 19.12 (2013), pp. 2783–2791. DOI: 10.1109/TVCG.2013.147.

[Cof+12] Dane Coffey, Nicholas Malbraaten, Trung Bao Le, Iman Borazjani, Fotis Sotiropoulos, Arthur G. Erdman, and Daniel F. Keefe. "Interactive Slice WIM: Navigating and Interrogating Volume Data Sets Using a Multisurface, Multitouch VR Interface." In: *IEEE Transactions on Visualization and Computer Graphics* 18.10 (2012), pp. 1614–1626. DOI: 10.1109/TVCG.2011.283.

[CAK93] Michael Cohen, Shigeaki Aoki, and Nobuo Koizumi. "Augmented audio reality: telepresence/VR hybrid acoustic environments." In: *Proceedings of the IEEE International Workshop on Robot and Human Communication*. 1993, pp. 361–364. DOI: 10.1109/ROMAN.1993.367692.

[Con+23] Ben J. Congdon, Gun Woo (Warren) Park, Jingyi Zhang, and Anthony Steed. "Comparing Mixed Reality Agent Representations: Studies in the Lab and in the Wild." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST '23. Christchurch, New Zealand: ACM, 2023. DOI: 10.1145/3611659.3615719.

[CB04] Thomas Convard and Patrick Bourdot. "History Based Reactive Objects for Immersive CAD." In: *Proceedings of the Symposium on Solid Modeling and Applications*. SM '04. Genoa, Italy: Eurographics Association, 2004, pp. 291–296. DOI: 10.2312/sm.20041404.

[CAC21] Emmanuel Courtoux, Caroline Appert, and Olivier Chapuis. "Wall-Tokens: Surface Tangibles for Vertical Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: ACM, 2021. DOI: 10.1145/3411764.3445404.

[Cru+92] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. "The CAVE: Audio Visual Experience Automatic Virtual Environment." In: *Communication of the ACM* 35.6 (1992), pp. 64–72. DOI: 10.1145/129888.129892.

[Cze+03] Mary Czerwinski, Greg Smith, Tim Regan, Brian Meyers, George Robertson, and Gary Starkweather. "Toward Characterizing the Productivity Benefits of Very Large Displays." In: *Proceedings of the IFIP TC13 International Conference On Human-Computer Interaction*. INTERACT '03. IOS Press, 2003, pp. 9–16.

[De +13] Bruno R. De Araújo, Géry Casiez, Joaquim A. Jorge, and Martin Hachet. "Mockup Builder: 3D modeling on and above the surface." In: *Computers & Graphics* 37.3 (2013), pp. 165–178. DOI: 10.1016/j.cag.2012.12.005.

[De 67] Edward De Bono. *The use of lateral thinking*. London, England: Jonathan Cape, 1967.

[Déto6] Françoise Détienne. "Collaborative design: Managing task interdependencies and multiple perspectives." In: *Interacting with Computers* 18.1 (2006), pp. 1–20. DOI: 10.1016/j.intcom.2005.05.001.

[Dou+12] Mingsong Dou, Ying Shi, Jan-Michael Frahm, Henry Fuchs, Bill Mauchly, and Mod Marathe. "Room-sized informal telepresence system." In: *Proceedings of the IEEE Virtual Reality Workshops (VRW)*. 2012, pp. 15–18. DOI: 10.1109/VR.2012.6180869.

[DLTo6] Thierry Duval, Anatole Lécuyer, and Sébastien Thomas. "SkeweR: a 3D Interaction Technique for 2-User Collaborative Manipulation of Objects in Virtual Environments." In: *Proceedings of the 3D User Interfaces symposium (3DUI'o6)*. 2006, pp. 69–72. DOI: 10.1109/VR.2006.119.

[Elr+92] Scott Elrod et al. "Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations, and Remote Collaboration." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. Monterey, California, USA: ACM, 1992, pp. 599–607. DOI: 10.1145/142750.143052.

[FFT22a] Arthur Fages, Cédric Fleury, and Theophanis Tsandilas. "ARgus : système multi-vues pour collaborer à distance avec un utilisateur en réalité augmentée." In: *Proceedings of the conférence francophone sur l'Interaction Homme-Machine*. Demonstration. Namur, Belgium, 2022. URL: https://hal.science/hal-03762816.

[FFT22b] Arthur Fages, Cédric Fleury, and Theophanis Tsandilas. "Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users." In: *Proceedings of the ACM on Human-Computer Interaction* 6. CSCW2 (2022). DOI: 10.1145/3555607.

[FFT23] Arthur Fages, Cédric Fleury, and Theophanis Tsandilas. "Conception collaborative au travers de versions parallèles en réalité augmentée." In: *Proceedings of the Conference on l'Interaction Humain-Machine*. IHM '23. TROYES, France: ACM, 2023. DOI: 10.1145/3583961.3583978.

[FMS93] Steven Feiner, Blair Macintyre, and Dorée Seligmann. "Knowledge-Based Augmented Reality." In: *Communications of the ACM* 36.7 (1993), pp. 53–62. DOI: 10.1145/159544.159587.

[FH22] Andreas Rene Fender and Christian Holz. "Causality-Preserving Asynchronous Reality." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '22. New Orleans, LA, USA: ACM, 2022. DOI: 10.1145/3491102.3501836.

[Fio+02] Michele Fiorentino, Raffaele de Amicis, Giuseppe Monno, and Andre Stork. "Spacedesign: A Mixed Reality Workspace for Aesthetic Industrial Design." In: *Proceedings of the International Symposium on Mixed and Augmented Reality*. ISMAR '02. IEEE Computer Society, 2002, p. 86. DOI: 10.1109/ISMAR.2002.1115077.

[Fle+04] Timo Fleisch, Gino Brunetti, Pamela Santos, and André Stork. "Stroke-input methods for immersive styling environments." In: *Proceedings of the Shape Modeling Applications conference*. 2004, pp. 275–283. DOI: 10.1109/SMI.2004.1314514.

[Fle+15a] Cédric Fleury, Ignacio Avellino, Michel Beaudouin-Lafon, and Wendy E. Mackay. "Telepresence Systems for Large Interactive Spaces." In: *Workshop on Everyday Telepresence: Emerging Practices and Future Research Directions at the ACM conference on Human Factors in Computing Systems*. Seoul, South Korea, 2015. URL: https://hal.science/hal-01242304.

[Fle+10a] Cédric Fleury, Alain Chauffaut, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. "A Generic Model for Embedding Users' Physical Workspaces into Multi-Scale Collaborative Virtual Environments." In: *Proceedings of the International Conference on Artificial Reality and Telexistence*. ICAT'10. Adelaide, Australia, 2010. URL: https://hal.science/inria-00534096.

[Fle+10b] Cédric Fleury, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. "A New Adaptive Data Distribution Model for Consistency Maintenance in Collaborative Virtual Environments." In: *Proceedings of the Joint Virtual Reality Conference of EuroVR - EGVE - VEC*. EGVE - JVRC'10. Stuttgart, Germany: Eurographics Association, 2010, pp. 29–36. DOI: 10.2312/EGVE/JVRC10/029-036.

[Fle+10c] Cédric Fleury, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. "Architectures and Mechanisms to Maintain efficiently Consistency in Collaborative Virtual Environments." In: *Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS) at the IEEE Conference on Virtual Reality*. Waltham, United States, 2010. URL: https://hal.science/inria-00534082.

[Fle+12] Cédric Fleury, Thierry Duval, Valérie Gouranton, and Anthony Steed. "Evaluation of Remote Collaborative Manipulation for Scientific Data Analysis." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST'12. Toronto, Ontario, Canada: ACM, 2012, pp. 129–136. DOI: 10.1145/2407336.2407361.

[Fle+15b] Cédric Fleury, Nicolas Férey, Jean-Marc Vézien, and Patrick Bourdot. "Remote collaboration across heterogeneous large interactive spaces." In: *Workshop on Collaborative Virtual Environments (3DCVE)*

***at the IEEE Conference on Virtual Reality*. 2015, pp. 9–10. DOI: 10.1109/3DCVE.2015.7153591.**

[Fle+14] **Cédric Fleury, Tiberiu Popa, Tat Jen Cham, and Henry Fuchs. "Merging Live and pre-Captured Data to support Full 3D Head Reconstruction for Telepresence." In: *Proceedings of the Conference of the European Association for Computer Graphics*. EG'14. Strasbourg, France, 2014, pp. 9–12. DOI: 10.2312/egsh.20141002.**

[FM21] Guo Freeman and Divine Maloney. "Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality." In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW3 (2021). DOI: 10.1145/3432938.

[Fri+21] Sebastian J. Friston, Ben J. Congdon, David Swapp, Lisa Izzouzi, Klara Brandstätter, Daniel Archer, Otto Olkkonen, Felix Johannes Thiel, and Anthony Steed. "Ubiq: A System to Build Flexible Social Virtual Reality Experiences." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST '21. Osaka, Japan: ACM, 2021. DOI: 10.1145/3489849.3489871.

[FSB14] Henry Fuchs, Andrei State, and Jean-Charles Bazin. "Immersive 3D Telepresence." In: *Computer* 47.7 (2014), pp. 46–52. DOI: 10.1109/MC.2014.185.

[Fuc+96] Henry Fuchs, Andrei State, Etta D. Pisano, William F. Garrett, Gentaro Hirota, Mark A. Livingston, Mary C. Whitton, and Stephen M. Pizer. "Towards Performing Ultrasound-Guided Needle Biopsies from within a Head-Mounted Display." In: *Proceedings of the International Conference on Visualization in Biomedical Computing*. VBC '96. Berlin, Heidelberg: Springer-Verlag, 1996, pp. 591–600. DOI: 10.1007/BFb0047002.

[FMG11] Philippe Fuchs, Guillaume Moreau, and Pascal Guitton. *Virtual Reality: Concepts and Technologies*. 1st. CRC Press, 2011. DOI: 10.1201/b11612.

[Fus+04] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D.I. Kramer. "Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks." In: *Human–Computer Interaction* 19.3 (2004), pp. 273–309. DOI: 10.1207/s15327051hci1903_3.

[Fyf+18] **Lawrence Fyfe, Olivier Gladin, Cédric Fleury, and Michel Beaudouin-Lafon. "Combining Web Audio Streaming, Motion Capture, and Binaural Audio in a Telepresence System." In: *Proceedings of the International Web Audio conference*. WAC'18. Berlin, Germany, 2018. URL: https://hal.science/hal-01957843.**

[Gau+14] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. "World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: ACM, 2014, pp. 449–459. DOI: 10.1145/2642918.2647372.

[GBR12] Florian Geyer, Jochen Budzinski, and Harald Reiterer. "IdeaVis: A Hybrid Workspace and Interactive Visualization for Paper-Based Collaborative Sketching Sessions." In: *Proceedings of the Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. NordiCHI '12. Copenhagen, Denmark: ACM, 2012, pp. 331–340. DOI: 10.1145/2399016.2399069.

[Gig+14] Dominik Giger, Jean-Charles Bazin, Claudia Kuster, Tiberiu Popa, and Markus Gross. "Gaze correction with a single webcam." In: *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*. 2014, pp. 1–6. DOI: 10.1109/ICME.2014.6890306.

[Got+18] Daniel Gotsch, Xujing Zhang, Timothy Merritt, and Roel Vertegaal. "TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-Size 3D Human Telepresence." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, 2018, pp. 1–10. DOI: 10.1145/3173574.3174096.

[Grø+21] Jens Emil Grønbæk, Banu Saatçi, Carla F. Griggio, and Clemens Nylandsted Klokmose. "MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: ACM, 2021. DOI: 10.1145/3411764.3445698.

[Gro+01] Tovi Grossman, Ravin Balakrishnan, Gordon Kurtenbach, George Fitzmaurice, Azam Khan, and Bill Buxton. "Interaction Techniques for 3D Modeling on Large Displays." In: *Proceedings of the Symposium on Interactive 3D Graphics*. I3D '01. ACM, 2001, pp. 17–23. DOI: 10.1145/364338.364341.

[Gru01] Jonathan Grudin. "Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. Seattle, Washington, USA: ACM, 2001, pp. 458–465. DOI: 10.1145/365024.365312.

[Gug+17] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. "ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, 2017, pp. 4021–4033. DOI: 10.1145/3025453.3025683.

[Gui+16] Hao Gui, Rong Zheng, Chao Ma, Hao Fan, and Liya Xu. "An Architecture for Healthcare Big Data Management and Analysis." In: *Health Information Science (HIS)*. Vol. 10038. Lecture Notes in Computer Science. Springer, 2016, pp. 154–160. DOI: 10.1007/978-3-319-48335-1_17.

[Hab+18] Jacob Habgood, David Moore, David Wilson, and Sergio Alapont. "Rapid, Continuous Movement Between Nodes as an Accessible Virtual Reality Locomotion Technique." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018, pp. 371–378. DOI: 10.1109/VR.2018.8446130.

[HCC07] Mark Hancock, Sheelagh Carpendale, and Andy Cockburn. "Shallow-depth 3D Interaction: Design and Evaluation of One-, Two- and Three-touch Techniques." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. San Jose, California, USA: ACM, 2007, pp. 1147–1156. DOI: 10.1145/1240624.1240798.

[HCC09] Mark Hancock, Thomas ten Cate, and Sheelagh Carpendale. "Sticky Tools: Full 6DOF Force-based Interaction for Multi-touch Tables." In: *Proceedings of the International Conference on Interactive Tabletops and Surfaces*. ITS '09. Banff, Alberta, Canada: ACM, 2009, pp. 133–140. DOI: 10.1145/1731903.1731930.

[Han97] Chris Hand. "A Survey of 3D Interaction Techniques." In: *Computer Graphics Forum* 16.5 (1997), pp. 269–281. DOI: 10.1111/1467-8659.00194.

[HC02] Denise Y.P. Henriques and John D. Crawford. "Role of Eye, Head, and Shoulder Geometry in the Planning of Accurate Arm Movements." In: *Journal of Neurophysiology* 87.4 (2002), pp. 1677–1685. DOI: 10.1152/jn.00509.2001.

[HYS16] Keita Higuch, Ryo Yonetani, and Yoichi Sato. "Can Eye Help You? Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, 2016, pp. 5180–5190. DOI: 10.1145/2858036.2858438.

[Hig+15] Keita Higuchi, Yinpeng Chen, Philip A. Chou, Zhengyou Zhang, and Zicheng Liu. "ImmerseBoard: Immersive Telepresence Experience Using a Digital Whiteboard." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015, pp. 2383–2392. DOI: 10.1145/2702123.2702160.

[Hof98] Hunter G. Hoffman. "Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments." In: *Proceedings of the IEEE Virtual Reality Annual International Symposium*. 1998, pp. 59–63. DOI: 10.1109/VRAIS.1998.658423.

[HS92] Jim Hollan and Scott Stornetta. "Beyond Being There." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. Monterey, California, USA: ACM, 1992, pp. 119–125. DOI: 10.1145/142750.142769.

[HRB97] Michael P. Hollier, Andrew N. Rimell, and D. Burraston. "Spatial audio technology for telepresence." In: *BT Technology Journal* 15 (1997), pp. 33–41. DOI: 10.1023/A:1018675327815.

[Hor+18] Tom Horak, Sriram Karthik Badam, Niklas Elmqvist, and Raimund Dachselt. "When David Meets Goliath: Combining Smartwatches with a Large Vertical Display for Visual Data Exploration." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, 2018, pp. 1–13. DOI: 10.1145/3173574.3173593.

[IIH07] Takeo Igarashi, Takeo Igarashi, and John F. Hughes. "A Suggestive Interface for 3D Drawing." In: *ACM SIGGRAPH 2007 Courses*. SIGGRAPH '07. San Diego, California: ACM, 2007. DOI: 10.1145/1281500.1281531.

[Irl+16] Andrew Irlitti, Ross T. Smith, Stewart Von Itzstein, Mark Billinghurst, and Bruce H. Thomas. "Challenges for Asynchronous Collaboration in Augmented Reality." In: *Adjunct proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 2016, pp. 31–35. DOI: 10.1109/ISMAR-Adjunct.2016.0032.

[IT93] Ellen A. Isaacs and John C. Tang. "What Video Can and Can'T Do for Collaboration: A Case Study." In: *Proceedings of the International Conference on Multimedia*. MULTIMEDIA '93. Anaheim, California, USA: ACM, 1993, pp. 199–206. DOI: 10.1145/166266.166289.

[IK92] Hiroshi Ishii and Minoru Kobayashi. "ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. Monterey, California, USA: ACM, 1992, pp. 525–532. DOI: 10.1145/142750.142977.

[Isr+09] Johann Israel, Eva Wiese, Magdalena Mateescu, Christian Zöllner, and Rainer Stark. "Investigating three-dimensional sketching for early conceptual design—Results from expert discussions and user studies." In: *Computers & Graphics* 33.4 (2009), pp. 462–473. DOI: 10.1016/j.cag.2009.05.005.

[Iza+11] Shahram Izadi et al. "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '11. ACM, 2011, pp. 559–568. DOI: 10.1145/2047196.2047270.

[JSK12] Bret Jackson, David Schroeder, and Daniel F. Keefe. "Nailing Down Multi-Touch: Anchored Above the Surface Interaction for 3D Modeling and Navigation." In: *Proceedings of the Graphics Interface 2012*. GI '12. Canadian Information Processing Society, 2012, pp. 181–184. URL: https://dl.acm.org/doi/10.5555/2305276.2305306.

[JFH94] Richard H. Jacoby, Mark Ferneau, and Jim Humphries. "Gestural interaction in a virtual environment." In: *Stereoscopic Displays and Virtual Reality Systems*. Ed. by Scott S. Fisher, John O. Merritt, and Mark T. Bolas. Vol. 2177. International Society for Optics and Photonics. SPIE, 1994, pp. 355–364. DOI: 10.1117/12.173892.

[JH14] Mikkel R. Jakobsen and Kasper Hornbæk. "Up Close and Personal: Collaborative Work on a High-Resolution Multitouch Wall Display." In: *ACM Transactions on Computer-Human Interaction* 21.2 (2014). DOI: 10.1145/2576099.

[JH16] Mikkel R. Jakobsen and Kasper Hornbæk. "Negotiating for Space? Collaborative Work Using a Wall Display with Mouse and Touch Input." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, 2016, pp. 2050–2061. DOI: 10.1145/2858036.2858158.

[JBC23] Raphaël James, Anastasia Bezerianos, and Olivier Chapuis. "Evaluating the Extension of Wall Displays with AR for Collaborative Work." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: ACM, 2023. DOI: 10.1145/3544548.3580752.

[Jan+20] Pascal Jansen, Fabian Fischbach, Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. "ShARe: Enabling Co-Located Asymmetric Multi-User Interaction for Augmented Reality Head-Mounted Displays." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '20. Virtual Event, USA: ACM, 2020, pp. 459–471. DOI: 10.1145/3379337.3415843.

[JSH19] Yvonne Jansen, Jonas Schjerlund, and Kasper Hornbæk. "Effects of Locomotion and Visual Overview on Spatial Memory When Interacting with Wall Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, pp. 1–12. DOI: 10.1145/3290605.3300521.

[Jon+21] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. "Belonging There: VROOM-Ing into the Uncanny Valley of XR Telepresence." In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (2021). DOI: 10.1145/3449133.

[Kam+18] Adhitya Kamakshidasan, José Galaz, Rodrigo Cienfuegos, Antoine Rousseau, and Emmanuel Pietriga. "Comparative Visualization of Deep Water Asteroid Impacts on Ultra-high-resolution Wall Displays with Seawall." In: *Proceedings of the IEEE Scientific Visualization Conference*. SciVis'18. 2018, pp. 142–145. DOI: 10.1109/SciVis.2018.8823616.

[KRF11] Anette von Kapri, Tobias Rick, and Steven Feiner. "Comparing steering-based travel techniques for search tasks in a CAVE." In: *Proceedings of the IEEE Virtual Reality Conference (VR)*. 2011, pp. 91–94. DOI: 10.1109/VR.2011.5759443.

[Kaz+17] Rubaiat Habib Kazi, Tovi Grossman, Hyunmin Cheong, Ali Hashemi, and George Fitzmaurice. "DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '17. Québec City, QC, Canada: ACM, 2017, pp. 401–414. DOI: 10.1145/3126594.3126662.

[KZL07] Daniel F. Keefe, Robert C. Zeleznik, and David H. Laidlaw. "Drawing on Air: Input Techniques for Controlled 3D Line Illustration." In: *IEEE Transactions on Visualization and Computer Graphics* 13.5 (2007), pp. 1067–1081. DOI: 10.1109/TVCG.2007.1060.

[KD07] Fakheredine Keyrouz and Klaus Diepold. "Binaural Source Localization and Spatial Audio Reproduction for Telepresence Applications." In: *Presence: Teleoperators and Virtual Environments* 16.5 (2007), pp. 509–522. DOI: 10.1162/pres.16.5.509.

[Kha+05] Azam Khan, Justin Matejka, George Fitzmaurice, and Gordon Kurtenbach. "Spotlight: Directing Users' Attention on Large Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: ACM, 2005, pp. 791–798. DOI: 10.1145/1054972.1055082.

[Kis+17] Ulrike Kister, Konstantin Klamka, Christian Tominski, and Raimund Dachselt. "GraSp: Combining Spatially-Aware Mobile Devices and a Display Wall for Graph Visualization and Interaction." In: *Computer Graphics Forum* 36.3 (2017), pp. 503–514. DOI: 10.1111/cgf.13206.

[Kiy+02] Kiyoshi Kiyokawa, Mark Billinghurst, Sean Hayes, Anoop Gupta, Yuki Sannohe, and Hirokazu Kato. "Communication behaviors of co-located users in collaborative AR interfaces." In: *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*. 2002, pp. 139–148. DOI: 10.1109/ISMAR.2002.1115083.

[KMI19] Anya Kolesnichenko, Joshua McVeigh-Schultz, and Katherine Isbister. "Understanding Emerging Design Practices for Avatar Systems in the Commercial Social VR Ecology." In: *Proceedings of the Designing Interactive Systems Conference*. DIS '19. San Diego, CA, USA: ACM, 2019, pp. 241–252. DOI: 10.1145/3322276.3322352.

[KS04] Vladislav Kraevoy and Alla Sheffer. "Cross-parameterization and Compatible Remeshing of 3D Models." In: *ACM Transactions on Graphics (siggraph'04)* 23.3 (2004), pp. 861–869. DOI: 10.1145/1015706.1015811.

[KK06] Martin Kuechler and Andreas Kunz. "HoloPort - A Device for Simultaneous Video and Data Conferencing Featuring Gaze Awareness." In: *Proceedings of IEEE Virtual Reality Conference (VR)*. 2006, pp. 81–88. DOI: 10.1109/VR.2006.71.

[Kul+11] Alexander Kulik, André Kunert, Stephan Beck, Roman Reichel, Roland Blach, Armin Zink, and Bernd Froehlich. "C1x6: A Stereoscopic Six-User Display for Co-Located Collaboration in Shared Virtual Environments." In: *ACM Transactions on Graphics* 30.6 (2011), pp. 1–12. DOI: 10.1145/2070781.2024222.

[Kun+14] André Kunert, Alexander Kulik, Stephan Beck, and Bernd Froehlich. "Photoportals: Shared References in Space and Time." In: *Proceedings of the Conference on Computer Supported Cooperative Work & Social Computing*. CSCW '14. Baltimore, Maryland, USA: ACM, 2014, pp. 1388–1399. DOI: 10.1145/2531602.2531727.

[LaV+17] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. *3D User Interfaces: Theory and Practice, 2nd edition*. Addison-Wesley usability and HCI series. Addison Wesley Longman Publishing Co., Inc., 2017. ISBN: 9780134034324.

[Le+19] Khanh-Duy Le, Ignacio Avellino, Cédric Fleury, Morten Fjeld, and Andreas M. Kunz. "GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration." In: *Proceedings of the IFIP TC13 International Conference On Human-Computer Interaction (INTERACT)*. Vol. 11747. Lecture Notes in Computer Science. Paphos,

Cyprus: Springer, 2019, pp. 282–303. DOI: 10.1007/978-3-030-29384-0_18.

[Léc+23] **Aurélien Léchappé, Aurélien Milliat, Cédric Fleury, Mathieu Chollet, and Cédric Dumas. "Characterization of Collaboration in a Virtual Environment with Gaze and Speech Signals." In: *Companion Publication of the International Conference on Multimodal Interaction*. ICMI '23 Companion. Paris, France: ACM, 2023, pp. 21–25. DOI: 10.1145/3610661.3617149.**

[LMR14] Nicolas H. Lehment, Daniel Merget, and Gerhard Rigoll. "Creating automatically aligned consensus realities for AR videoconferencing." In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2014, pp. 201–206. DOI: 10.1109/ISMAR.2014.6948428.

[LD97] Valerie D. Lehner and Thomas A. DeFanti. "Distributed virtual reality: supporting remote collaboration in vehicle design." In: *IEEE Computer Graphics and Applications* 17.2 (1997), pp. 13–17. DOI: 10.1109/38.574654.

[Lei+96] Jason Leigh, Andrew E. Johnson, Christina A. Vasilakis, and Thomas A. DeFanti. "Multi-perspective collaborative design in persistent networked virtual environments." In: *Proceedings of the IEEE Virtual Reality Annual International Symposium*. 1996, pp. 253–260. DOI: 10.1109/VRAIS.1996.490535.

[Li+13] Hao Li, Jihun Yu, Yuting Ye, and Chris Bregler. "Realtime Facial Animation with On-the-fly Correctives." In: *ACM Transactions on Graphics (siggraph'13)* 32.4 (2013), 42:1–10. DOI: 10.1145/2461912.2462019.

[Lis+15] Lars Lischke, Sven Mayer, Katrin Wolf, Niels Henze, Albrecht Schmidt, Svenja Leifert, and Harald Reiterer. "Using Space: Effect of Display Size on Users' Search Performance." In: *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. CHI EA '15. Seoul, Republic of Korea: ACM, 2015, pp. 1845–1850. DOI: 10.1145/2702613.2732845.

[Liu15] Can Liu. "Embodied Interaction for Data Manipulation Tasks on Wall-sized Displays." PhD thesis. Université Paris-Saclay, 2015. URL: https://theses.hal.science/tel-01264670v2.

[Liu+16] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. "Shared Interaction on a Wall-Sized Display in a Data Manipulation Task." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, 2016, pp. 2075–2086. DOI: 10.1145/2858036.2858039.

[Liu+17] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. "CoReach: Cooperative Gestures for Data Manipulation on Wall-Sized Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, 2017, pp. 6730–6741. DOI: 10.1145/3025453.3025594.

[Liu+14] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E. Mackay. "Effects of Display Size and Navigation Type on a Classification Task." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: ACM, 2014, pp. 4147–4156. DOI: 10.1145/2556288.2557020.

[Liu+18] James Liu, Hirav Parekh, Majed Al-Zayer, and Eelke Folmer. "Increasing Walking in VR Using Redirected Teleportation." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '18. Berlin, Germany: ACM, 2018, pp. 521–529. DOI: 10.1145/3242587.3242601.

[LBC04] Julian Looser, Mark Billinghurst, and Andy Cockburn. "Through the Looking Glass: The Use of Lenses as an Interface Tool for Augmented Reality Interfaces." In: *Proceedings of the International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*. GRAPHITE '04. Singapore: ACM, 2004, pp. 204–211. DOI: 10.1145/988834.988870.

[Lop+16] David Lopez, Lora Oehlberg, Candemir Doger, and Tobias Isenberg. "Towards An Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration." In: *IEEE Transactions on Visualization and Computer Graphics* 22.5 (2016), pp. 1616–1629. DOI: 10.1109/TVCG.2015.2440233.

[LF16] Jean-Baptiste Louvet and Cédric Fleury. "Combining Bimanual Interaction and Teleportation for 3D Manipulation on Multi-Touch Wall-Sized Displays." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST'16. Munich, Germany: ACM, 2016, pp. 283–292. DOI: 10.1145/2993369.2993390.

[Luf+15] Paul K. Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. "Flexible Ecologies And Incongruent Locations." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015, pp. 877–886. DOI: 10.1145/2702123.2702286.

[Ma+21] Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. "Pixel Codec Avatars." In: *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2021, pp. 64–73. DOI: 10.1109/CVPR46437.2021.00013.

[Mac99] Wendy E. Mackay. "Media Spaces: Environments for multimedia interaction." In: *Computer-Supported Cooperative Work*. Ed. by Michel Beaudouin-Lafon. Trends in Software Series. Chichester: Wiley and Sons, 1999, pp. 55–82. ISBN: 0-471-96736-X.

[Mah+10] Morad Mahdjoub, Davy Monticolo, Samuel Gomes, and Jean-Claude Sagot. "A Collaborative Design for Usability Approach Supported by Virtual Reality and a Multi-Agent System Embedded in a PLM Environment." In: *Computer-Aided Design* 42.5 (2010), pp. 402–413. DOI: 10.1016/j.cad.2009.02.009.

[Mah+19] Tahir Mahmood, Willis Fulmer, Neelesh Mungoli, Jian Huang, and Aidong Lu. "Improving Information Sharing and Collaborative Analysis for Remote GeoSpatial Visualization Using Mixed Reality." In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2019, pp. 236–247. DOI: 10.1109/ISMAR.2019.00021.

[MHG12] Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. "Cross-Device Interaction via Micro-Mobility and f-Formations." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '12. Cambridge, Massachusetts, USA: ACM, 2012, pp. 13–22. DOI: 10.1145/2380116.2380121.

[Mar+17] Pierre Martin, Stéphane Masfrand, Yujiro Okuya, and Patrick Bourdot. "A VR-CAD Data Model for Immersive Design." In: *Augmented Reality, Virtual Reality, and Computer Graphics*. Cham: Springer International Publishing, 2017, pp. 222–241. DOI: 10.1007/978-3-319-60922-5_17.

[MCG10a] Anthony Martinet, Gery Casiez, and Laurent Grisoni. "The Design and Evaluation of 3D Positioning Techniques for Multi-touch Displays." In: *Proceedings of the IEEE Symposium on 3D User Interfaces*. 3DUI '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 115–118. DOI: 10.1109/3DUI.2010.5444709.

[MCG10b] Anthony Martinet, Géry Casiez, and Laurent Grisoni. "The Effect of DOF Separation in 3D Manipulation Tasks with Multi-touch Displays." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST '10. Hong Kong: ACM, 2010, pp. 111–118. DOI: 10.1145/1889863.1889888.

[Mil+95] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. "Augmented reality: a class of displays on the reality-virtuality continuum." In: *Telemanipulator and Telepresence Technologies*. Ed. by Hari Das. Vol. 2351. International Society for Optics and Photonics. SPIE, 1995, pp. 282–292. DOI: 10.1117/12.197321.

[Mil+93] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. "Applications of augmented reality for human-robot communication." In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93)*. Vol. 3. 1993, 1467–1472 vol.3. DOI: 10.1109/IROS.1993.583833.

[Min+20] Dae-Hong Min, Dong-Yong Lee, Yong-Hun Cho, and In-Kwon Lee. "Shaking Hands in Virtual Space: Recovery in Redirected Walking for Direct Interaction between Two Users." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 164–173. DOI: 10.1109/VR46266.2020.00035.

[Min95] Mark R. Mine. *Virtual Environment Interaction Techniques*. Tech. rep. University of North Carolina at Chapel Hill, 1995.

[Moh+20] Peter Mohr, Shohei Mori, Tobias Langlotz, Bruce H. Thomas, Dieter Schmalstieg, and Denis Kalkofen. "Mixed Reality Light Fields for Interactive Remote Assistance." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: ACM, 2020, pp. 1–12. DOI: 10.1145/3313831.3376289.

[MG02] Andrew F. Monk and Caroline Gale. "A Look Is Worth a Thousand Words: Full Gaze Awareness in Video-Mediated Conversation." In: *Discourse Processes* 33.3 (2002), pp. 257–278. DOI: 10.1207/S15326950DP3303_4.

[MMK12] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. "The Uncanny Valley [From the Field]." In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100. DOI: 10.1109/MRA.2012.2192811.

[MSH04] Tariq Mujber, Tamas Szecsi, and Saleem Hashmi. "Virtual reality applications in manufacturing process simulation." In: *Journal of Materials Processing Technology* 155-156 (2004), pp. 1834–1838. DOI: 10.1016/j.jmatprotec.2004.04.401.

[Nan+15] Mathieu Nancel, Emmanuel Pietriga, Olivier Chapuis, and Michel Beaudouin-Lafon. "Mid-Air Pointing on Ultra-Walls." In: *ACM Transactions on Computer-Human Interaction* 22.5 (2015). DOI: 10.1145/2766448.

[NC05] David Nguyen and John Canny. "MultiView: Spatially Faithful Group Video Conferencing." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: ACM, 2005, pp. 799–808. DOI: 10.1145/1054972.1055084.

[NM97] Haruo Noma and Tsutomu Miyasato. "Cooperative Object Manipulation in Virtual Space Using Virtual Physics." In: *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*. Vol. Dynamic Systems and Control. Nov. 1997, pp. 101–106. DOI: 10.1115/IMECE1997-0383.

[Obs22] Rubin Observatory. *Data Management*. https://www.lsst.org/about/dm, accessed: 2022-03-23. 2022.

[OF09] Ohan Oda and Steven Feiner. "Interference avoidance in multi-user hand-held augmented reality." In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. 2009, pp. 13–22. DOI: 10.1109/ISMAR.2009.5336507.

[Oku+20] Yujiro Okuya, Olivier Gladin, Nicolas Ladévèze, Cédric Fleury, and Patrick Bourdot. "Investigating Collaborative Exploration of Design Alternatives on a Wall-Sized Display." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'20. Virtual Event: ACM, 2020, pp. 1–12. DOI: 10.1145/3313831.3376736.

[Oku+18a] Yujiro Okuya, Nicolas Ladévèze, Cédric Fleury, and Patrick Bourdot. "ShapeGuide: Shape-Based 3D Interaction for Parameter Modification of Native CAD Data." In: *Frontiers in Robotics and AI* 5 (2018). DOI: 10.3389/frobt.2018.00118.

[Oku+18b] Yujiro Okuya, Nicolas Ladévèze, Olivier Gladin, Cédric Fleury, and Patrick Bourdot. "Distributed Architecture for Remote Collaborative Modification of Parametric CAD Data." In: *Workshop on Collaborative Virtual Environments (3DCVE) at the IEEE Conference on Virtual Reality*. Reutlingen, Germany, 2018, pp. 1–4. DOI: 10.1109/3DCVE.2018.8637112.

[Oku+21] **Yujiro Okuya, Nicolas Ladévèze, Olivier Gladin, Cédric Fleury, and Patrick Bourdot. "Collaborative VR-CAD for Industrial Product Design: CAD Parameter Modification with 3D Interaction on Heterogeneous Immersive Platforms." In: *Manufacturing in the Era of 4th Industrial Revolution*. Ed. by Monica Bordegoni, Satyandra K. Gupta, and James Ritchie. Vol. 3. World Scientific, 2021. Chap. 2, pp. 17–47. DOI: 10.1142/9789811222863_0002.**

[Ort+16] Sergio Orts-Escolano et al. "Holoportation: Virtual 3D Teleportation in Real-Time." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '16. Tokyo, Japan: ACM, 2016, pp. 741–754. DOI: 10.1145/2984511.2984517.

[Ots16] Kazuhiro Otsuka. "MMSpace: Kinetically-augmented telepresence for small group-to-group conversations." In: *Proceedings of IEEE Virtual Reality (VR)*. 2016, pp. 19–28. DOI: 10.1109/VR.2016.7504684.

[PS14] Ye Pan and Anthony Steed. "A Gaze-Preserving Situated Multiview Telepresence System." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: ACM, 2014, pp. 2173–2176. DOI: 10.1145/2556288.2557320.

[PL06] Fred Pighin and John P. Lewis. "Performance-Driven Facial Animation." In: *ACM Siggraph Courses*. 2006. DOI: 10.1145/1185657.1185841. URL: http://scribblethink.org/Courses/PDFA/2006_COURSE30.pdf.

[PBF08] Marcio S. Pinho, Doug A. Bowman, and Carla M. Dal Sasso Freitas. "Cooperative object manipulation in collaborative virtual environments." In: *Journal of the Brazilian Computer Society* 14 (2008), pp. 53–67. DOI: 10.1007/BF03192559.

[PBF02] Márcio S. Pinho, Doug A. Bowman, and Carla M.D.S. Freitas. "Cooperative Object Manipulation in Immersive Virtual Environments: Framework and Techniques." In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '02. Hong Kong, China: ACM, 2002, pp. 171–178. DOI: 10.1145/585740.585769.

[Piu+19] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. "On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, pp. 1–17. DOI: 10.1145/3290605.3300458.

[PBN22] Carole Plasson, Renaud Blanch, and Laurence Nigay. "Selection Techniques for 3D Extended Desktop Workstation with AR HMD." In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2022, pp. 460–469. DOI: 10.1109/ISMAR55827.2022.00062.

[Pou+96] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. "The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '96. Seattle, Washington, USA: ACM, 1996, pp. 79–80. DOI: 10.1145/237091.237102.

[PBC17] Arnaud Prouzeau, Anastasia Bezerianos, and Olivier Chapuis. "Evaluating Multi-User Selection for Exploring Graph Topology on Wall-Displays." In: *IEEE Transactions on Visualization and Computer Graphics* 23.8 (2017), pp. 1936–1951. DOI: 10.1109/TVCG.2016.2592906.

[PBC18] Arnaud Prouzeau, Anastasia Bezerianos, and Olivier Chapuis. "Awareness Techniques to Aid Transitions between Personal and Shared Workspaces in Multi-Display Environments." In: *Proceedings of the International Conference on Interactive Surfaces and Spaces*. ISS'18. Tokyo, Japan: ACM, 2018, pp. 291–304. DOI: 10.1145/3279778.3279780.

[RR14] Wullianallur Raghupathi and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." In: *Health Information Science and Systems* 2.3 (2014). DOI: 10.1186/2047-2501-2-3.

[Ran+01] Dirk Rantzau, Franz Maurer, C. Mayer, Robin Löffler, Oliver Riedel, Holger R. Scharm, and D. Banek. "The integration of immersive Virtual Reality applications into Catia V5." In: *Immersive Projection Technology and Virtual Environments*. Vienna: Springer Vienna, 2001, pp. 93–102. DOI: 10.1007/978-3-7091-6221-7_10.

[Rap+09] Alberto Raposo, Ismael Santos, Luciano Soares, Gustavo Wagner, Eduardo Corseuil, and Marcelo Gattass. "Environ: Integrating VR and CAD in Engineering Projects." In: *IEEE Computer Graphics and Applications* 29.6 (2009), pp. 91–95. DOI: 10.1109/MCG.2009.118.

[RKW01] Sharif Razzaque, Zachariah Kohn, and Mary C. Whitton. "Redirected Walking." In: *Proceedings of the Eurographics conference - Short Presentations*. Eurographics Association, 2001. DOI: 10.2312/egs.20011036.

[RS02] Holger Regenbrecht and Thomas Schubert. "Real and illusory interactions enhance presence in virtual environments." In: *Presence: Teleoperators & Virtual Environments* 11.4 (2002), pp. 425–434. DOI: 10.1162/105474602760204318.

[RDH09] Jason L. Reisman, Philip L. Davidson, and Jefferson Y. Han. "A Screen-space Formulation for 2D and 3D Direct Manipulation." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '09. Victoria, BC, Canada: ACM, 2009, pp. 69–78. DOI: 10.1145/1622176.1622190.

[Rek96] Jun Rekimoto. "Transvision: A hand-held augmented reality system for collaborative design." In: *Proceedings of Virtual Systems and Multi-Media (VSMM '96)*. Jan. 1996.

[Rie+06] Kai Riege, Thorsten Holtkamper, Gerold Wesche, and Bernd Froehlich. "The Bent Pick Ray: An Extended Pointing Technique for Multi-User Interaction." In: *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI'06)*. 2006, pp. 62–65. DOI: 10.1109/VR.2006.127.

[Rim22] Meghan Rimol. *Gartner Survey Reveals a 44% Rise in Workers' Use of Collaboration Tools Since 2019*. https://www.gartner.com/en/newsroom/press-releases/2021-08-23-gartner-survey-reveals-44-percent-rise-in-workers-use-of-collaboration-tools-since-2019, accessed: 2022-03-23. 2022.

[RH17] Joan Sol Roo and Martin Hachet. "One Reality: Augmenting How the Physical World is Experienced by Combining Multiple Mixed Reality Modalities." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '17. Québec City, QC, Canada: ACM, 2017, pp. 787–795. DOI: 10.1145/3126594.3126638.

[RSJ02] Roy A. Ruddle, Justin C. D. Savage, and Dylan M. Jones. "Symmetric and Asymmetric Action Integration during Cooperative Object Manipulation in Virtual Environments." In: *ACM Transactions on Computer-Human Interaction* 9.4 (2002), pp. 285–308. DOI: 10.1145/586081.586084.

[Rud+16] Roy A. Ruddle, Rhys G. Thomas, Rebecca Randell, Philip Quirke, and Darren Treanor. "The Design and Evaluation of Interfaces for Navigating Gigapixel Images in Digital Pathology." In: *ACM Transactions on Computer-Human Interaction* 23.1 (2016). DOI: 10.1145/2834117.

[SJF09] Holger Salzmann, Jan Jacobs, and Bernd Froehlich. "Collaborative Interaction in Co-Located Two-User Scenarios." In: *Proceedings of the Joint Virtual Reality Eurographics Conference on Virtual Environments*. JVRC'09. Lyon, France: Eurographics Association, 2009, pp. 85–92. DOI: 10.2312/EGVE/JVRC09/085-092.

[Sch03] Thomas W. Schubert. "The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness." In: *Zeitschrift für Medienpsychologie* 15.2 (2003), pp. 69–71. DOI: 10.1026//1617-6383.15.2.69.

[Sch+12] Tobias Schwarz, Simon Butscher, Jens Mueller, and Harald Reiterer. "Content-Aware Navigation for Large Displays in Context of Traffic Control Rooms." In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI'12. Capri Island, Italy: ACM, 2012, pp. 249–252. DOI: 10.1145/2254556.2254601.

[SBA92] Abigail Sellen, Bill Buxton, and John Arnott. "Using Spatial Cues to Improve Videoconferencing." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. Monterey, California, USA: ACM, 1992, pp. 651–652. DOI: 10.1145/142750.143070.

[Ser+22] Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J. McGuffin, and Tobias Isenberg. "Collaborative Work in Augmented Reality: A Survey." In: *IEEE Transactions on Visualization and Computer Graphics* 28.6 (2022), pp. 2530–2549. DOI: 10.1109/TVCG.2020.3032761.

[SVG15] Adalberto L. Simeone, Eduardo Velloso, and Hans Gellersen. "Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences." In: *Proceedings of the Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015, pp. 3307–3316. DOI: 10.1145/2702123.2702389.

[SSW21] Richard Skarbez, Missie Smith, and Mary C. Whitton. "Revisiting Milgram and Kishino's Reality-Virtuality Continuum." In: *Frontiers in Virtual Reality* 2 (2021). DOI: 10.3389/frvir.2021.647997.

[SH16] Drew Skillman and Patrick Hackett. *Tilt Brush application, Google Inc.* 2016. URL: https://www.tiltbrush.com/.

[Sra+16] Misha Sra, Sergio Garrido-Jurado, Chris Schmandt, and Pattie Maes. "Procedurally generated virtual reality from 3D reconstructed physical space." In: *Proceedings of the Conference on Virtual Reality Software and Technology*. 2016, pp. 191–200. DOI: 10.1145/2993369.2993372.

[Ste+10] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. "Estimation of Detection Thresholds for Redirected Walking Techniques." In: *IEEE Transactions on Visualization and Computer Graphics* 16.1 (2010), pp. 17–27. DOI: 10.1109/TVCG.2009.62.

[SB02] Joachim Stempfle and Petra Badke-Schaub. "Thinking in design teams - an analysis of team communication." In: *Design Studies* 23.5 (2002), pp. 473–496. DOI: 10.1016/S0142-694X(02)00004-2.

[SCP95] Richard Stoakley, Matthew J. Conway, and Randy Pausch. "Virtual Reality on a WIM: Interactive Worlds in Miniature." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '95. Denver, Colorado, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 265–272. DOI: 10.1145/223904.223938.

[Str+99] Norbert A. Streitz, Jörg Geißler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Seitz, and Ralf Steinmetz. "I-LAND: An Interactive Landscape for Creativity and Innovation." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'99. Pittsburgh, Pennsylvania, USA: ACM, 1999, pp. 120–127. DOI: 10.1145/302979.303010.

[Sum+12] Evan A. Suma, Zachary Lipps, Samantha Finkelstein, David M. Krum, and Mark Bolas. "Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture." In: *IEEE Transactions on Visualization and Computer Graphics* 18.4 (2012), pp. 555–564. DOI: 10.1109/TVCG.2012.47.

[SWK16] Qi Sun, Li-Yi Wei, and Arie Kaufman. "Mapping virtual and physical reality." In: *ACM Transactions on Graphics (TOG)* 35.4 (2016), pp. 1–12. DOI: 10.1145/2897824.2925883.

[Sza+98] Zsolt Szalavári, Dieter Schmalstieg, Anton Fuhrmann, and Michael Gervautz. ""Studierstube": An environment for collaboration in augmented reality." In: *Virtual Reality* 3.1 (1998), pp. 37–48. DOI: 10.1007/BF01409796.

[TB15] Matthew Tait and Mark Billinghurst. "The Effect of View Independence in a Collaborative AR System." In: *Computer Supported Cooperative Work (CSCW)* 24.6 (2015), pp. 563–589. DOI: 10.1007/s10606-015-9231-8.

[TM90] John C. Tang and Scott L. Minneman. "VideoDraw: A Video Interface for Collaborative Drawing." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. Seattle, Washington, USA: ACM, 1990, pp. 313–320. DOI: 10.1145/97243.97302.

[TM91] John C. Tang and Scott L. Minneman. "VideoWhiteboard: Video Shadows to Support Remote Collaboration." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '91. New Orleans, Louisiana, USA: ACM, 1991, pp. 315–322. DOI: 10.1145/108844.108932.

[Teo+19] Theophilus Teo, Louise Lawrence, Gun A. Lee, Mark Billinghurst, and Matt Adcock. "Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, pp. 1–14. DOI: 10.1145/3290605.3300431.

[Tho+20] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. "TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '20. Virtual Event, USA: ACM, 2020, pp. 182–195. DOI: 10.1145/3379337.3415827.

[Uso+99] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks. "Walking > Walking-in-Place > Flying, in Virtual Environments." In: *Proceedings of the Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '99. ACM, 1999, pp. 359–364. DOI: 10.1145/311535.311589.

[Val+11] Dimitar Valkov, Frank Steinicke, Gerd Bruder, and Klaus Hinrichs. "2D Touching of 3D Stereoscopic Objects." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. Vancouver, BC, Canada: ACM, 2011, pp. 1353–1362. DOI: 10.1145/1978942.1979142.

[Vei+99] Elizabeth S. Veinott, Judith Olson, Gary M. Olson, and Xiaolan Fu. "Video Helps Remote Work: Speakers Who Need to Negotiate Common Ground Benefit from Seeing Each Other." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. Pittsburgh, Pennsylvania, USA: ACM, 1999, pp. 302–309. DOI: 10.1145/302979.303067.

[Ver99] Roel Vertegaal. "The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. Pittsburgh, Pennsylvania, USA: ACM, 1999, pp. 294–301. DOI: 10.1145/302979.303065.

[Ver+01] Roel Vertegaal, Robert Slagter, Gerrit Van der Veer, and Anton Nijholt. "Eye Gaze Patterns in Conversations: There is More to Conversational Agents than Meets the Eyes." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. Seattle, Washington, USA: ACM, 2001, pp. 301–308. DOI: 10.1145/365024.365119.

[VVV00] Roel Vertegaal, Gerrit Van der Veer, and Harro Vons. "Effects of Gaze on Multiparty Mediated Communication." In: *Proceedings of the Graphics Interface 2000 Conference*. GI '00. Montr'eal, Qu'ebec, Canada, 2000, pp. 95–102. DOI: 10.20380/GI2000.14.

[Wal+20] Shaun Wallace, Brendan Le, Luis A. Leiva, Aman Haq, Ari Kintisch, Gabrielle Bufrem, Linda Chang, and Jeff Huang. "Sketchy: Drawing Inspiration from the Crowd." In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW2 (2020). DOI: 10.1145/3415243.

[Wan+20a] Chiu-Hsuan Wang, Chia-En Tsai, Seraphina Yong, and Liwei Chan. "Slice of Light: Transparent and Integrative Transition Among Realities in a Multi-HMD-User Environment." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '20. Virtual Event, USA: ACM, 2020, pp. 805–817. DOI: 10.1145/3379337.3415868.

[Wan+20b] Chiu-Hsuan Wang, Seraphina Yong, Hsin-Yu Chen, Yuan-Syun Ye, and Liwei Chan. "HMD Light: Sharing In-VR Experience via Head-Mounted Projector for Asymmetric Interaction." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '20. Virtual Event, USA: ACM, 2020, pp. 472–486. DOI: 10.1145/3379337.3415847.

[Wei+11] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. "Realtime Performance-Based Facial Animation." In: *ACM Transactions on Graphics (siggraph'11)* 30.4 (2011), 77:1–10. DOI: 10.1145/2010324.1964972.

[WBF20] Tim Weissker, Pauline Bimberg, and Bernd Froehlich. "Getting There Together: Group Navigation in Distributed Virtual Environments." In: *IEEE Transactions on Visualization and Computer Graphics* 26.5 (2020), pp. 1860–1870. DOI: 10.1109/TVCG.2020.2973474.

[WF21] Tim Weissker and Bernd Froehlich. "Group Navigation for Guided Tours in Distributed Virtual Environments." In: *IEEE Transactions on Visualization and Computer Graphics* 27.5 (2021), pp. 2524–2534. DOI: 10.1109/TVCG.2021.3067756.

[WKF19] Tim Weissker, Alexander Kulik, and Bernd Froehlich. "Multi-Ray Jumping: Comprehensible Group Navigation for Collocated Users in Immersive Virtual Reality." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 136–144. DOI: 10.1109/VR.2019.8797807.

[Wei+18] Tim Weissker, André Kunert, Bernd Fröhlich, and Alexander Kulik. "Spatial Updating and Simulator Sickness During Steering and Jumping in Immersive Virtual Environments." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018, pp. 97–104. DOI: 10.1109/VR.2018.8446620.

[Wil+10] Malte Willert, Stephan Ohl, Anke Lehmann, and Oliver Staadt. "The Extended Window Metaphor for Large High-resolution Displays." In: *Proceedings of the Eurographics Conference on Virtual Environments & Second Joint Virtual Reality*. EGVE - JVRC'10. Stuttgart, Germany: Eurographics Association, 2010, pp. 69–76. DOI: 10.2312/EGVE/JVRC10/069-076.

[WG10] Nelson Wong and Carl Gutwin. "Where Are You Pointing?: The Accuracy of Deictic Pointing in CVEs." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'10*. Atlanta, Georgia, USA: ACM, 2010, pp. 1029–1038. DOI: 10.1145/1753326.1753480.

[Xia+18] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. "Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality." In: *Proceedings of the Symposium on User Interface Software and Technology*. UIST '18. Berlin, Germany: ACM, 2018, pp. 853–866. DOI: 10.1145/3242587.3242597.

[XZ15] Junjie Xue and Gang Zhao. "Interactive Rendering and Modification of Massive Aircraft CAD Models in Immersive Environment." In: *Computer-Aided Design and Applications* 12.4 (2015), pp. 393–402. DOI: 10.1080/16864360.2014.997635.

[Yam+02] Toshio Yamada, Daisuke Tsubouchi, Tetsuro Ogi, and Michitaka Hirose. "Desk-sized immersive workplace using force feedback grid interface." In: *Proceedings IEEE Virtual Reality (VR)*. 2002, pp. 135–142. DOI: 10.1109/VR.2002.996516.

[Yan+22] Longqi Yang et al. "The effects of remote work on collaboration among information workers." In: *Nature Human Behaviour* 6 (2022), pp. 43–54. DOI: 10.1038/s41562-021-01196-4.

[Yao+18] Nancy Yao, Jeff Brewer, Sarah D'Angelo, Mike Horn, and Darren Gergle. "Visualizing Gaze Information from Multiple Students to Support Remote Instruction." In: *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: ACM, 2018, pp. 1–6. DOI: 10.1145/3170427.3188453.

[Yoo+19] Boram Yoon, Hyung-il Kim, Gun A. Lee, Mark Billinghurst, and Woontack Woo. "The Effect of Avatar Appearance on Social Presence in an Augmented Reality Remote Collaboration." In: *Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 547–556. DOI: 10.1109/VR.2019.8797719.

[Yu+22] Kevin Yu, Ulrich Eck, Frieder Pankratz, Marc Lazarovici, Dirk Wilhelm, and Nassir Navab. "Duplicated Reality for Co-located Augmented Reality Collaboration." In: *IEEE Transactions on Visualization and Computer Graphics* 28.5 (2022), pp. 2190–2200. DOI: 10.1109/TVCG.2022.3150520.

[Yu+10] Lingyun Yu, Pjotr Svetachov, Petra Isenberg, Maarten H. Everts, and Tobias Isenberg. "FI3D: Direct-Touch Interaction for the Exploration of 3D Scientific Visualization Spaces." In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1613–1622. DOI: 10.1109/TVCG.2010.157.

[Zha+23] Lei Zhang, Ashutosh Agrawal, Steve Oney, and Anhong Guo. "VRGit: A Version Control System for Collaborative Content Creation in Virtual Reality." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: ACM, 2023. DOI: 10.1145/3544548.3581136.

[Zha+04] Li Zhang, Noah Snavely, Brian Curless, and Steven M. Seitz. "Spacetime Faces: High Resolution Capture for Modeling and Animation." In: *ACM Transactions on Graphics (siggraph'04)* 23.3 (2004), pp. 548–558. DOI: 10.1145/1015706.1015759.

[Zha+21] Yiran Zhang, Sy-Thanh Ho, Nicolas Ladévèze, Huyen Nguyen, Cédric Fleury, and Patrick Bourdot. "In Touch with Everyday Objects: Teleportation Techniques in Virtual Environments Supporting Tangibility." In: *Workshop on Everyday Virtual Reality (WEVR) at the IEEE Conference on Virtual Reality and 3D User Interfaces*. Virtual Event, 2021, pp. 278–283. DOI: 10.1109/VRW52623.2021.00057.

[Zha+19] Yiran Zhang, Nicolas Ladévèze, Cédric Fleury, and Patrick Bourdot. "Switch Techniques to Recover Spatial Consistency Between Virtual and Real World for Navigation with Teleportation." In: *Proceedings of the EuroVR International Conference*. Vol. 11883. Lecture Notes in Computer Science. Tallinn, Estonia: Springer, 2019, pp. 3–23. DOI: 10.1007/978-3-030-31908-3_1.

[Zha+20] Yiran Zhang, Nicolas Ladévèze, Huyen Nguyen, Cédric Fleury, and Patrick Bourdot. "Virtual Navigation Considering User Workspace: Automatic and Manual Positioning before Teleportation." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST'20. Virtual Event: ACM, 2020. DOI: 10.1145/3385956.3418949.

[Zha+22] Yiran Zhang, Huyen Nguyen, Nicolas Ladevèze, Cédric Fleury, and Patrick Bourdot. "Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE VR'22. 2022, pp. 674–682. DOI: 10.1109/VR51125.2022.00088.

CURRICULUM VITAE

Cédric FLEURY - https://cedricfleury.fr

Current Situation & Research Positions

2021 - present	**Assistant Professor**, IMT Atlantique / Lab-STICC, Brest (France) *Member of the **INUIT** research group*
2013 - 2021	**Assistant Professor**, Université Paris-Saclay / LISN, Orsay (France) *Member of the **ex)situ** Inria research group*
2012 - 2013	**Postdoctoral researcher**, University of North Carolina (UNC), Chapel Hill (USA) Visiting researcher at Nanyang Technological University (NTU), Singapour *Supervisors: Henry Fuchs, Tat Jen Cham*
2008 - 2012	**Ph.D. student**, IRISA / Inria Rennes (France), funded by the "ANR Collaviz" project *Supervisors: Bruno Arnaldi (thesis director), Thierry Duval, Valérie Gouranton*
2011 Sept. - Dec.	**Visiting Ph.D. student**, University College of London (UK) - *supervised by Anthony Steed*
2008 Feb. - June	**Master student**, IRISA / Inria Rennes (France) - *supervised by Thierry Duval*
2007 Feb. - June	**Master student**, IRISA / Inria Rennes (France) - *supervised by Thierry Duval*
2006 July - Sept.	**Intern**, INSA de Rennes (France) - *supervised by Pierre-Yves Glorennec*

Education

2012	**Ph.D. in Computer Science**, INSA de Rennes (France)
2008	**Master of Research degree** in *Interaction, Image Processing and AI*, INSA de Rennes (France)
2008	**Master of Science degree** in *Computer Science*, INSA de Rennes (France)
2004	**Bachelor of Science degree in Mathematics**, Université de Bretagne Sud, Lorient (France)
2002	**Baccalauréat in sciences** (high school diploma), Lycée St Charles, St Brieuc, (France)

Teaching

Courses at undergraduate level: intro. to computer science, object-oriented programming, intro. to HCI
Courses at graduate level: VR and AR, Advanced HCI, software engineering for HCI, CSCW

2021 - present	**Assistant Professor of Computer Science** at IMT Atlantique engineering school and in the human-computer interaction Master at Université Bretagne Occidentale
2021 - present	**Co-chair of the HCI Major** (Master's level), IMT Atlantique
2013 - 2021	**Assistant Professor of Computer Science** at Polytech Paris-Saclay engineering school and in the human-computer interaction Masters at Université Paris-Saclay
2013 - 2021	**Head of a technological platform** used for teaching mixte reality, Polytech Paris-Saclay
2019 - 2021	**Chair of the 5th year** of engineering school in computer science, Polytech Paris-Saclay
2017 - 2019	**Chair of the 3rd year** of apprenticeship in computer science, Polytech Paris-Saclay
2013 - 2018	**Head of internships** in the human-computer interaction Masters, Université Paris-Saclay
2006 - 2011	Teaching assistant, INSA de Rennes (France)
2008 - 2010	Teaching assistant, Université de Rennes 1 (France)

Research

My research interests are in the fields of human-computer interaction, computer-supported cooperative work, and mixte reality including virtual and augmented reality. I am interested in the collaboration and interaction of multiple users in large interactive spaces such as wall-sized displays, immersive virtual reality systems and augmented reality setups. I worked on various projects on video-conferencing, telepresence, collaborative virtual environments and 3D interaction.

Supervision: I supervised 14 master's level interns, 6 engineers and 6 Ph.D. students (2 Ph.D.s are in progress)

Ph.D. students (6)

2022 - present Aurélien Léchappé, supervised at 40% with C. Dumas and M. Chollet
"Modeling common ground knowledge for real-time analysis of collaboration in a virtual env."

2020 - present Thomas Rinnert, supervised at 30% with T. Duval et B. Thomas
"Perceiving distant collaborative activity with mixed reality"

2019 - 2023 Arthur Fages, supervised at 50% with T. Tsandilas - **thesis defended**
"Supporting collaborative 3D modeling through augmented-reality spaces"

2017 - 2021 Yiran Zhang, supervised at 50% with P. Bourdot - **thesis defended**
"Telepresence for remote and heterogeneous collaborative virtual environments"

2015 - 2019 Yujiro Okuya, supervised at 50% with P. Bourdot - **thesis defended**
"*CAD modification techniques for design reviews on heterogeneous interactive systems"*

2014 - 2017 Ignacio Avellino, supervised at 80% with M. Beaudouin-Lafon - **thesis defended**
"Supporting collaborative practices across wall-sized displays with video-mediated communication"

Master's level interns (14)

- Clément Jézéquel (M2), co-supervised with E. Peillard (2023). *Keywords: remote collab., AR*
- Michele Romani (M1), co-supervised with M. Beaudouin-Lafon (2019). *Keywords: telepresence, AR*
- Clément Sauvard (M2), co-supervised with T. Tsandilas (2019). *Keywords: 2D/3D interaction, AR*
- Cyril Crebouw (M2), co-supervised with M. Beaudouin-Lafon (2019). *Keywords: telepresence, WSD*
- Antonin Cheymol (M1), supervised at 100% (2019). *Keywords: 3D interaction, VR*
- Jiannan Li (visiting PhD), co-supervised with M. Beaudouin-Lafon (2018). *Keywords: telepresence, WSD*
- Kévin Ahson (M1), co-supervised with T. Tsandilas (2018). *Keywords: collaborative interaction, AR*
- Krishnan Chandran (M2), co-supervised with T. Tsandilas (2018). *Keywords: collaborative interaction, AR*
- Brennan Jones (M2), co-supervised with I. Avellino et M. Beaudouin-Lafon (2016). *Keywords: telepresence, WSD*
- Jean-Baptiste Louvet (M2), supervised at 100% (2015). *Keywords: 3D interaction, WSD*
- Ignacio Avellino (M2), co-supervised with M. Beaudouin-Lafon (2014). *Keywords: telepresence, WSD*
- Hugo Marchadour (M1), co-supervised with T. Duval et V. Gouranton (2011). *Keywords: visualization, VR*
- Florent Goetz (M1), co-supervised with T. Duval et V. Gouranton (2011). *Keywords: collaborative interaction, VR*
- Charles Perin (M1), co-supervised with T. Duval (2010). *Keywords: 3D interaction, VR*

Engineers (6)

- Léo Colombaro (intern), co-supervised with O. Gladin (2018)
- Gabriel Tézier (2014), Amani Kooli, Jonathan Thorpe, Rémi Hellequin (2014 - 2016) and Lawrence Fyfe (2016 - 2018), co-supervised with J. Vézien, O. Gladin, S. Huot and M. Beaudouin-Lafon in the context of DIGISCOPE

Research projects

2022 - present Industrial chair "Région Pays-de-la-Loire" - Ph.D. funding of Aurélien Léchappé

2021 - present Participation in the EquipEx+ "Continuum" (scientific and technical committees)

2014 - 2020 Co-chair of the **technical committee** (with J. Vézien) and budget monitoring of the EquipEx "DIGISCOPE" partially funded by the French National Research Agency (ANR) (22M€)

2015 - 2019 Co-investigator of "SensoMotorCVE" project (with P. Bourdot) - Ph.D. funding of Y. Okuya by Labex Digicosme - ANR (109K€)

2018 Participation in "VR-BatIM" project (main investigator: J. Vézien) - Univ. Paris-Saclay (20K€)
"Virtual reality for the interaction with building information model at Paris-Saclay"

2017 Participation in "ARSCPMD" project (main investigator: T. Tsandilas) - Univ. Paris-Saclay (8K€)
"An augmented-reality system for collaborative physical modeling and design"

2015 Main investigator of an "Attractivité" project - Université Paris-Saclay (7,3K€)
"Telepresence systems preserving eye contact between remote users located in large interactive spaces"

2012 - 2013	Participation in the *BeingThere Center,* a collaborative project between University of North Carolina (UNC) at Chapel Hill (USA), Nanyang Technological University (NTU) in Singapour and Swiss Federal Institute of Technology (ETH) in Zurich (Switzerland)
2009 - 2012	Ph.D. student for the ANR "Collaviz" project
2007 - 2010	Intern, then Ph.D. student for the RNTL / ANR "Part@ge" project

Organization and research evaluation

- Co-chair of the doctoral consortium at IEEE ISMAR 2022 conference
- Co-chair of the program committees for the demonstrations at IHM 2021 conference
- Co-chair of the program committees for the "work in progress" at IHM 2018 conference
- Member of the program committees of the following conferences: ACM VRST 2019, EuroVR (2017 and 2020), GI 2016, IEEE VR 2016, 3DCVE@IEEE VR (2015 - 2018), GRAPP (2014 - 2016)
- Reviewer for the following journals: TVCG, Frontiers, CG&A, JOCCH, JVRB
- Reviewer for the following conferences: ACM CHI, ACM UIST, ACM VRST, ACM SIGGRAPH ASIA, IEEE VR, IEEE ISMAR, 3DUI, 3DVIS@IEEE VIS, 3DCVE@IEEE VR, CGI, JVRC, IHM
- Reviewer for project proposals of the French National Research Agency (ANR) (2014, 2015, 2020, 2023)
- Member of the PhD. committee of Romain Terrier (IRISA - Inria Rennes) (2020)

Professional service

- Elected member of the **LRI lab council** (joint research lab between Univ. Paris-Saclay & CNRS) (2014 - 2020)
- Member of a **hiring committee for an assistant professor position** at Université Paris-Saclay (2019)

Other experiences

2002 - 2009	**High level athlete in sailing** - registered on the list of the French Ministry of Sports International level in Olympic sailing (49er) and offshore racing ("Tour de France" sailing races)
2002 - 2005	**Treasurer of the association 29er - 49er France:** this association federates French competitors sailing on 49er and 29er boats, and organizes international events.

PUBLICATIONS

BOOK CHAPTER

[Oku+21] Yujiro Okuya, Nicolas Ladévèze, Olivier Gladin, **Cédric Fleury**, and Patrick Bourdot. "Collaborative VR-CAD for Industrial Product Design: CAD Parameter Modification with 3D Interaction on Heterogeneous Immersive Platforms." In: *Manufacturing in the Era of 4th Industrial Revolution*. Ed. by Monica Bordegoni, Satyandra K. Gupta, and James Ritchie. Vol. 3. World Scientific, 2021. Chap. 2, pp. 17–47. DOI: 10.1142/9789811222863_0002.

JOURNALS

[FFT22b] Arthur Fages, **Cédric Fleury**, and Theophanis Tsandilas. "Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users." In: *Proceedings of the ACM on Human-Computer Interaction* 6. CSCW2 (2022). DOI: 10.1145/3555607.

[Oku+18a] Yujiro Okuya, Nicolas Ladévèze, **Cédric Fleury**, and Patrick Bourdot. "ShapeGuide: Shape-Based 3D Interaction for Parameter Modification of Native CAD Data." In: *Frontiers in Robotics and AI* 5 (2018). DOI: 10.3389/frobt.2018.00118.

[Duv+14] Thierry Duval, Huyen Nguyen, **Cédric Fleury**, Alain Chauffaut, Georges Dumont, and Valérie Gouranton. "Improving awareness for 3D virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues." In: *Journal on Multimodal User Interfaces* 8 (2014), pp. 187–197. DOI: 10.1007/s12193-013-0134-z.

INTERNATIONAL CONFERENCES

[Zha+22] Yiran Zhang, Huyen Nguyen, Nicolas Ladevèze, **Cédric Fleury**, and Patrick Bourdot. "Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs." In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE VR'22. 2022, pp. 674–682. DOI: 10.1109/VR51125.2022.00088.

[Oku+20] Yujiro Okuya, Olivier Gladin, Nicolas Ladévèze, **Cédric Fleury**, and Patrick Bourdot. "Investigating Collaborative Exploration of Design Alternatives on a Wall-Sized Display." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'20. Virtual Event: ACM, 2020, pp. 1–12. DOI: 10.1145/3313831.3376736.

[Zha+20] Yiran Zhang, Nicolas Ladévèze, Huyen Nguyen, **Cédric Fleury**, and Patrick Bourdot. "Virtual Navigation Considering User Workspace: Automatic and Manual Positioning before Teleportation." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST'20. Virtual Event: ACM, 2020. DOI: 10.1145/3385956.3418949.

[Le+19] Khanh-Duy Le, Ignacio Avellino, **Cédric Fleury**, Morten Fjeld, and Andreas M. Kunz. "GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration." In: *Proceedings of the IFIP TC13 International Conference On Human-Computer Interaction (INTERACT)*. Vol. 11747. Lecture Notes in Computer Science. Paphos, Cyprus: Springer, 2019, pp. 282–303. DOI: 10.1007/978-3-030-29384-0_18.

[Zha+19] Yiran Zhang, Nicolas Ladévèze, **Cédric Fleury**, and Patrick Bourdot. "Switch Techniques to Recover Spatial Consistency Between Virtual and Real World for Navigation with Teleportation." In: *Proceedings of the EuroVR International Conference*. Vol. 11883. Lecture Notes in Computer Science. Tallinn, Estonia: Springer, 2019, pp. 3–23. DOI: 10.1007/978-3-030-31908-3_1.

[Fyf+18] Lawrence Fyfe, Olivier Gladin, **Cédric Fleury**, and Michel Beaudouin-Lafon. "Combining Web Audio Streaming, Motion Capture, and Binaural Audio in a Telepresence System." In: *Proceedings of the International Web Audio conference*. WAC'18. Berlin, Germany, 2018. URL: https://hal.science/hal-01957843.

[Ave+17] Ignacio Avellino, **Cédric Fleury**, Wendy E. Mackay, and Michel Beaudouin-Lafon. "CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'17. Denver, Colorado, USA: ACM, 2017, pp. 6718–6729. DOI: 10.1145/3025453.3025604.

[LF16] Jean-Baptiste Louvet and **Cédric Fleury**. "Combining Bimanual Interaction and Teleportation for 3D Manipulation on Multi-Touch Wall-Sized Displays." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST'16. Munich, Germany: ACM, 2016, pp. 283–292. DOI: 10.1145/2993369.2993390.

[AFB15] Ignacio Avellino, **Cédric Fleury**, and Michel Beaudouin-Lafon. "Accuracy of Deictic Gestures to Support Telepresence on Wall-Sized Displays." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'15. Seoul, Republic of Korea: ACM, 2015, pp. 2393–2396. DOI: 10.1145/2702123.2702448.

[Fle+14] **Cédric Fleury**, Tiberiu Popa, Tat Jen Cham, and Henry Fuchs. "Merging Live and pre-Captured Data to support Full 3D Head Reconstruction for Telepresence." In: *Proceedings of the Conference of the European Association for Computer Graphics*. EG'14. Strasbourg, France, 2014, pp. 9–12. DOI: 10.2312/egsh.20141002.

[NDF13b] Thi Thuong Huyen Nguyen, Thierry Duval, and **Cédric Fleury**. "Guiding Techniques for Collaborative Exploration in Multi-Scale Shared Virtual Environments." In: *Proceedings of the International Conference*

on Computer Graphics Theory and Applications. GRAPP'13. Barcelone, Spain, 2013, pp. 327–336. URL: https://hal.science/hal-00755313.

[Duv+12] Thierry Duval, Huyen Nguyen, **Cédric Fleury**, Alain Chauffaut, Georges Dumont, and Valérie Gouranton. "Embedding the features of the users' physical environments to improve the feeling of presence in collaborative Virtual Environments." In: *Proceedings of the IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. 2012, pp. 243–248. DOI: 10.1109/CogInfoCom.2012.6421987.

[Fle+12] **Cédric Fleury**, Thierry Duval, Valérie Gouranton, and Anthony Steed. "Evaluation of Remote Collaborative Manipulation for Scientific Data Analysis." In: *Proceedings of the Symposium on Virtual Reality Software and Technology*. VRST'12. Toronto, Ontario, Canada: ACM, 2012, pp. 129–136. DOI: 10.1145/2407336.2407361.

[DF11b] Thierry Duval and **Cédric Fleury**. "PAC-C3D: A New Software Architectural Model for Designing 3D Collaborative Virtual Environments." In: *Proceedings of the International Conference on Artificial Reality and Telexistence*. ICAT'11. Osaka, Japan, 2011, pp. 53–60. URL: https://hal.science/inria-00636143.

[Fle+10a] **Cédric Fleury**, Alain Chauffaut, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. "A Generic Model for Embedding Users' Physical Workspaces into Multi-Scale Collaborative Virtual Environments." In: *Proceedings of the International Conference on Artificial Reality and Telexistence*. ICAT'10. Adelaide, Australia, 2010. URL: https://hal.science/inria-00534096.

[Fle+10b] **Cédric Fleury**, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. "A New Adaptive Data Distribution Model for Consistency Maintenance in Collaborative Virtual Environments." In: *Proceedings of the Joint Virtual Reality Conference of EuroVR - EGVE - VEC*. EGVE - JVRC'10. Stuttgart, Germany: Eurographics Association, 2010, pp. 29–36. DOI: 10.2312/EGVE/JVRC10/029-036.

[DF09] Thierry Duval and **Cédric Fleury**. "An Asymmetric 2D Pointer/3D Ray for 3D Interaction within Collaborative Virtual Environments." In: *Proceedings of the International Conference on 3D Web Technology*. Web3D'09. Darmstadt, Germany: ACM, 2009, pp. 33–41. DOI: 10.1145/1559764.1559769.

WORKSHOP AND LATE-BREAKING WORK PAPERS

[Léc+23] Aurélien Léchappé, Aurélien Milliat, **Cédric Fleury**, Mathieu Chollet, and Cédric Dumas. "Characterization of Collaboration in a Virtual Environment with Gaze and Speech Signals." In: *Companion Publication of the International Conference on Multimodal Interaction*. ICMI '23 Companion. Paris, France: ACM, 2023, pp. 21–25. DOI: 10.1145/3610661.3617149.

[Zha+21] Yiran Zhang, Sy-Thanh Ho, Nicolas Ladévèze, Huyen Nguyen, **Cédric Fleury**, and Patrick Bourdot. "In Touch with Everyday Objects: Teleportation Techniques in Virtual Environments Supporting Tangibility." In: *Workshop on Everyday Virtual Reality (WEVR) at the IEEE Conference on Virtual Reality and 3D User Interfaces*. Virtual Event, 2021, pp. 278–283. DOI: 10.1109/VRW52623.2021.00057.

[Oku+18b] Yujiro Okuya, Nicolas Ladévèze, Olivier Gladin, **Cédric Fleury**, and Patrick Bourdot. "Distributed Architecture for Remote Collaborative Modification of Parametric CAD Data." In: *Workshop on Collaborative Virtual Environments (3DCVE) at the IEEE Conference on Virtual Reality*. Reutlingen, Germany, 2018, pp. 1–4. DOI: 10.1109/3DCVE.2018.8637112.

[Fle+15a] **Cédric Fleury**, Ignacio Avellino, Michel Beaudouin-Lafon, and Wendy E. Mackay. "Telepresence Systems for Large Interactive Spaces." In: *Workshop on Everyday Telepresence: Emerging Practices and Future Research Directions at the ACM conference on Human Factors in Computing Systems*. Seoul, South Korea, 2015. URL: https://hal.science/hal-01242304.

[Fle+15b] **Cédric Fleury**, Nicolas Férey, Jean-Marc Vézien, and Patrick Bourdot. "Remote collaboration across heterogeneous large interactive spaces." In: *Workshop on Collaborative Virtual Environments (3DCVE) at the IEEE Conference on Virtual Reality*. 2015, pp. 9–10. DOI: 10.1109/3DCVE.2015.7153591.

[Fle+10c] **Cédric Fleury**, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. "Architectures and Mechanisms to Maintain efficiently Consistency in Collaborative Virtual Environments." In: *Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS) at the IEEE Conference on Virtual Reality*. Waltham, United States, 2010. URL: https://hal.science/inria-00534082.

POSTERS AND DEMONSTRATIONS

[FFT22a] Arthur Fages, **Cédric Fleury**, and Theophanis Tsandilas. "ARgus : système multi-vues pour collaborer à distance avec un utilisateur en réalité augmentée." In: *Proceedings of the conférence francophone sur l'Interaction Homme-Machine*. Demonstration. Namur, Belgium, 2022. URL: https://hal.science/hal-03762816.

[Oku+17] Yujiro Okuya, Nicolas Ladévèze, **Cédric Fleury**, and Patrick Bourdot. "VR-CAD Framework for Parametric Data Modification with a 3D Shape-based Interaction." In: *Proceedings of the EuroVR International Conference (poster)*. Laval, France, 2017. URL: https://hal.science/hal-04327386.

[NDF13a] Thi Thuong Huyen Nguyen, Thierry Duval, and **Cédric Fleury**. "Demonstration of Guiding Techniques for Collaborative Exploration in Multi-Scale Shared Virtual Environments." In: *Proceedings of the conférence francophone sur l'Interaction Homme-Machine (demo)*. Demonstration. Bordeaux, France, 2013. URL: https://inria.hal.science/hal-00879475.

[NFD12] Thi Thuong Huyen Nguyen, **Cédric Fleury**, and Thierry Duval. "Collaborative exploration in a multi-scale shared virtual environment." In: *Proceedings of the Symposium on 3D User Interfaces*. Demonstration. 2012, pp. 181–182. DOI: 10.1109/3DUI.2012.6184221.

[DF11a] Thierry Duval and **Cédric Fleury**. "Collaboration between Networked Heterogeneous 3D Viewers through a PAC-C3D Modeling of the Shared Virtual Environment." In: *Proceedings of the International Conference on Artificial Reality and Telexistence*. Demonstration. Osaka, Japan, 2011. URL: https://hal.science/inria-00638638v1.

[Dup+10] Florent Dupont, Thierry Duval, **Cédric Fleury**, Julien Forest, Valérie Gouranton, Pierre Lando, Thibaut Laurent, Guillaume Lavoué, and Alban Schmutz. "Collaborative Scientific Visualization: The COLLAVIZ Framework." In: *Proceedings of the Joint Virtual Reality Conference of EuroVR - EGVE - VEC*. Demonstration. Fellbach, Germany, 2010. URL: https://hal.science/inria-00534105.

[Duv+08] Thierry Duval, **Cédric Fleury**, Bernard Nouailhas, and Laurent Aguerreche. "Collaborative Exploration of 3D Scientific Data." In: *Proceedings of the Symposium on Virtual Reality Software and Technology (VRST'08)*. Demonstration. Bordeaux, France: ACM, 2008, pp. 303–304. DOI: 10.1145/1450579.1450664.

NATIONAL CONFERENCES

[FFT23] Arthur Fages, **Cédric Fleury**, and Theophanis Tsandilas. "Conception collaborative au travers de versions parallèles en réalité augmentée." In: *Proceedings of the Conference on l'Interaction Humain-Machine*. IHM '23. TROYES, France: ACM, 2023. DOI: 10.1145/3583961.3583978.

[FDC08] **Cédric Fleury**, Thierry Duval, and Alain Chauffaut. "Cabine Virtuelle d'Immersion (CVI) : naviguer et interagir en immersion dans les univers virtuels collaboratifs multi-échelles." In: *Journées de l'Association Française de Réalité Virtuelle (AFRV)*. Bordeaux, France, 2008. URL: https://hal.science/inria-00433843.

SELECTED PUBLICATIONS

p. 117 J.-B. Louvet, C. Fleury (2016). Combining Bimanual Interaction and Teleportation for 3D Manipulation on Multi-Touch Wall-sized Displays. Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST), 8 pages.

p. 127 Y. Okuya, O. Gladin, N. Ladévèze, C. Fleury, P. Bourdot (2020). Investigating Collaborative Exploration of Design Alternatives on a Wall-Sized Display. Proceedings of the ACM conference on Human Factors in Computing Systems (CHI), 12 pages.

p. 139 Y. Zhang, N. Ladévèze, H. Nguyen, C. Fleury, P. Bourdot (2020). Virtual Navigation considering User Workspace: Automatic and Manual Positioning before Teleportation. Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST), 10 pages.

p. 149 Y. Zhang, T.T.H. Nguyen, N. Ladévèze, C. Fleury, P. Bourdot (2022). Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs, Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR), 9 pages.

p. 158 I. Avellino, C. Fleury, M. Beaudouin-Lafon (2015). Accuracy of deictic gestures to support telepresence on wall-sized displays. Proceedings of the ACM conference on Human Factors in Computing Systems (CHI), 4 pages.

p. 162 I. Avellino, C. Fleury, W. Mackay, M. Beaudouin-Lafon (2017). CamRay: Camera Arrays Support Remote Collaboration on Wall- Sized Displays. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 12 pages.

p. 174 K.-D. Le, I. Avellino, C. Fleury, M. Fjeld, A. Kunz (2019). GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration. Proc. of the IFIP TC13 Conference on Human-Computer Interaction (INTERACT), 21 pages.

p. 195 A. Fages, C. Fleury, T. Tsandilas (2022). Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users, Proc. of the ACM Conference on Computer-Supported Cooperative Work (CSCW), 27 pages.

Combining Bimanual Interaction and Teleportation for 3D Manipulation on Multi-Touch Wall-sized Displays

Jean-Baptiste Louvet [1,2,3] * Cédric Fleury [2,1] †

[1] Inria, Université Paris-Saclay, 91405, Orsay, France
[2] LRI, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405, Orsay, France
[3] Normandie Univ, INSA Rouen, UNIHAVRE, UNIROUEN, LITIS, 76000 Rouen, France

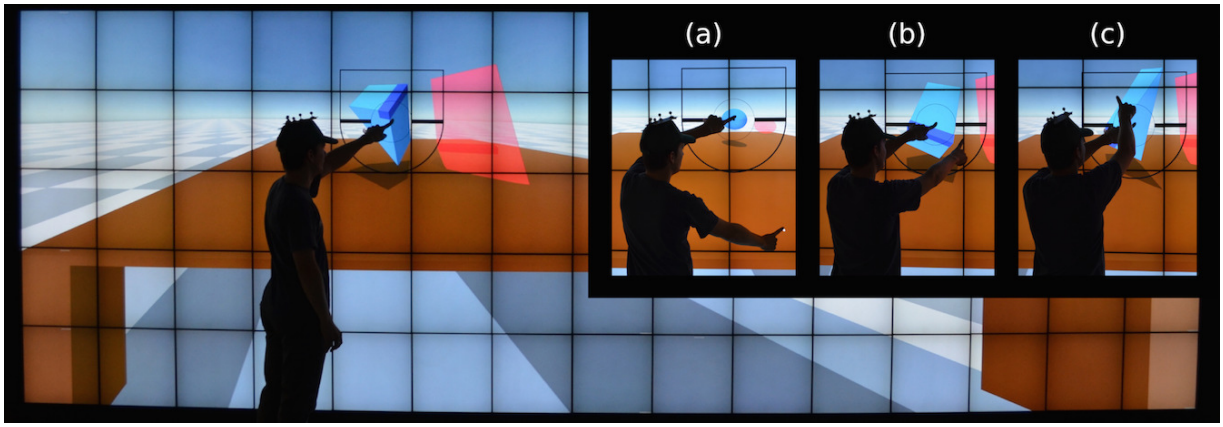

Figure 1: *6 degrees of freedom manipulation of a 3D object on a multi-touch wall-sized display combining bimanual interaction and teleportation. The user is performing a* xy *translation (main picture),* z *translation (a), roll rotation (b), and pitch & yaw rotation (c).*

Abstract

While multi-touch devices are well established in our everyday life, they are currently becoming larger and larger. Large screens such as wall-sized displays are now equipped with multi-touch capabilities. Multi-touch wall-sized displays will become widespread in a near future in various places such as public places or meeting rooms. These new devices are an interesting opportunity to interact with 3D virtual environments: the large display surface offers a good immersion, while the multi-touch capabilities could make interaction with 3D content accessible to the general public.

In this paper, we aim to explore touch-based 3D interaction in the situation where users are immersed in a 3D virtual environment and move in front of a vertical wall-sized display. We design In(SITE), a bimanual touch-based technique combined with object teleportation features which enables users to interact on a large wall-sized display. This technique is compared with a standard 3D interaction technique for performing 6 degrees of freedom manipulation tasks on a wall-sized display. The results of two controlled experiments show that participants can reach the same level of performance for completion time and a better precision for fine adjustments of object position with the In(SITE) technique. They also suggest that combining object teleportation with both techniques improves translations in terms of ease of use, fatigue, and user preference.

Keywords: multi-touch interaction, 3D manipulation, wall-sized display, virtual reality

Concepts: •**Human-centered computing** → **Virtual reality; Touch screens;** *Usability testing;*

1 Introduction

Even if touch has been reserved for small devices such as smartphones and tablets in its early years, more and more large screens, such as computer screens, televisions, whiteboards, and wall-sized displays, are now equipped with multi-touch capabilities. Large multi-touch screens will become widespread in a near future in various locations such as public places, meeting rooms, classrooms. For example, we can already see such large multi-touch displays in shopping malls or airports to enable visitors to browse information. At the same time, Microsoft is currently launching its new Surface Hub[1] which is designed for meeting rooms.

Large screen displays are powerful tools to visualize the increasing amount of data from science, industry, business, and society. They are also appropriate to support collaboration among small groups of users. However, these benefits should not be limited to 2D content. Large multi-touch devices are a relevant solution to interact with 3D virtual environments. While the large display surface increases

*e-mail: jeanbaptiste.louvet@insa-rouen.fr
†e-mail: cedric.fleury@lri.fr

VRST '16, November 02-04, 2016, Garching bei München, Germany
ISBN: 978-1-4503-4491-3/16/11...$15.00
DOI: http://dx.doi.org/10.1145/2993369.2993390

[1] https://www.microsoft.com/microsoft-surface-hub/

immersion in virtual environments, multi-touch input is a simple and efficient way to perform 3D interaction. Touch-based interaction can be easy to use and to learn for none-expert users since most of the people are familiar with touch devices. It also does not require additional external equipment since everything is embedded in the display device, which makes it particularly suitable for public areas, classrooms, and meeting rooms.

Most of the current touch-based 3D interaction techniques are designed either for small screens or consider that users stay static in front of the screen. Consequently, these techniques are not well adapted to the situation where users are immersed in a 3D virtual environment and freely move in front of a large vertical screen. In particular, finger combinations could be hard to achieve if users have to perform actions at the top or the bottom of the screen. In this paper, we propose In(SITE), a bimanual interaction technique designed to perform 6-DOF manipulation of 3D virtual objects using multi-touch input on a wall-sized display. This technique takes advantage of the large interaction space available on the display, in addition to the fact that users can easily move along the screen and perform actions with their two hands. On large wall-sized displays, it can be uncomfortable and inefficient to have to drag objects all along the screen when performing large translations. Consequently, we also propose to combine the In(SITE) technique with object teleportation features in order to overcome this drawback.

Two controlled experiments were performed to assess the usability of the In(SITE) technique in comparison with a standard 3D interaction technique which is used as a baseline for interaction in virtual reality. The first experiment evaluates the two techniques for 3D manipulation tasks which require only translations, while the second experiment focuses mainly on rotations. The first experiment also compares both techniques with and without teleportation features to determine the benefits of teleportation for each technique.

After reviewing the related work (section 2), we present the In(SITE) technique (section 3). Section 4 describes the usability testing of the In(SITE) technique through two controlled experiments. Then, we discuss the implications for touch-based 3D interaction on wall-sized displays (section 5). Finally, we conclude and propose a set of future research directions for 3D interaction on multi-touch wall-sized displays (section 6).

2 Related Work

The increasing accessibility to multi-touch technology has drawn the attention of researchers on the capabilities of touch-based interaction for 3D manipulation. On a state of the art about 3D interaction, Jankowski and Hachet [2015] point out that "*3D transformation widgets need to be reinvented to adapt to the tactile paradigm*", as users of touchscreens do not have access to an unobstructed view of the screen, an accurate pointing, and a direct access to buttons. The emergence of large wall-sized displays with multi-touch capabilities also introduces new constraints which should be taken into account when designing 3D touch-based interaction techniques.

2.1 3D Touch-based Interaction

Multiple approaches using touch input for 3D interaction have been explored. Cohé and Hachet [2012] conducted an experiment to understand how non-technical users *a-priori* interact with multi-touch screens to manipulate 3D objects. The experiment was conducted by showing participants a video of a cube transformation (rotation, scaling, translation) and then asking them to draw a gesture that best matches this transformation on a static image displayed on a touchscreen. The results of the study made them define a taxonomy based on the analysis of the users' gestures. It contains different strategies applied by participants to perform the actions for rotation, scaling, and translation. This study gave us some guidelines on how to design the interaction technique presented in this paper.

Some techniques try to mimic direct manipulation of an object in 3D such as the *screen-space* technique presented by Reisman et al. [2009]. This technique was designed to ensure that the screen-space projection of the object-space point "touched" by users always remains under their fingertip. It uses a constraint solver to minimize a function that measures the error between the points in the object local-space and the points in screen-space and update the object position and rotation according to the result. Even if these manipulation techniques seem more intuitive, the users can have troubles to perform specific manipulation (especially for rotation) because it is not always easy to predict how the manipulated object will react and to determine what is the correct action to perform. In a similar way, *sticky fingers* presented by Hancock et al. [2009] is a force-based interaction technique designed to maintain the feeling of physical interaction with a virtual object. It is a transposition of the 2D multi-touch interaction paradigm (rotate, spin, scale) to 3D interaction. Jackson et al. [2012] proposes to extend touch-based interface techniques by using hand gestures above the surface.

Some other techniques use multiple simultaneous touch inputs from several fingers or/and from both hands to control each one of the 6 DOF. In particular, Hancock et al. [2007] have tested different 3D manipulation techniques in "shallow-depth" (i.e., in limited depth) involving 1, 2, or 3 touch points, making possible to control 5 or 6 DOF. The user study carried out shows that user preference and performance are higher with techniques involving multiple touch points which enable users to separate DOF. To facilitate the control of the depth, the *multi-touch viewport* presented by Martinet et al. [2010a] divides the screen in four viewports, each one corresponding to a different viewpoint on the 3D scene. Interacting in the first viewport allows a 2 DOF translation of the object, while interacting at the same time in a second viewport makes it possible to control the third DOF of translation of the object. Martinet et al. [2010a] also proposed the *Z-technique* in order to only use a single view of the scene for 3D positioning. When a first finger touches the screen, a raycast is done from the camera center and passing by the touch position and the first object intersected can then be translated in the plane parallel to the camera plane passing through the object center. The depth of the manipulated object is controlled in an indirect way using a second finger. Its up and down movements on the screen are mapped to forward and backward movements of the object relatively to the user position. A docking task experiment did not show significant results on performance but the *Z-technique* was preferred by a majority of participants in comparison to the *multi-touch viewport* technique. Dividing the screen into four viewports does not seem a solution suitable for large wall-sized displays. In addition, Martinet et al. [2010b] combined the *Z-technique* to the *screen-space* technique [Reisman et al. 2009] into a technique called *DS3* and compared it to the *sticky fingers* [Hancock et al. 2009] and *screen-space* technique. The results of this study show that separating the control of translation and rotation, as it is done with DS3, makes the interaction significantly faster.

To improve the understanding of 3D manipulation and the separation of the 6 DOF, Cohé et al. [2011] introduced *tBox*, a 3D transformation widget designed for touchscreens. This widget is inspired by standard box-shaped 3D transformation widgets operated from mouse and keyboard in desktop 3D applications. The widget is a wireframe box that appears on top of the manipulated object. The rotation is carried out with one finger, rotating the widget (with the object) around one of its primary axes, depending of the direction of the gesture. The translation is carried out with one finger on an edge of the widget, allowing to translate the object in the direction of the selected edge. The scaling is carried out on the direction of the primary axes of the widget, with two fingers selecting two opposed edges of a face and and moving them away from each other.

2.2 Touch-based Interaction on Large Displays

Several previous works started to explore touch-based interaction techniques for 3D manipulation in the context of large vertical displays. For example, Yu et al. [2010] proposed *FI3D*, a touch-based interaction technique to navigate in a 3D scene on a large vertical screen. Traditional pan and zoom can be performed with 2-finger interaction with the display, but manipulation of other DOF can be achieved either by starting the drag gesture on the frame of the screen or just by touching the frame of the screen with the second finger. However, this technique is specific to the exploration of a single object in the 3D scene which does not require any object selection. Touching the frame of the screen is also not suitable with wall-sized displays because they are too large. In most recent work, Lopez et al. [2016] presented a variation of the previous technique in which the *FI3D* technique is used on a separate tablet to explore data displayed a large stereoscopic display. While this variation enables the user to control data on a larger display and deals with stereoscopic screen, the user has to frequently switch between the tablet view and the stereoscopic view. It also requires that the user hold an additional device, which does not seem appropriate when large displays are used in public places. In a similar way, Coffey et al. [2012] manipulated 3D medical data displayed on a large stereoscopic screen through 2D widget on a tabletop. Consequently, this technique implies indirect manipulation of the data displayed on the stereoscopic screen and also requires an additional device. Finally, Gilliot et al. [2014] presented a touch-based interaction technique called *WallPad* to perform direct and indirect manipulation on large wall-sized displays. However, this technique mainly focuses on touch-based interaction with a 2D graphical user interface.

2.3 Touch-based Interaction on Stereoscopic Displays

Some other related work focus on 3D touch-based interaction in the context of stereoscopic vision. For example, Benko and Feiner [2007] introduced the *Balloon Selection* which makes the selection of 3D objects possible in augmented reality settings. This multi-touch interaction technique consists in controlling the 2D position in the screen plane with a finger and to adjust the depth with the finger of the second hand using the metaphor of the string of an helium balloon which can be pulled down or released according to the second finger position. This technique only deals with translations and not with rotations. To improve the technique and to perform rotations, Strothoff et al. [2011] proposed a variation of the *Balloon Selection* called the *Triangle Cursor* in which two fingers of the user (same hand) define the base of an isosceles triangle which is perpendicular to the screen. The top of the triangle is the cursor for selection and manipulation. The user can control the height and the orientation of the triangle by moving these two fingers. The user can performed the additional rotation with his second hand. However, only objects which "come out" of the screen can be selected with both of these techniques. Consequently, they cannot be useful in the general context of immersive systems in which the 3D scene stands behind and in front of the display area.

On a more theoretical point of view, Valkov et al. [2011] studied touch-based interaction with a 2D screen to interact with 3D stereoscopic objects. They showed that users tend to touch between the projections for the two eyes with an offset towards the projection for the dominant eye. They also provided guidelines on how to design touch-based interaction for systems with 3D stereoscopic vision.

2.4 Synthesis

To sum up, many different touch-based interaction techniques have been proposed to perform 6 DOF manipulation of 3D objects. Several exhaustive studies have been achieved to better understand 3D touch-based interaction and provide useful guidelines to design a 3D touch-based interaction technique. In particular, Hancock et al. [2007] provide as guidelines to separate DOF and enable users to simultaneously control them. Even if some previous works have explored large and/or stereoscopic displays, most of the proposed techniques consider systems in which users stay static in front of the display and none of them deal with large wall-sized display in front of which users can freely move. In particular, touch input techniques which combine several fingers from the same hand are not convenient when users have to perform actions at the top or bottom of a large wall-sized display. Long drags all along the screen must also be avoided. In addition, techniques suitable with large displays often require control devices such a tablet or a tabletop. In the context of wall-sized displays in public area, we want to find solutions which avoid these additional devices and we would rather prefer solutions which only use direct interaction on the display.

3 In(SITE) Technique

We designed a new interface for interacting with a 3D virtual environment using multi-touch input: In(SITE), Interface for Spatial Interaction in Tactile Environments. We decided to focus on selection, translation and rotation of objects in virtual environments, so scaling is not present in this first version of the interface.

3.1 Manipulation Mode

The In(SITE) technique provides a widget, presented in Figure 2, which separates manipulation of the 6-DOF for 3D interaction. When the user touches the screen to select an object, a raycast is performed starting from his head and passing by his finger tip, which makes it possible to select the object under his finger tip according to his point of view on the 3D scene. If it reaches an object of the virtual environment, the widget appears on the screen under the user's finger.

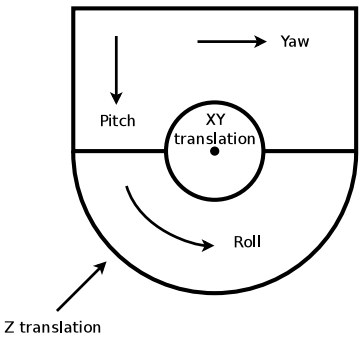

Figure 2: *In(SITE) manipulation widget.*

This widget allows the user to fully control the 6-DOF of the selected object using multi-touch input. With his primary finger (i.e., the one used for the selection), positioned at the center of the central circle of the widget, the user controls in a direct way the two x and y translations of the object in a plane parallel to the screen (see Figure 1(*main picture*)). This method allows co-planar dragging and is called *ObjectCorrection* by Möllers et al. [2012] in the literature. It implies that if the user moves his head, the object will also move to remain under his finger. The central circle of widget is a dead zone to avoid false-positive detections of secondary fingers while dragging, especially when the touch-detection system is not very precise (as it is the case on our wall-sized display).

The z translation is controlled in an indirect way by interacting with the area outside the widget with a secondary finger (see Figure 1(a)). This interaction design has been strongly influenced by the *Z-technique* [Martinet et al. 2010a]. The main difference is that the gesture of the secondary finger does not have to be vertical but towards the primary finger: when the user's secondary finger gets

closer to his primary finger mimicking a large pinch gesture, the object gets closer to him, and when his secondary finger gets away from his primary finger mimicking a large unpinch gesture, the object gets away from him. The direction of the z translation is given by the ray starting from the head of the user and passing through his primary finger. The mapping between the finger displacement and the object displacement is linear.

The upper and lower areas of the widget are used to control the rotation of the object. The lower area allows users to control the roll of the object by doing curve gestures which follow the object rotation (see Figure 1(b)). These curve gestures are suggested to the user by the round shape of the area. The upper area allows users to manipulate the yaw and pitch of the object by doing respectively horizontal and vertical movements (see Figure 1(c)). In the upper area, the yaw and pitch rotations can be combined by doing diagonal movements. The square shape of this area aims to highlight these 2D interaction possibilities. For both rotations, the rotation axis are defined by the gravity center of the manipulated object.

	Translation			Rotation		
Mode	Tx	Ty	Tz	Rx	Ry	Rz
1d(center)	○——○					
1d(center) + 1i(out)	○——○——Ⓘ					
2d(center + up)	○——○—			○——○		
2d(center + down)	○——○—					○

Table 1: *Description of the In(SITE) technique using the taxonomy described by Martinet et al. [2010b].*

Table 1 presents the different DOF manipulation modes offered by In(SITE) using the taxonomy introduced by Martinet et al. [2010b]. The primary finger is used to control Tx and Ty in a direct way, as the projection on the screen of the object is directly mapped to the position of the finger. The secondary finger can control Tz, Rx and Ry, or Rz. Tz is controlled in an indirect way using the external area of the widget, Rx and Ry are controlled in a direct way using the upper area of the widget, and Rz is controlled in a direct way using the lower area of the widget.

3.2 Teleportation

As In(SITE) aims to be used in large displays, we combine teleportation with the In(SITE) technique to improve its effectiveness for translations. Indeed, teleportation avoids that the user drags objects all along the screen when he needs to perform large translations.

To perform teleportation, the first step consists in selecting the object that the user wants to teleport. If he does a short touch (less than 1s) on the object, it is selected. As a feedback of the selection, the object color is changed. If he does a long touch (more than 1s), the interface enters directly in the manipulation mode of the In(SITE) technique, without selection. If the user selects an object by mistake, he can deselect it just by touching it again. In this case, the color feedback is removed and the user can select another object.

Once the user has selected an object, he can teleport it anywhere in the environment by doing a short touch on the destination location (either on the floor or on another object). This instantly moves the object above the destination and makes it fall downwards with an animation. The animation stops when the teleported object collides with another object of the environment. This animation is particularly useful for physically simulated virtual environments because the object can be piled up on other objects. Once the animation is over, the color feedback is still present for a short period of time. During this period, the user can catch the object again by touching it. If the object is caught, the interface enters in manipulation mode.

It is also possible to teleport the object without animation by doing a long touch on the destination. In this case the object is directly teleported at the location designated by the intersection between the environment and the ray starting from the user's head and passing by the finger used for the touch. Once the object has been teleported under the user's finger, the interface enters in manipulation mode. The teleportation is especially useful when the user wants to put the object under some other objects.

4 Usability Testing

We conducted two experiments to assess the usability of the In(SITE) technique combined with object teleportation in the context of an immersive virtual environment based on a large wall-sized display. The goal was to determine if the In(SITE) technique could reach the same level of performance than a standard 3D interaction technique for 6-DOF manipulation tasks of 3D objects in terms of completion time of the task and precision. We also considered precision, ease of use, frustration and fatigue of users as important evaluation criteria. In particular, we wanted to assess the four following hypotheses:

H1 Users can reach the same level of performance with In(SITE) in comparison to standard 3D interaction technique for tasks which require only translations.

H2 Users still can reach the same level of performance with In(SITE) in comparison to standard 3D interaction technique if the task requires also rotations.

H3 In(SITE) have additional advantages, in particular in terms of precision, ease of use, frustration, and fatigue.

H4 The teleportation can be beneficial to both techniques.

For the standard 3D interaction technique, we decided to use the virtual ray technique (also known as the ray-casting technique [Jacoby et al. 1994; Mine 1995]) as a baseline for 3D interaction in virtual reality (see Figure 3). Some other techniques have been developed to improve the virtual ray technique such as HOMER [Bowman and Hodges 1997] or Voodoo Dolls [Pierce et al. 1999], but they are still hard to manage for non-expert users and require additional tracking of the users' hand. Consequently, we have chosen to stick to the more standard version of the virtual ray technique since this version is the most commonly used as a mainstream interaction technique in various virtual reality systems such as CAVE. However, a virtual ray is not very efficient for pitch and yaw rotations because the manipulated object is attached to the ray and cannot easily rotate in that case. After pilot experiments, we decided to add a feature to the virtual ray technique for rotation along a vertical axis in order to be as fair as possible in our comparison between the two techniques. Users can thus easily perform yaw rotations and even pitch rotations by turning objects of 90° on the roll axis before using this additional feature (see section 4.1.1 for implementation details).

The two experiments followed the same experimental method, so the common parts are described in the following subsection. The first experiment focused on tasks which require only translations, while the second one concentrated on tasks which require rotations.

4.1 Experimental Method

4.1.1 Apparatus

The experiments were performed on a $5.90m \times 1.96m$ wall-sized display composed of 75 thin-bezel screens for a total resolution of 14400×4800 pixels. Applications ran on a server that distributes the environment to the 10 machines running the wall-sized display. Each machine has one Intel Xeon CPU at 3.7 GHz and a Nvidia Quadro K5000 graphic card. The virtual environment was designed

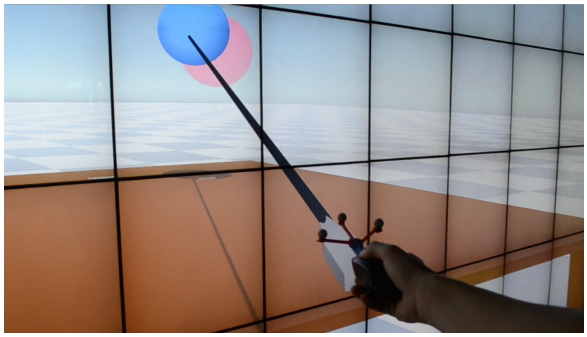

Figure 3: *For the virtual ray technique, the ray is co-located with the wireless mouse held by the user.*

using the game engine *Unity*[2].

While this wall-sized display offers a very large surface with multi-touch capabilities, it does not support stereoscopy. Even if the ideal case would have been to have stereoscopic vision to support immersion in the virtual environment, the device still provide a interesting level of immersion because it supports other cues of immersion: a wide field of view, a complete visual immersion when users are close to the screens and a good resolution. The user's head was also tracked in front of the screen by a *Vicon*[3] system. Displayed images were deformed according to the user head position to match the viewing frustum defined by the user's head and the four corners of the wall-sized displays. Consequently, motion parallax was respected and participants had a sense of depth when they moved.

For the In(SITE) technique, the touch was detected by a *PQLabs*[4] infrared frame surrounding the wall-sized display. This frame is composed of infrared emitters and sensors which shoot and capture infrared light along the screen on the two directions (up-down and right-left). It can thus determine the 2D position of the fingers when they intersect with the infrared light. The main drawback of this system is that fingers can be detected slightly before they touch the screen (around $0.5cm$ from the screen). It can induce false-positive touch detections when fingers or parts of the hand are close to the screen. This system has an accuracy of $1cm$ all over the screen surface and can detect up to 32 contact points.

For the virtual ray technique, the participants held a wireless mouse attached to a Vicon target which was the starting point of the virtual ray (see Figure 3). The Vicon tracking system has an accuracy of $1mm$, but can suffer from jitter, especially for rotations. To reduce the jitter of the virtual ray, we used a *1€ filter* [Casiez et al. 2012] with the following parameters: for translations, min cut-off = 5 and beta = 0.8; for rotations, min cut-off = 0.5 and beta = 0.5. The virtual ray technique was designed as similar as possible to the In(SITE) technique. The first difference was that participants did not select objects by touching them, but by left-clicking while the virtual ray was pointing at them. The second difference was that participants did not use the In(SITE) widget, but had objects directly attached to the ray when they manipulated them. To perform rotations along the vertical axis, participants used the scroll wheel of the mouse. The right button of the mouse was used to fix the rotation of the selected object, and enabled participants to move the virtual ray without modifying the object rotation.

[2]http://unity3d.com/
[3]http://www.vicon.com/
[4]http://www.multitouch.com/

4.1.2 General Task

We asked participants to perform a docking task using the two interaction techniques. We set up a simple environment that enabled participants to perform the task without any navigation actions: a floor and a 1-meter high, 6-meter wide and 3-meter deep table, positioned just behind the screen. Participants had to put an object present at the left of the table in a target placed on the right of the table as fast as possible. The target, which was slightly bigger than the moved object (set after pilot experiments to 10% for experiment 1 and 20% for experiment 2), was red and turned green when the object was correctly placed inside. In order to validate the completion of the task, the object had to stay in the target for one second, whether it was released or not by the participant. A directional light was placed in the scene, casting shadows vertically on the table and the floor to help participants to perceive the depth in the scene. Physics, including gravity, were disabled during the experiment to avoid bias due to the physics engine.

4.1.3 Participants

16 adults (5 females and 11 males) with a mean age of 24.4 (SD 5.0) participated in both experiments. There were all right-handed. 13 of them use touch devices every day. They had variable experience with 3D visualization systems. They all performed the second experiment after the first one. Participants were not remunerated for their participation.

4.1.4 Data Collection

For each trial, we collected the task completion time. The completion time measurement started when the participant selects the manipulated object for the first time and stopped when the task was accomplished (i.e., when the object stayed in the target for one second). We also collected the number of overshoots. An overshoot was defined by the fact that the manipulated object reached the targeted position (which means that the target turns green), but could not stay within the target during the mandatory 1 second. This usually happened when participants moved the object too fast and could not adjust precisely enough its final position. After the experiments, participants filled out a subjective questionnaire.

4.2 Experiment 1: Translations

The first experiment compared the two interaction techniques and assessed the benefits of the teleportation for 3D manipulation tasks which required only translations. Consequently, we set the TECHNIQUE and the TELEPORTATION as two primary factors. Table 2 describes the TECHNIQUE×TELEPORTATION combinations.

	3D Virtual Ray	In(SITE)
No teleportation	*VRay, NoTele*	*Touch, NoTele*
Teleportation	*VRay+Tele*	*Touch+Tele*

Table 2: TECHNIQUE×TELEPORTATION *combinations in exp1.*

For teleportation, both techniques used the process described in section 3.2. The touch and long touch for In(SITE) were replaced by a click and a long click on the mouse in the case of the virtual ray technique. When the participants used teleportation, they could make the object fall down from the top in the direction of the indicated destination. The object stopped when it reached the table in the particular case of this experiment. The vertical position of the target could thus have an impact on the performance. Indeed, the object stopped at a position close to the target when the target was on the table, while it still stopped on the table even if the target was in the air. Consequently, we decided to set the VERTICAL POSITION of the target as a secondary factor to evaluate the influence of the height of the target on the performance.

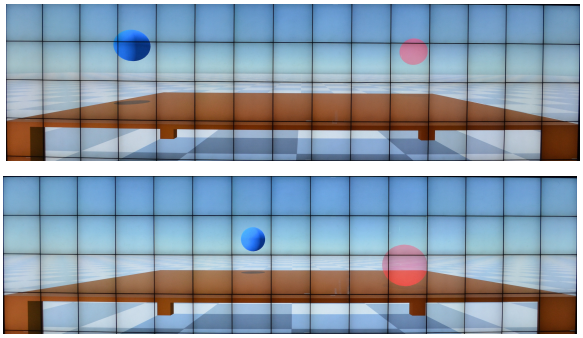

Figure 4: *Relative positions of the manipulated sphere and its associated target in Floor condition (bottom) and Midair condition (top) for experiment 1.*

4.2.1 Task

The manipulation task followed the general task described in section 4.1.2. The manipulated objects were spheres to insure that participants did not need to rotate the objects for putting them in the targets (see Figure 4). We picked 4 random positions for each object, on the left of the table for the manipulated object and on the right for the target, making sure that the objects were at least 50 cm away from each other.

4.2.2 Experimental Design & Procedure

The experiment had a [2 × 2 × 2] within-subject design with the following factors:

- TECHNIQUE, with the two treatments: *VRay* for the virtual ray technique and *Touch* for the In(SITE) technique.
- TELEPORTATION, with the two treatments: *NoTele* when the technique was not combined with teleportation and *+Tele* when the technique was combined with teleportation.
- VERTICAL POSITION, the two levels: *Floor* condition for which the position of the target was on the table and *Midair* condition for which the position of the target was above the table at a random altitude of at least 50 cm (see Figure 4).

Trials were grouped by TECHNIQUE and TELEPORTATION. The order of TECHNIQUE×TELEPORTATION was counterbalanced across participants using a balanced latin square. For each combination of TECHNIQUE×TELEPORTATION, the session began by 4 training trials, during which we explained to the participant how to use the TECHNIQUE with or without the TELEPORTATION. After the training, the participant performed 8 trials: 4 trials for the *Midair* condition using the 4 different object/target positions and 4 trials for the *Floor* condition using the same 4 object/target positions, but with a zero altitude for the target (target laying on the table). The order of the 8 trials was randomly chosen. Consequently, each one of the 16 participant performed 2 TECHNIQUES × 2 TELEPORTATIONS × 2 VERTICAL POSITIONS × 4 trials = 32 trials (bringing the total to 512 trials for the whole experiment). The participants were authorized to take a break whenever they wanted between trials, but we encouraged them to take a break between each block of the experiment corresponding to a particular combination TECHNIQUE×TELEPORTATION. Sessions lasted from 25 to 40 minutes depending on the participants.

4.2.3 Results

Task Completion Time An analysis of outliers detected two extreme outliers. The global mean for *completion time* was 14.34*s* (SD: 11.3) while the two trials corresponding to the two outliers lasted more than 130*s*. Since we had recorded during the experiment that two participants faced critical problems during these two trials, we chose to remove these two extreme outliers. One participant faced technique issues, while the other did not understand that the depth was wrong and kept trying to adjust the position during a very long time. Figure 5 illustrates the results for each TECHNIQUE×TELEPORTATION combination and VERTICAL POSITION with regard to task *completion time*.

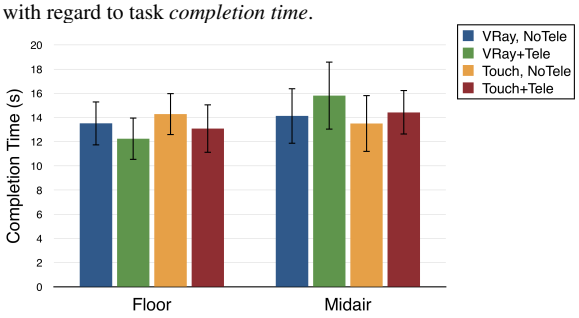

Figure 5: *Mean completion time by* TECHNIQUE×TELEPORTATION *and* VERTICAL POSITION *in experiment 1. Error bars represent 95% confidence intervals.*

A repeated measures ANOVA[5] on *completion time* with the model TECHNIQUE×TELEPORTATION×VERTICAL POSITION revealed no significant effect of TECHNIQUE, TELEPORTATION, and VERTICAL POSITION, but it showed a significant interaction effect of TELEPORTATION×VERTICAL POSITION ($F(1,14.3) = 8.3$, $p = 0.01$). Pairwise comparisons revealed that techniques combined with teleportation (*+Tele*) were significantly faster in *Floor* condition (avg. 12.66*s*) than in *Midair* condition (avg. 15.12*s* , $p < 0.001$). This showed that teleportation was more efficient when the target was laying on another horizontal object (*Floor* condition). In addition, the overall mean values of *completion time* for the two TECHNIQUEs were very close to each other (avg. *VRay*: 13.92*s* and *Touch*: 13.83*s*), which suggests that there is no practical difference between *VRay* and *Touch* with respect to the *completion time*.

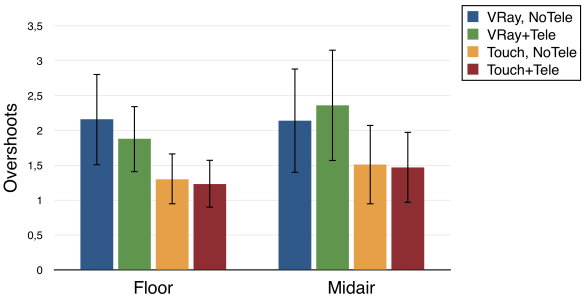

Figure 6: *Mean numbers of overshoots by* TECHNIQUE×TELEPORTATION *and* VERTICAL POSITION *in experiment 1. Error bars represent 95% confidence intervals.*

Overshoots For the *overshoots* analysis, we also excluded the two extreme outliers detected during the *completion time* analysis in order to be consistent. Figure 6 shows the mean number of *overshoots* for each TECHNIQUE×TELEPORTATION combination and VERTICAL POSITION. A repeated measures ANOVA on *overshoots* with the model TECHNIQUE×TELEPORTATION×VERTICAL POSITION revealed a significant effect of TECHNIQUE ($F(1,14.96) = 3.63$, $p = 0.002$), but no significant effect of

[5] All analyses except the ART procedure were performed with the SAS JMP statistical platform. The ART procedure [Wobbrock et al. 2011] was performed with R.

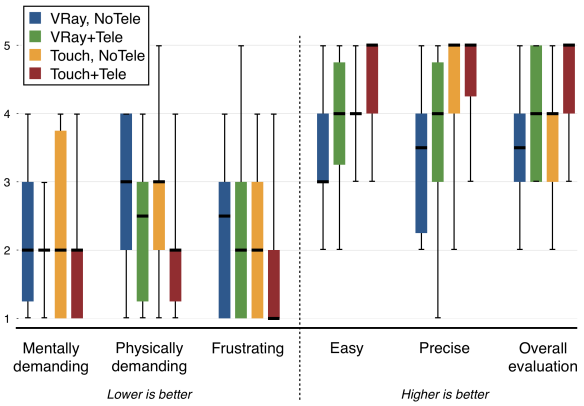

Figure 7: *Boxplots for the answers to the subjective questionnaire of experiment 1 using a 5-point Likert scale. Each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the effect over the participants. The median is also represented for each* TECHNIQUE×TELEPORTATION *combination. Error bars represent data minimum and maximum.*

TELEPORTATION and VERTICAL POSITION, as well as no significant interaction effects. The *Touch* technique (avg. 1.38) led to significantly less overshoots than the *VRay* technique (avg. 2.13).

Subjective Questionnaire After the experiment, participants had to rate on a 5-point Likert scale if each TECHNIQUE×TELEPORTATION combination was *mentally demanding*, *physically demanding*, *frustrating*, *easy* to use, *precise* (1:not at all/5:very) and they had also to give an *overall evaluation* (1:bad/5:good). Figure 7 illustrates the results of the subjective questionnaire. To analyze the results of the subjective questionnaire, we used the aligned rank transform (ART) procedure proposed by Wobbrock et al. [2011] with the model TECHNIQUE×TELEPORTATION. It showed a significant effect of TECHNIQUE on the *easy* ($F(1,45) = 19.21$, $p < 0.001$) and *precise* ($F(1,45) = 30.49$, $p < 0.001$) criteria. *Touch* was perceived easier to use (avg. 4.31 vs. 3.61) and more precise (avg. 4.47 vs. 3.5) than *VRay*. The analysis also showed a significant effect of TELEPORTATION on the *mentally demanding* ($F(1,45) = 5.36$, $p = 0.03$), *physically demanding* ($F(1,45) = 7.64$, $p < 0.001$) and *easy* ($F(1,45) = 18.41$, $p < 0.001$) criteria, as well as on the *overall evaluation* ($F(1,45) = 14.57$, $p < 0.001$). *+Tele* was perceived less mentally demanding (avg. 1.91 vs. 2.31), less physically demanding (avg. 2.19 vs. 2.75) and easier to use (avg. 4.28 vs. 3.69) than *NoTele*. *+Tele* was also preferred to *NoTele* (avg. 4.21 vs. 3.47) according to the overall evaluation. Finally, the analysis did not reveal any significant effects of TECHNIQUE and TELEPORTATION on the *frustrating* criterion, as well as any significant interaction effects of TECHNIQUE×TELEPORTATION on all the criteria.

4.3 Experiment 2: Rotations

The second experiment compared the In(SITE) technique with the virtual ray technique for 3D manipulation tasks which required rotations. In order to minimize the influence of translation on task performance, the object and the target were placed close to each other with different orientations. Since the translations to perform were really small and the teleportation did not have an effect on rotations, it seemed reasonable to consider that teleportation did not impact the task performance. Consequently, only the *VRay* and *Touch* TECHNIQUEs both without teleportation were evaluated in this experiment to shorten the global duration.

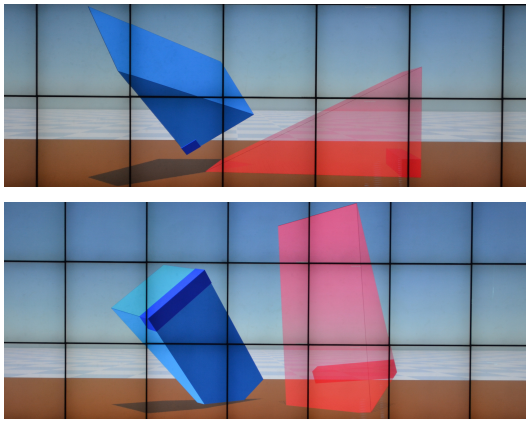

Figure 8: *Relative positions of the manipulated wedge and its associated target in Floor condition (top) and Midair condition (bottom) for experiment 2.*

4.3.1 Task

The manipulation task followed the general task described in section 4.1.2. The manipulated objects were wedges (triangular prisms with one right angle) to insure that there was only one correct rotation to fit the object in the target. Since it was not always easy to determinate the angle values in the 3D virtual environment, the right angle of the wedge was marked on both the object and the target to help participants identifying the rotation to perform (see Figure 8). The manipulated object and the target were placed 1*m* away from each other and the manipulated object was located on the left of the target. We picked random values for the orientations of both the object and the target. The rotation was considered as correct if the dot product between the manipulated object and the target's up, right, and forward were bigger than or equal to 0.98 (set after pilot experiments).

4.3.2 Experimental Design & Procedure

The experiment had a [2 × 2] within-subject design with the following factors:

- TECHNIQUE, with the two treatments: *VRay* and *Touch* both without teleportation.
- VERTICAL POSITION, the two levels: *Floor* and *Midair* conditions. For the *Midair* positions, the target height was set at 50 cm above the table. For the *Floor* positions, the target position was defined with one side of the target laying on the virtual table (see Figure 8).

Trials were grouped by TECHNIQUE and the order of the TECHNIQUEs were counterbalanced across participants. For each TECHNIQUE, the session began by 4 training trials. After the training, the participant performed 8 trials: 4 trials for the *Floor* condition and 4 trials for the *Midair* condition. The order of the 8 trials was randomly chosen. Consequently, each one of the 16 participant performed 2 TECHNIQUEs × 2 VERTICAL POSITIONs × 4 trials = 16 trials (bringing the total to 256 trials for the whole experiment). The participants were authorized to take a break whenever they wanted between trials, but we encouraged them to take a break between the two blocks of the two TECHNIQUEs. Sessions lasted from 15 to 20 minutes depending on the participants.

4.3.3 Results

Task Completion Time Figure 9 illustrates the results for each TECHNIQUE and VERTICAL POSITION with respect to task *com-*

pletion time. A repeated measures ANOVA on *completion time* with the model TECHNIQUE×VERTICAL POSITION showed a significant effect of VERTICAL POSITION ($F(1,15) = 25.55$, $p < 0.001$). Participants performed the task significantly faster in *Floor* condition (avg. 27.4*s*) than in *Midair* condition (avg. 43.61*s*). However, the analysis did not reveal any significant effects of TECHNIQUE or interaction effect of TECHNIQUE×VERTICAL POSITION.

Overshoots Figure 10 shows the mean number of *overshoots* for each TECHNIQUE and VERTICAL POSITION. A repeated measures ANOVA with the model TECHNIQUE×VERTICAL POSITION showed a significant effect of TECHNIQUE ($F(1,15) = 20.25$, $p < 0.001$) on *overshoots*. *Touch* (avg. 0.45) led to significantly less overshoots than *VRay* (avg. 0.84). However, the analysis revealed neither a significant effect of VERTICAL POSITION nor a significant interaction effect of TECHNIQUE×VERTICAL POSITION.

Subjective Questionnaire After the experiment, participants filled out a similar questionnaire to the one of experiment 1, but only for the two TECHNIQUE conditions (*Touch* and *VRay*). Pairwise Wilcoxon rank sum tests with Bonferroni corrections among TECHNIQUEs did not reveal any significant differences for all the criteria, and for the *overall evaluation*.

5 Discussion

On one hand, the two experiments did not reveal any significant effects of TECHNIQUE on completion time. This result, in addition to the fact that the overall mean values of completion time were very similar for the two techniques, suggests that users can reach the same level of performance with the In(SITE) technique in comparison with a standard 3D interaction technique even if the manipulation tasks require both translations and rotations. It suggests a validation of **H1** and **H2**.

On the other hand, the results of both experiments revealed that the In(SITE) technique led to a significant smaller number of overshoots in comparison to the technique based on a virtual ray. It shows that the In(SITE) technique was more precise for fine adjustments of 3D object position in the immersive context of large vertical displays. This result is consistent with the fact that participants perceived the In(SITE) technique as more precise than the virtual ray technique for manipulation tasks which required only translations. Concerning **H3**, we can note that the In(SITE) technique was more precise for fine adjustments of 3D object position, which was confirmed by the subjective questionnaire only for translations. For translations, the In(SITE) technique was also perceived easier to use by the participants.

Moreover, the results of the first experiments revealed that teleportation significantly improved the performance of the two techniques in terms of completion time for the *Floor* condition. This result seems coherent since users can select a destination for teleportation close to the targeted position in this condition. In addition, the subjective questionnaire of the first experiment showed that teleportation significantly reduced the mental and physical loads. For the mental load, we cannot formulate a strong claim, but it might be explained by the fact that the participants only had to focus on the destination with teleportation and not on the full path to reach the target. For the physical load, it was probably due to the gorilla-arm effect [Hincapié-Ramos et al. 2014] which could affect participants if they hold the 3D virtual ray during a long time or drag an object all along the screen with In(SITE). Teleportation avoids to always raise the arm during interaction and reduces the arm fatigue. This is also confirmed by the fact that participants preferred and found easier to use the two techniques when they were combined with teleportation. These results show that **H4** is confirmed and that the teleportation has a real benefit on large wall-sized displays.

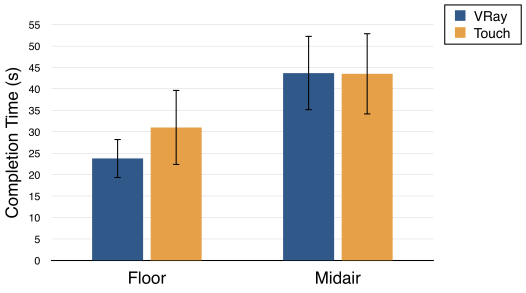

Figure 9: *Mean completion time by* TECHNIQUE *and* VERTICAL POSITION *in experiment 2. Error bars represent 95% confidence intervals.*

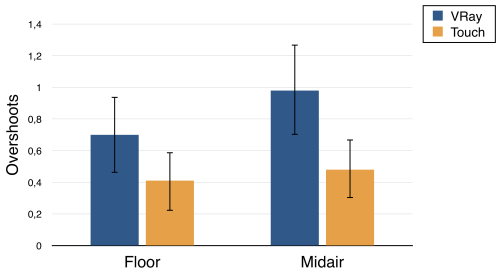

Figure 10: *Mean numbers of overshoots by* TECHNIQUE *and* VERTICAL POSITION *in exp2. Error bars represent 95% confidence intervals.*

Finally, the difference on *completion time* between the *Floor* and *Midair* conditions in the second experiment can be explained by the fact that the targeted positions in *Floor* condition were defined with one side of the object laying on the virtual table. Consequently, the performed rotations were less complicated and the participants achieved more quickly the manipulation task in *Floor* condition.

6 Conclusion and Future Work

This paper aims to assess the usability of the In(SITE) technique combined with teleportation for 3D manipulation on multi-touch wall-sized displays. The In(SITE) technique is a bimanual touch-based technique for 3D interaction adapted to the situation where users are immersed in a 3D virtual environment and move in front of a large vertical screen. The usability testing shows that the In(SITE) technique can reach the same level of performance than a standard 3D interaction technique with respect to completion time. In addition, the In(SITE) technique improves precision for fine adjustments of 3D object position. The results also show that techniques combined with teleportation were found easier to use, less tiring and were globally preferred by participants for translation tasks.

This paper is a first study of combining a bimanual touch-based interaction with teleportation on large wall-sized display. Since we demonstrated that the In(SITE) technique can be beneficial in such context, we want now to improve it by adding scaling functionalities and to explore different variations of the proposed widget to question the design choices we have done. For example, it will be interesting to study if the choice of the rotation axis of the manipulated object (object gravity center or selection point of the user) or if the choice of the transfer function (which is currently linear) used to control the translation along the z-axis as studied by Casiez et al. [2008] have some impacts on performance. In addition, we want to

study more in depth the different teleportation strategies such as the animated teleportation (objects falling down when a physical simulation is used) or the direct teleportation (under the user's finger).

The wall-sized display used during the experiments does not support stereoscopy, but we think that the In(SITE) technique could be extended to stereoscopic vision using the guidelines defined by Valkov et al. [2011]. Consequently, it would be interesting to study the In(SITE) technique with another system which support stereoscopy and to compare the results. In addition, we want to adapt and study the In(SITE) technique for systems which do not use head-tracking and viewing frustum deformation. These systems are especially relevant to enable multiple users to collaborate together in front of a same wall-sized display. We think that In(SITE) could be easily adapted to this context by using the center of the virtual camera instead of the user's head as the starting of the raycast when performing object selection and manipulation.

Finally, touch-based 3D interaction techniques could be an issue for remote collaboration in a shared virtual environment. Indeed, when a user interacts with 3D objects through his own multi-touch device, it is impossible for the remote users to understand which objects he is manipulating. It is an issue for the communication between the remote users. We need to provide a feedback which links the user's 2D touch inputs to the manipulated 3D objects. A similar solution to the 2D pointer / 3D ray proposed by Duval and Fleury [2009] could be an interesting way to implement this feedback.

Acknowledgements

This research was supported by the French National Research Agency under grant ANR-10-EQPX-26-01 DIGISCOPE.

References

BENKO, H., AND FEINER, S. 2007. Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selection. In *2007 IEEE Symposium on 3D User Interfaces*, IEEE Computer Society, 3DUI '07.

BOWMAN, D. A., AND HODGES, L. F. 1997. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proc. 1997 Symposium on Interactive 3D Graphics*, ACM, I3D '97, 35–ff.

CASIEZ, G., VOGEL, D., BALAKRISHNAN, R., AND COCKBURN, A. 2008. The Impact of Control-Display Gain on User Performance in Pointing Tasks. *Human-Computer Interaction 23*, 3, 215–250.

CASIEZ, G., ROUSSEL, N., AND VOGEL, D. 2012. 1€ filter: A simple speed-based low-pass filter for noisy input in interactive systems. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '12, 2527–2530.

COFFEY, D., MALBRAATEN, N., LE, T. B., BORAZJANI, I., SOTIROPOULOS, F., ERDMAN, A. G., AND KEEFE, D. F. 2012. Interactive Slice WIM: Navigating and Interrogating Volume Data Sets Using a Multisurface, Multitouch VR Interface. *IEEE Transactions on Visualization and Computer Graphics 18*, 10, 1614–1626.

COHÉ, A., AND HACHET, M. 2012. Understanding User Gestures for Manipulating 3D Objects from Touchscreen Inputs. In *Proc. Graphics Interface 2012*, Canadian Information Processing Society, GI '12, 157–164.

COHÉ, A., DÈCLE, F., AND HACHET, M. 2011. tBox: A 3D Transformation Widget Designed for Touch-screens. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '11, 3005–3008.

DUVAL, T., AND FLEURY, C. 2009. An Asymmetric 2D Pointer/3D Ray for 3D Interaction Within Collaborative Virtual Environments. In *Proc. 14th International Conference on 3D Web Technology*, ACM, Web3D '09, 33–41.

GILLIOT, J., CASIEZ, G., AND ROUSSEL, N. 2014. Direct and Indirect Multi-touch Interaction on a Wall Display. In *Proc. 26th Conference on L'Interaction Homme-Machine*, ACM, IHM '14, 147–152.

HANCOCK, M., CARPENDALE, S., AND COCKBURN, A. 2007. Shallow-depth 3D Interaction: Design and Evaluation of One-, Two- and Three-touch Techniques. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '07, 1147–1156.

HANCOCK, M., TEN CATE, T., AND CARPENDALE, S. 2009. Sticky Tools: Full 6DOF Force-based Interaction for Multi-touch Tables. In *Proc. ACM International Conference on Interactive Tabletops and Surfaces*, ACM, ITS '09, 133–140.

HINCAPIÉ-RAMOS, J. D., GUO, X., MOGHADASIAN, P., AND IRANI, P. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '14, 1063–1072.

JACKSON, B., SCHROEDER, D., AND KEEFE, D. F. 2012. Nailing Down Multi-Touch: Anchored Above the Surface Interaction for 3D Modeling and Navigation. In *Proc. Graphics Interface 2012*, Canadian Information Processing Society, GI '12, 181–184.

JACOBY, R. H., FERNEAU, M., AND HUMPHRIES, J. 1994. Gestural interaction in a virtual environment. In *Proc. of Stereoscopic Displays and Virtual Reality Systems (SPIE)*, 355–364.

JANKOWSKI, J., AND HACHET, M. 2015. Advances in Interaction with 3D Environments. *Computer Graphics Forum 34*, 1, 152–190.

LOPEZ, D., OEHLBERG, L., DOGER, C., AND ISENBERG, T. 2016. Towards An Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration. *IEEE Transactions on Visualization and Computer Graphics 22*, 5, 1616–1629.

MARTINET, A., CASIEZ, G., AND GRISONI, L. 2010. The Design and Evaluation of 3D Positioning Techniques for Multi-touch Displays. In *Proc. 2010 IEEE Symposium on 3D User Interfaces*, IEEE Computer Society, 3DUI '10, 115–118.

MARTINET, A., CASIEZ, G., AND GRISONI, L. 2010. The Effect of DOF Separation in 3D Manipulation Tasks with Multi-touch Displays. In *Proc. 17th ACM Symposium on Virtual Reality Software and Technology*, ACM, VRST '10, 111–118.

MINE, M. R. 1995. Virtual Environment Interaction Techniques. Tech. rep., University of North Carolina, Department of CS.

MÖLLERS, M., ZIMMER, P., AND BORCHERS, J. 2012. Direct Manipulation and the Third Dimension: Co-planar Dragging on 3D Displays. In *Proc. 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ACM, ITS '12, 11–20.

PIERCE, J. S., STEARNS, B. C., AND PAUSCH, R. 1999. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. In *Proc. 1999 Symposium on Interactive 3D Graphics*, ACM, I3D '99, 141–145.

REISMAN, J. L., DAVIDSON, P. L., AND HAN, J. Y. 2009. A Screen-space Formulation for 2D and 3D Direct Manipulation.

In *Proc. 22Nd Annual ACM Symposium on User Interface Software and Technology*, ACM, UIST '09, 69–78.

STROTHOFF, S., VALKOV, D., AND HINRICHS, K. 2011. Triangle Cursor: Interactions with Objects Above the Tabletop. In *Proc. ACM International Conference on Interactive Tabletops and Surfaces*, ACM, ITS '11, 111–119.

VALKOV, D., STEINICKE, F., BRUDER, G., AND HINRICHS, K. 2011. 2D Touching of 3D Stereoscopic Objects. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '11, 1353–1362.

WOBBROCK, J. O., FINDLATER, L., GERGLE, D., AND HIGGINS, J. J. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '11, 143–146.

YU, L., SVETACHOV, P., ISENBERG, P., EVERTS, M. H., AND ISENBERG, T. 2010. FI3D: Direct-Touch Interaction for the Exploration of 3D Scientific Visualization Spaces. *IEEE Transactions on Visualization and Computer Graphics 16*, 6, 1613–1622.

Investigating Collaborative Exploration of Design Alternatives on a Wall-Sized Display

Yujiro Okuya[1,2] **Olivier Gladin**[2] **Nicolas Ladevèze**[1] **Cédric Fleury**[2] **Patrick Bourdot**[1]

[1] Université Paris-Saclay, CNRS, LIMSI
VENISE team, F-91405 Orsay, France

[2] Université Paris-Saclay, CNRS, Inria, LRI
F-91405 Orsay, France

yujiro.okuya@limsi.fr, olivier.gladin@inria.fr, nicolas.ladeveze@limsi.fr, cedric.fleury@lri.fr, patrick.bourdot@limsi.fr

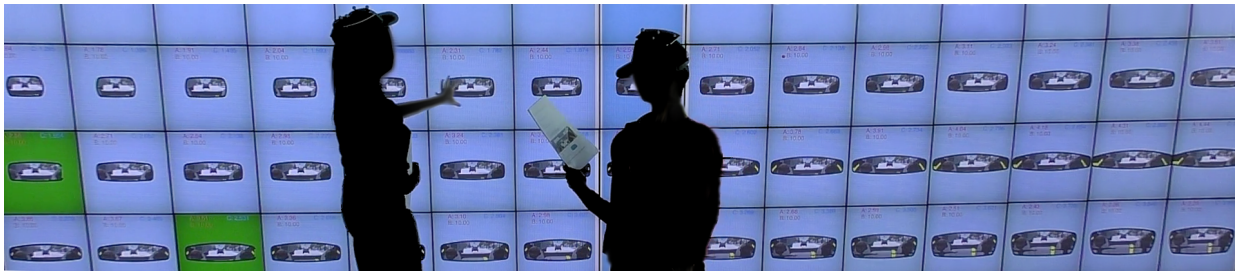

Figure 1. ***ShapeCompare*** **enables non-CAD experts to generate and explore multiple design alternatives on a wall-sized display.**

ABSTRACT

Industrial design review is an iterative process which mainly relies on two steps involving many stakeholders: design discussion and CAD data adjustment. We investigate how a wall-sized display could be used to merge these two steps by allowing multidisciplinary collaborators to simultaneously generate and explore design alternatives. We designed *ShapeCompare* based on the feedback from a usability study. It enables multiple users to compute and distribute CAD data with touch interaction. To assess the benefit of the wall-sized display in such context, we ran a controlled experiment which aims to compare *ShapeCompare* with a visualization technique suitable for standard screens. The results show that pairs of participants performed a constraint solving task faster and used more deictic instructions with *ShapeCompare*. From these findings, we draw generic recommendations for collaborative exploration of alternatives.

Author Keywords

Computer-aided design; collaboration; wall-sized display.

CCS Concepts

•**Human-centered computing** → **Graphical user interfaces; Collaborative interaction;**

Yujiro Okuya, Olivier Gladin, Nicolas Ladevèze, Cédric Fleury and Patrick Bourdot. Investigating Collaborative Exploration of Design Alternatives on a Wall-Sized Display. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI 2020)*, ACM, April 2020. Paper 607.

ISBN 9781450367080
http://dx.doi.org/10.1145/3313831.3376736

INTRODUCTION

The industrial design increasingly relies on digital tools to review aesthetic properties, user satisfaction, and technical feasibility of products before building physical prototypes. Computer-aided design (CAD) is now an essential part of the design process. At specific stages of the process, multidisciplinary teams (e.g. designers, engineers, ergonomists) evaluate and adjust product design using digital mock-ups [37]. These experts need a shared workspace to review design alternatives. With advances in technology, large displays and virtual reality systems are becoming useful in such context [10, 43, 7]. They offer opportunities to visualize digital mock-ups and foster collaboration. For example, automotive industries are now using CAVE systems or large screens to review virtual cars at full scale or in a realistic environment.

However, allowing all experts to modify CAD data in such interactive systems is challenging, as mastering CAD skills is complex, time-consuming and costly [9]. Consequently, the review process is often iterative. For example, we identified the following design practices by interviewing engineers at PSA Group[1]: engineers prepare design alternatives from CAD data based on expert recommendations; project members review prepared digital mock-ups within an interactive system; then, engineers apply post-modifications on the CAD data and prepare new alternatives based on annotations. As experts cannot directly reflect their ideas, miscommunication could occur resulting in unnecessary iterations and increased development time. A side-by-side setup of an interactive system and a CAD workstation [46] could reduce iterations, but still, only the engineers can apply the modifications.

[1]French multinational automotive company gathering Peugeot, Citroën, DS, Opel and Vauxhall brands (https://www.groupe-psa.com).

The challenge addressed in the paper is to merge the design discussion and CAD data adjustment steps by providing a shared interactive workspace in which experts can explore several design alternatives. Wall-sized displays [6] are promising in such context because they offer new ways to collaborate and interact with large data sets. Previous work demonstrated their benefits over traditional desktop displays for a co-located collaboration, as several people can simultaneously interact with information [4, 24, 31].

In this paper, we investigate how wall-sized displays can improve the review process in industrial design. In particular, we design *ShapeCompare* which allows users to generate and distribute multiple alternatives of CAD data on a wall-sized display using touch interaction (Fig. 1). *ShapeCompare* is linked to a commercial CAD engine (Catia V5) and enables non-CAD experts to modify native CAD data by directly retrieving multiple design alternatives from the CAD engine. Users are thus able to explicitly express their design ideas and compare the proposed alternatives. To assess the benefit of wall-sized displays for collaborative design tasks, we ran a controlled experiment which compared *ShapeCompare* with a visualization technique suitable for standard screens. It evaluates whether comparing many design alternatives on a wall-sized display is more beneficial than exploring them one by one, as it can be done on a standard screen.

The main contributions of this paper include: i) a system that enables non-CAD experts to modify native CAD data by generating multiple design alternatives and distributing them on a wall-sized display; ii) a controlled experiment which shows that wall-sized displays can improve collaboration for design alternative exploration; iii) design recommendations which generalize the use of a wall-sized display to other types of alternative exploration in various contexts.

Section 2 examines related work. Section 3 describes the design of *ShapeCompare* based on the feedback from a usability study. Section 4 details the system implementation. Section 5 reports the controlled experiment. Section 6 discusses the results and section 7 presents the design recommendations. Section 8 concludes by highlighting future work.

RELATED WORK

The challenge addressed in this paper is to create a shared interactive workspace where non-CAD experts could generate multiple design alternatives and collaboratively explore them. We first review techniques to modify CAD data for non-experts. We then present the use of large interactive platforms in industrial design. Finally, we examine how wall-sized displays foster collaboration on complex data sets.

Design for Non-CAD Experts

Many interaction techniques were proposed to facilitate drawing and sketching at early stages of the design, such as immersive drawing [23, 45], surface modeling [17], digital tape-drawing [2, 20, 18, 25] and rapid prototyping with bimanual interaction [1]. However, much fewer studies target detailed design stages where modification of parametric CAD models is mandatory. A CAD model is a solid model defined by a set of mathematical operations (e.g. extrusion, boolean operations) applied on 2D sketches. Unlike drawing or surface modeling, users need to interact with parameters which requires extensive training. Consequently, modifications of native CAD data is cumbersome for non-CAD experts.

Martin et al. [36] presented a data pipeline which allows CAD data modification from a virtual reality platform. Based on this work, Okuya et al. [39] proposed a shape-based interaction, which enables non-CAD experts to modify parameter values of CAD models by simply pushing and pulling their 3D shape. Coffey et al. [13] proposed to browse pre-computed design of a medical device by dragging its surface on a tablet.

Although some solutions enable non-experts to modify CAD data, they are limited to a single CAD model and none of them supports the generation of multiple design alternatives.

Design Reviews in Interactive Platforms

Collaborative reviews is a critical part of the industrial design process. Such meetings are often conducted on interactive platforms, such as large screens or virtual reality systems. They offer a full-scale design visualization, a large interactive space, and a collaborative environment. For example, "*the ability to display and interact with large-scale representations of vehicles has always been a fundamental requirement*" [10] in the automotive industry. Portfolio Wall [10] displayed multiple different designs as tiled thumbnails on the screen. It was designed to compare various concepts, like a traditional wall-mounted corkboard. Khan et al. [26] also studied a tool that highlights the area of the user attention on a projected display to facilitate group meetings.

Previous works explored 3D visualization to reduce time and costs for manufacturing physical mock-ups in product development process [29, 49]. While many virtual reality systems can display CAD data [7, 42, 43, 27, 46], only a few can modify native data from commercial CAD systems [36, 39]. Other works addressed remote collaboration: vehicle design reviews between remote CAVE [30], collaborative material/texture editions [11], or object manipulations [35].

While these systems provide a 3D visualization and multi-user context to facilitate discussions, only a few static design alternatives can be compared during each review meetings. Generating and modifying new design alternatives of CAD data is currently not possible.

Wall-sized Displays

The benefits of wall-sized displays have been demonstrated on various tasks. Ball et al. [4] showed that users' physical navigation induced by a large display improves performance on navigation, search and pattern finding. Ball and North found that peripheral vision offered by wall-sized displays contributed to task performance [3]. The large amount of data displayed improves task efficiency and accuracy [48]. Larger screens provide better performance on complex and cognitively loaded tasks [15, 33], and enhance peripheral application awareness [8] compared to workstations.

Wall-sized displays also enhance collaboration among co-located users. Previous studies investigated how user interaction affected collaboration [28, 16] for data analysis tasks. Liu

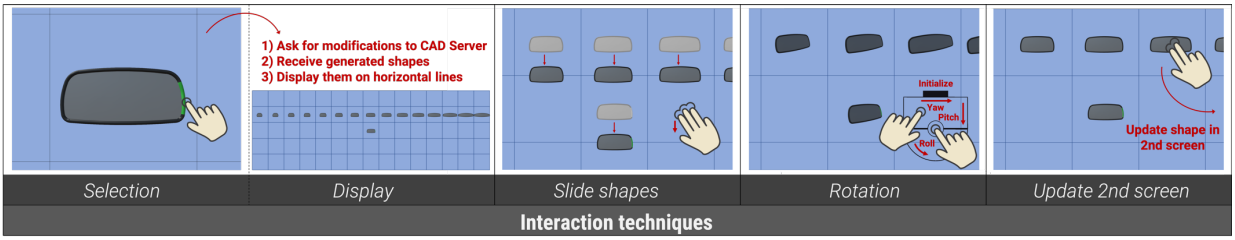

Figure 2. Interaction with *ShapeCompare*: when a part is selected (*Selection*), the system generates a set of design alternatives on a row of the wall-sized display (*Display*). All alternatives can be scrolled up and down with a three-finger drag (*Slide shapes*), rotated in 3D with a widget (*Rotation*), and displayed on the external screen with a two finger long press (*Update 2nd screen*).

et al. [31] compared various collaboration strategies on a wall-sized display. They also showed that cooperative interaction improves close and loose collaboration[32].

Since wall-sized displays are powerful tools for displaying and interacting with large data sets, they can allow to visualize multiple alternatives by creating "small multiple" representations of an object. For example, Beaudouin-Lafon [5] proposed to distribute a large number of brain scans on the screen. This enables neuroscientists to compare and classify this brain scan according to specific brain fold patterns. For industrial design reviews, the "small multiple" approach can be valuable for exploring and comparing design alternatives.

Summary and Approach

While some previous work explored CAD data modification for non-experts, they focused on the deformation of one particular CAD model and did not consider the generation of multiple design alternatives. Other works investigated reviews of static design alternatives, but they did not allow users to modify these alternatives or to generate new ones. As Wall-sized displays are efficient to show multiple variations of a same object and to foster collaboration, they could be an ideal tool for design discussions where a multidisciplinary team can review, compare, and also generate design alternatives without using a conventional CAD system.

DESIGN OF SHAPECOMPARE

To assess the benefit of a wall-sized display for collaborative exploration of design alternatives, we first need to design a system allowing non-CAD experts to modify native CAD data and distribute alternatives on the screen. We created *ShapeCompare* by using an iterative design process involving potential users. The prototypes and the usability study used during this process are described in this section.

First Prototype

The first prototype was designed to meet three criteria: i) interaction in a large space, ii) native CAD data modification and iii) multiple-design comparison. Based on previous work on CAD data modification [39, 40], we implemented a service which generates multiple alternative shapes by varying parameter values of a native CAD model.

User Interaction

Various interaction techniques have been studied on a Wall-sized display [19, 47, 38]. However, as the interaction technique in itself is not the main focus of the paper, we decided to simply use direct touch to interact with the CAD data displayed on the wall-sized display.

Shape Generation

To generate new design alternatives, users touch the part to modify on one of the displayed shapes (Fig. 2). If the part can be modified, it turns green. Each part is tied to an internal parameter defined in the native CAD data. The system then prepares a set of parameter values for the selected part and asks for the corresponding shapes. In this first prototype, we defined a minimum and maximum parameter value for each part, and chose a pre-defined number of values equally distributed in the range. As CAD models are defined by multiple geometric constraints, the CAD engine cannot always perform the modification for the full range of values. If a shape is not successfully generated, a "cross" is displayed to inform users of the failure.

Shape Visualization

The set of newly generated shapes is displayed on a full row of the screen, above previous versions of the CAD model. In this manner, each row represents a set of design alternatives for a specific part of the CAD model.

Design History

Users thus accumulate design history below the current design alternatives, and can navigate with a three-finger interaction to scroll up or down the design alternatives (Fig. 2). Users can select a part of any shape in the design history and start over modification from this shape.

3D Rotation

To interact with 3D objects, we implemented the In(SITE) technique [34]. It allows to perform 3D rotation on a wall-sized display with bimanual touch interaction (Fig. 2). The rotation of displayed alternatives is synchronized to maintain a similar view angle for all of them.

Usability Study

To assess the usability of the interaction technique proposed in the first prototype, we invited potential users to test it and observed their behaviors while achieving an individual CAD

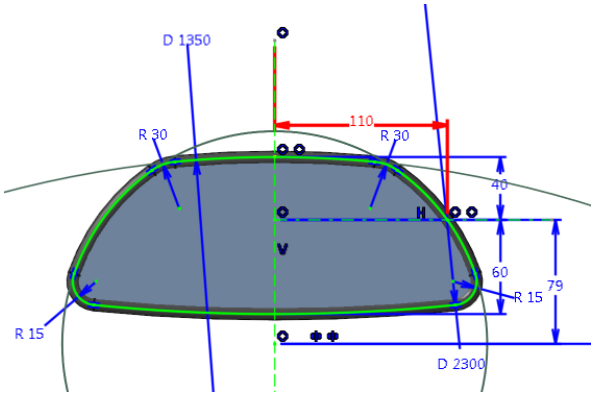

Figure 3. Sketch of the *rear-view mirror* designed using CATIA V5. Green line is a guide curve of a sweep operation generating the 3D shape. Constraints and parameter values are highlighted in blue: for example, the red parameter defines the width of the *rear-view mirror* (110 mm).

Figure 4. Usability study setup: the target shape is displayed with a transparent yellow color in a realistic environment on an external screen next to the wall-sized display.

modification task. After the task, we had short individual debriefings and a brainstorming session with all of them.

Participants

The 5 participants (1 female), aged 21 to 24 (mean 22.8), are students at the civil engineering department. 4 rated their expertise level of CAD system as 2 out of 5, and 1 rated as 1 (*1: Never used before - 5: Use almost every day*). All students have experience in AutoCAD to read construction plans and design. Although they are not CAD experts, they provided us relevant feedback as they are knowledgeable about parametric modeling and design process.

Task

We asked each participant to modify a CAD object with *ShapeCompare* to reach a given target shape within a time limit of 5 minutes. The CAD object is a rear-view mirror (Fig. 3) designed by an industrial designer at PSA Group. An external screen next to the wall-sized display was used to display a design alternative of the mirror within an automotive cockpit. Participants could display a particular alternative on this external screen by selecting it on the wall-sized display with a two finger long press. The target shape was shown with a transparent yellow color overlaying the design alternative on the external screen (Fig. 4). This simulated the fact that participants have design skills to evaluate the design alternatives in a realistic environment.

To investigate if the interaction technique used for CAD modification was straightforward, we did not provide any CAD parameter-related information to the participants. We just explained to them that the mirror is a parametric CAD model designed with a commercial CAD system, and what are the actions to modify it.

Results

Based on the interviews and observations of participant behaviors, we extracted the main issues they encountered. First, all participants mentioned that it was difficult and frustrating to figure out how the part selection affects the shape deformation. P2 explained, "*I didn't know the link between the parameter map and shapes. Without this information, I was afraid to modify the shape wrongly*". Most participants expected that each "length" of the selected part would change. For example, P3 said, "*I was surprised when I selected the right side. I expected that it would get higher but it got wider instead. I had to adjust my mind after each system response*". The links between parts and parameters are defined by the CAD data, and each parameter is mapped to the part that constraints the parameter. For example, when the right-side part is selected, it changes the parameter that constraints the width (Fig. 3).

Second, all participants often needed time to find out how the generated shapes on the new row are different from the one they selected. P3 detailed, "*I expected that the same shape will appear just above the selected shape, but it didn't. I was surprised of this behavior. I had to make a step behind to look for the exact same shape*" in the new row. In fact, as the parameter values used for generating shapes are always distributed between a static minimum and maximum, the initial shape in the range of variations is not displayed above the previously selected shape, but at a random position.

Aside from these critical issues, participants often had difficulties to find the difference also between neighbor shapes, especially on radii of corners, top and bottom parts. Some of them claimed that the difference of 4 shapes on a row was not noticeable. They needed to step back from the screen and check the smallest and the largest shapes (leftmost and rightmost ones) to grasp the global difference of the shapes on the row. P2 commented, "*I was looking at the first and the last shapes of the list to understand the modification. Then looked in the middle to find the one I want*".

Despite the above issues, we had positive reactions of all participants. They liked the simple interaction which does not require to understand and manipulate parameter values. In particular, they found it useful for novices: "*I found that it is very difficult in AutoCAD to find parameter values, remember the parameter information and how it affects the shape*

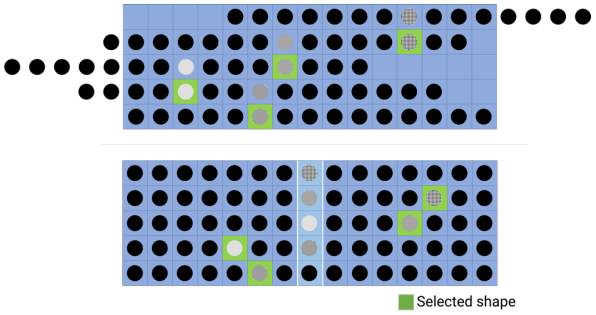

Figure 5. Two solutions for visualizing the design history.

modification, but this app made it much easier" (P5). They also appreciated the proposed shape visualization: "*Seeing many different designs on the wall-sized display, comparing them, and visualizing them within the realistic context was nice to generate new ideas*" (P1). As common ground, all of them agreed that *ShapeCompare* cannot be a substitute for traditional CAD software due to the limited functionalities, but it is valuable for a design adjustment task which does not require to change the whole design intent.

Updated Prototype

Based on the usability study results, we identified two main issues in the first prototype with respect to: *(i) Understanding of shape modification* and *(ii) Visualization of design history*. We addressed these issues as described in the following.

Understanding of Shape Modification

To solve this issue, we drew inspiration from the *Suggestive Interface* proposed by Igarashi [22]. This interface offers small thumbnails presenting results of geometric operations and encourages novice users to explore a new system and find unknown operations. We decided to compute and displayed a preview of the minimum and maximum shape modification for each part. As participants often checked the smallest and the largest shapes to understand the current modification during the usability study, these extreme shapes would have a similar effect and give a hint to the users before selecting the modification. In the updated prototype, when users select a shape, a *selection widget* appears and displays thumbnails of the two extreme shape modifications for all parameters of each part. (Fig. 6, (2)). Users can then select a thumbnail to generate the corresponding design alternatives. They no longer have to touch the parts directly, which prevents touch problems.

Visualization of Design History

To improve the shape visualization, we changed how the system generates design alternatives to ensure that we have the same number of design alternatives with lower and higher parameter values compared to the selected one. Instead of defining a static minimum and maximum for the parameter values, we defined a specific offset for each part. The system thus generates half of shapes with lower parameter values by incrementally decreasing by the offset the parameter of the selected shape, and conversely for the other half with higher parameter values. Consequently, the selected shape always appears in the middle of the alternative range, and not at a random position as in the first prototype. We then explored solutions to display each row of new design alternatives. We first showed the similar shape just above the selected one (Fig. 5, top) in order to meet the expectation of the participant of the usability study. However, it creates unused screen space and requires an additional interaction for horizontal panning. For these reasons, we shifted the new alternatives to the center of the screen to always fit them in. This solution displays all the selected shapes always in the central column (Fig. 5, bottom), which is beneficial for keeping track of the modification history. We highlighted the selected shapes with a green color and the central column with a different saturation.

SYSTEM IMPLEMENTATION

The front-end application of *ShapeCompare* communicates with a back-end server (*CAD Server*) linked to a CAD engine which computes tessellated meshes of design alternatives. Both components are located in remote locations on our campus and are connected by a distributed architecture [40].

Wall-Sized Display

The wall-sized display consists of a 15×5 grid of 21.6" LCD screens. It measures $5.9 \times 2m$ for a resolution of 14.400×4.800 pixels. It is controlled by a cluster of 10 PCs running Linux, each managing a row of 7 or 8 screens. Touch interaction is detected by a *PQLABs*[2] infrared frame surrounding the wall-sized display. A *VICON*[3] infrared tracking system tracks users' head positions and orientations.

CAD Server

To load and modify native CAD source files, we implement a *CAD Server* based on the work of Okuya et al. [39]. The *CAD Server* is a custom C++ application using the CAA API of CATIA V5[4]. It can load original CAD files, update the Constructive History Graph (CHG) and the Boundary Representation (B-Rep) when receiving modification requests, and send back the tessellated meshes. The main concept of the *CAD Server* is *labeling* [14], a direct linkage from 3D meshes to B-Rep elements and CHG nodes of the CAD object. With this linkage, users can access to the parameter values of the CHG node by selecting a relevant mesh displayed on a user interface. Once selected, the front-end application transmits the B-Rep ID, constraint ID and the new parameter value to the *CAD Server* to request a modified shape.

Software

The user interface on the wall-sized display is implemented with Unity[5]. Unity manages the clustered rendering. The master node of the cluster handles communication, receives and stores 3D meshes in a local cache.

[2] https://www.pqlabs.com/
[3] http://www.vicon.com/
[4] https://www.3ds.com/products-services/catia/
[5] http://unity3d.com/

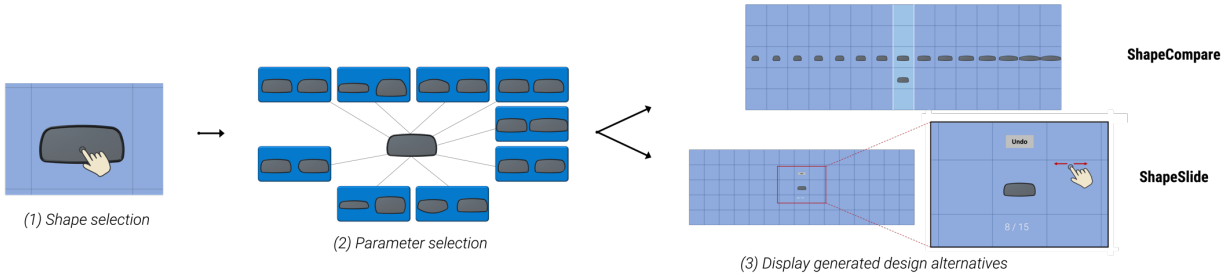

Figure 6. Interaction in experimental conditions: after the shape selection (1), a widget helps users to select a part by displaying the two extreme shape modifications (2). Then users can explore design alternatives using *ShapeCompare* or *ShapeSlide*.

EXPERIMENT

We conducted a controlled experiment to assess the benefit of a wall-sized display in the context of collaborative design reviews. In particular, we investigated how simultaneous visualization of multiple design alternatives affects the collaboration between participants. We compared *ShapeCompare* to another technique called *ShapeSlide*, which displays only one shape at a time (Fig. 6). With *ShapeSlide*, users can change the shape displayed at the center of the wall-sized display with a sliding gesture on the screen.

Although *ShapeSlide* is suitable for a standard screen, we implemented it on the wall-sized display to avoid bias which could be introduced by the devices, user positions, or interaction techniques. Consequently, we used the same device (i.e. the wall-sized display) for both conditions. Only a small portion of the wall-sized display was used for *ShapeSlide*, simulating the use of a smaller screen. This reduces bias that could be introduced by different devices or standing-sitting positions. It also simplifies the experiment since participants did not have to change devices. In addition, we decided to use the exact same widget and interaction technique for selecting the part to modify on the CAD model for both conditions. (Fig. 6). Only the way to browse the generated design alternatives differs in the two conditions: users had to walk in front of the wall-sized display with *ShapeCompare*, while they had to use a sliding gesture with *ShapeSlide*. Finally, to fairly compare the conditions, we provided the same functionalities in both cases. In particular, we imitated the design history of *ShapeCompare* by implementing an "Undo" button in *ShapeSlide*, which allows users to go back to the set of design alternatives that were previously generated.

Task

We designed the experimental task based on actual industrial practices collected through interviews with engineers at PSA Group. The task was to modify a car rear-view mirror and simulate expert negotiation on several design criteria. Since it was difficult to find and invite real experts involved in an actual industrial design process, we controlled users' expertise by giving individual design criteria to pairs of participants. We simulated two distinct expert: *Specialist 1* who focuses on general shape properties of the mirror, and *Specialist 2* who focuses on reflections from the mirror face.

Specialist 1 had to consider two criteria (Fig. 7):

- *Aspect ratio* (A) is the balance between the height H and the width W of the mirror such as $A = H/W$.
- *Asymmetric balance* (B) is the balance between either left-and-right (B_{LR}) or top-and-bottom (B_{TB}). One of the two asymmetric balances was chosen for the tasks. In both cases, the asymmetric balance was defined from the four corner radii: left-top (LT_R), left-bottom (LB_R), right-top (RT_R) and right-bottom (RB_R).

$$B_{LR} = |LT_R - RT_R| + |LB_R - RB_R| \quad (1)$$

$$B_{TB} = |LT_R - LB_R| + |RT_R - RB_R| \quad (2)$$

Specialist 2 had to consider two criteria (Fig. 7):

- *Rear visibility*: participants had to follow a given guideline that specifies which objects should or should not be visible in the reflective part of the mirror.
- *Surface size* (C) is the geometric area of the reflective surface of the mirror (computed in cm^2).

These criteria are good representatives of design challenges for each role. The *Aspect ratio* and the *Asymmetric balance* represent criteria used by designers to influence the overall appearance, whereas the *Rear visibility* and the *Surface size* are important factors for ergonomists to assess user experience. For A, B and C, participants had to reach a value within a given

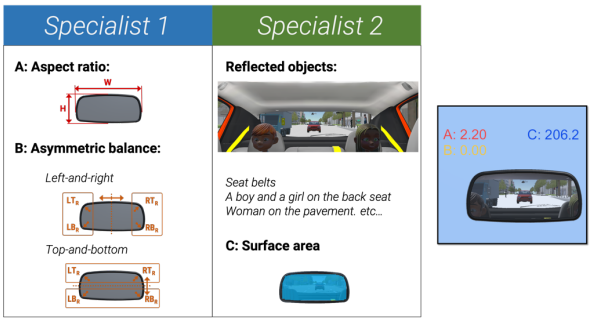

Figure 7. (Left) design criteria of *Specialist 1* and *Specialist 2*. (Right) design criteria values displayed next the mirror.

range. To help them, the current value of each criterion was displayed with different colors next to each design alternative: red for *A*, yellow for *B* and blue for *C* (Fig. 7).

To ensure a proper counterbalancing and avoid bias, we designed two tasks which resulted in different mirror shapes (*Small* and *Large*). We verified through pilot tests that they had similar difficulty and contradictory criteria which require pairs to find a trade-off. The criterion values are *A*: 2.0 – 2.3, *B*: > 50 (top-and-bottom), *C*: 110 – 140 for *Small*, and *A*: 3.1 – 3.5, *B*: > 55 (left-and-right), *C*: 200 – 230 for *Large*.

Once Participants agreed on a design, they saved it with a double tap gesture. They were instructed to finish the task as quickly as possible. We encouraged pairs to communicate together, but strongly forbade them to tell their own design criteria. For example, *Specialist 1* was not allowed to say "*Aspect ratio*" or "*Asymmetric balance*" to express requirements. Instead, they could use shape-related vocabularies, e.g. "*I want to make the mirror higher/wider/smaller/curvier/etc.*".

Hypotheses

We formulate hypotheses based on the usability study and previous work about collaboration on wall-sized displays:

- **H1**: participants find the right design faster with *ShapeCompare* than with *ShapeSlide*;
- **H2**: participants find the right design with fewer iterations with *ShapeCompare* than *ShapeSlide*;
- **H3**: participants prefer *ShapeCompare* for communicating with their partner;
- **H4**: overall, participants prefer *ShapeCompare* to achieve the task.

Method

The experiment is a [2×2] within-participant design with the following factors:

- VISUALIZATION with the two techniques: *ShapeCompare* and *ShapeSlide*;
- TASK with two set of design criteria resulting in different target shapes: *Small* and *Large*.

We first counterbalanced the order of the two VISUALIZATION conditions among pairs, then for each condition, we switched the TASK. To allow all participants to do the two TASKs with both VISUALIZATION conditions, but to avoid them remembering the task, we ran the experiment in two sessions separated from two to three weeks. For example, if participants did *Small* with *ShapeCompare* and *Large* with *ShapeSlide* in the 1st session, they did *Small* with *ShapeSlide* and *Large* with *ShapeCompare* in the 2nd session. Participant roles remained constant across the two sessions.

Participants

We recruited 24 participants, aged 20 to 32 (mean 25.4), with normal or corrected-to-normal vision. Pairs were formed at the time of recruitment leading to 6 male-male, 4 female-male and 2 female-female. 6 participants had previous experience with AutoCAD, 4 with SolidWorks, 2 with CATIA, 3 with other CAD systems, and 9 had with no experience.

Procedure

For each session, participants received written instructions. They filled out demographic questionnaires. They sat in distant places in a room and received design criteria for *Specialist 1* or *Specialist 2*. They could ask questions to the instructor without being heard by the partner. At the beginning of the experiment, each participant had a dedicated training to understand the given design criteria. For example, the instructor asked *Specialist 1* to modify the mirror and reach a specific *Aspect ratio* with two different sizes. During this training, the partner waited in a different room. For each VISUALIZATION condition, pairs also performed a common training to learn the interaction, followed by a measured trial. They filled out a questionnaire after each trial.

Data Collection

We registered 48 trials: 2 VISUALIZATION×2 TASK×12 pairs. We logged the task completion time (*TCT*) and the number of selections (*Selections*). For *TCT*, the instructor gave the starting signal and measurement stopped when pairs agreed on a design or after 30 minutes. *Selections* correspond to the number of iterations performed by the participants during the task. We recorded video. The questionnaire was based on the NASA TLX [21] with additional questions about communication with partner and overall preferences.

Data Analysis

We analyzed the video recording to investigate communication and the use of speech and gestures. We first transcribed participants' discussions. Based on the transcripts, we ignored the utterances which were not relevant to the task and grouped their design-related conversation into 5 categories:

- *Deictic instructions:* participants used deictic gestures to show something on the screen to the partner. This category has two subgroups:
  - *Deictic-specific*: participants indicated a specific shape (*e.g. "I want to modify the mirror like this"*). Most of the time, they used a pointing gesture.
  - *Deictic-range*: participants indicated a range of shapes (*e.g. "...from this shape to this one, it is OK"*).
- *Design expression*: participants expressed ideas either verbally or with gestures (excluding deictic gestures). This category has two subgroups:
  - *Expression-verbal*: participants used shape-related vocabularies (e.g. *"...wider, more curved, etc."*).
  - *Expression-gesture*: participants described the desired shape with hand or finger motions.
- *Magnitude:* participants quantified the size of the modification they wanted (e.g. *"much more..."* or *"a bit less..."*).

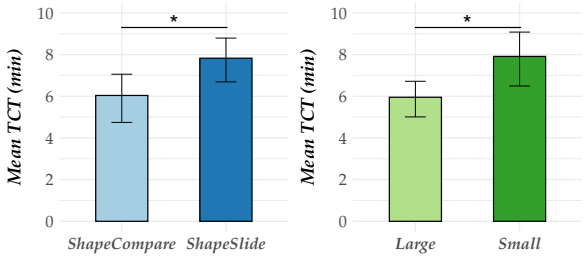

Figure 8. Mean *TCT* by VISUALIZATION (left) and by TASK (right). Error bars show 95% confidence intervals (CI).

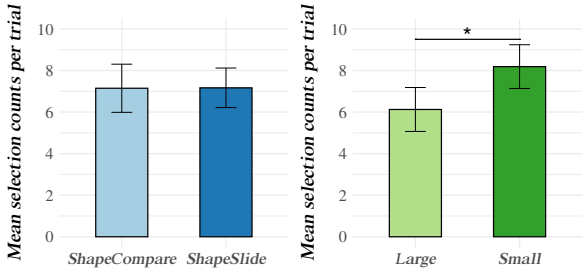

Figure 9. Mean *Selections* by VISUALIZATION (left) and by TASK (right). Error bars show 95% CI.

RESULTS

Task Completion Time (TCT)

We tested *TCT* for normality on the whole data set using a Shapiro-Wilk W test and found that it was not normally distributed[6]. We tested for goodness-of-fit with a log-normal distribution using Kolmogorov's *D*-test, which showed a non-significant result. Therefore, we ran the analysis using the log-transform of *TCT*, as recommended by Robertson & Kaptein [44](p. 316). We did not find any significant learning effects due to technique presentation order.

A repeated measures ANOVA on *TCT* with the model VISUALIZATION×TASK revealed significant effects on VISUALIZATION ($F_{1,47} = 4.83$, $p = 0.033$) and TASK ($F_{1,47} = 4.66$, $p = 0.036$) but no significant interaction effect. Pairs achieved the task faster with *ShapeCompare* (6.04±1.76 min) than with *ShapeSlide* (7.83±1.59 min) (Fig. 8, left). *Large* (5.95±1.49 min) was faster than *Small* (7.91±1.91 min) (Fig. 8, right).

Number of Selections

In conformity with count data, *Selections* did not follow normal or log-normal distribution. Consequently, we computed the mean *Selections* of each participant by levels for each factor and we used non-parametric tests. For VISUALIZATION, a Wilcoxon Signed Rank test did not reveal any significant differences ($p = 0.78$)(Fig. 9, left). For TASK, a Wilcoxon Signed Rank test showed ($p = 0.009$) that *Large* (6.13±1.66) led to fewer *Selections* than *Small* (8.19±1.66) (Fig. 9, right).

[6]All analyses were performed with R and we used a significance level of $\alpha = 0.05$ for all statistical tests.

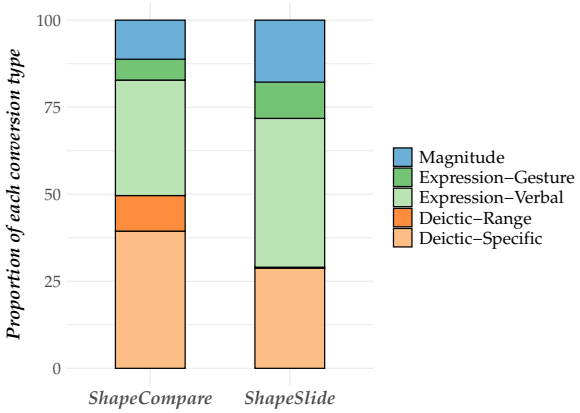

Figure 10. Tagged expression proportions by VISUALIZATION

Conversation Analysis

We analyze the communications between participants during the design exploration tasks. 2591 sentences were tagged for all trials (Fig. 10). We did not find difference in the total number of sentences between *ShapeCompare* (1343) and *ShapeSlide* (1248). The participants used more *Deictic instructions* with *ShapeCompare* (39.4% for *Deictic-Specific*, 10.22% for *Deictic-Range*) than with *ShapeSlide* (28.77% for *Deictic-Specific*, 0.27% for *Deictic-Range*). On the contrary, they used less *Shape-related expression* and *Magnitude* with *ShapeCompare* (33.17% for *Expression-Verbal*, 5.99% for *Expression-Gesture* and 11.22% for *Magnitude*) than with *ShapeSlide* (42.74% for *Expression-Verbal*, 10.42% for *Expression-Gesture* and 17.8% for *Magnitude*).

Qualitative Feedback

In questionnaires, participants graded each VISUALIZATION on a 5-point Likert scale. To avoid confusion, we phrased the questions so that they always had to give a high grade if they appreciated the condition. We asked them if they found the condition *efficient*, *not mentally demanding*, *not physically demanding*, *not difficult to use*, *not frustrating* and *helpful for communication* (1: strongly disagree, 5: strongly agree). They also gave an *overall evaluation* (1: bad, 5: good) for the technique itself and the communication with their partner.

We computed the mean grades of each participant and used a non-parametric Wilcoxon Signed Rank test (Fig. 11). *ShapeCompare* was perceived more helpful for communication (avg. 3.96 vs. 3.13, $p = 0.00014$) and preferred in general (avg. 4.08 vs. 3.25, $p = 0.014$) and for the communication (avg. 4.25 vs. 3.375, $p = 0.004$) in comparison to *ShapeSlide*. We did not find significant differences for the other criteria.

DISCUSSION

ShapeCompare is linked to a *CAD Server* which enables non-CAD experts to easily generate multiple design alternatives of native CAD data. Unlike the conventional design process, all project members can participate in design adjustment tasks. This new capability is complex to evaluate since no comparable systems exist. However, we can draw inspiration from

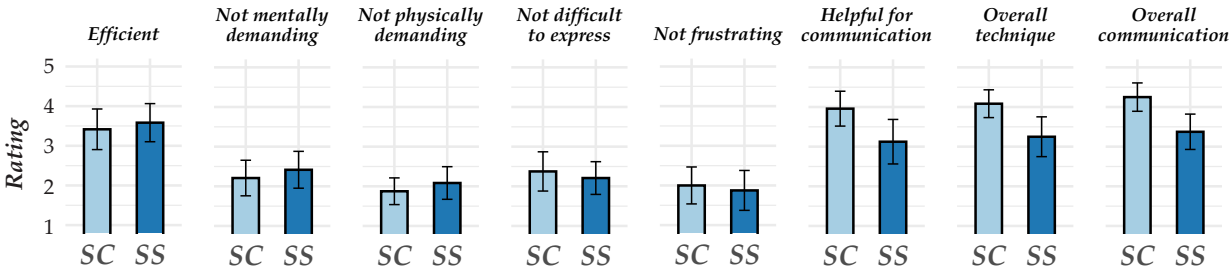

Figure 11. Ratings in questionnaires (5 is best, 1 is worst). SC=*ShapeCompare* and SS=*ShapeSlide*. Error bars show 95% CI.

some evaluation criteria proposed by Olsen [41] to verify the contribution of our system: *"Expressive Match"*—users can interact with the final shapes of the CAD object instead of a 2D sketch and a parameter tree as in a CAD software, and *"Empowers new design participants"*—non-CAD experts can achieve design adjustments which are currently done by CAD engineers in the industrial design process.

ShapeCompare also takes advantage of a wall-sized display to distribute design alternatives across the screen space. To assess the benefit of this visualization technique for collaborative design tasks, we examine the results from the controlled experiment with respect to initial hypotheses. While the number of *Selections* (i.e. iterations) was not significantly different with the two VISUALIZATION conditions, participants found the right design significantly faster with *ShapeCompare* than with *ShapeSlide*. This supports **H1**, but not **H2**. The results from the questionnaires show that *ShapeCompare* was perceived more helpful for communicating with the partner than *ShapeSlide*, and *ShapeCompare* was generally preferred. This supports **H3** and **H4**.

The smaller *TCT* and better communication with *ShapeCompare* could be explained by the large number of *Deictic instructions* used by participants. During the task, the multiple alternatives of *ShapeCompare* were often used as references for communication: participants used the displayed shapes to convey their design idea (e.g. "*I want a mirror like this*"), to show limitations (e.g. "*only shapes between this and this one*") or to ask for partner opinion (e.g. "*what do you think about this one?*"). Whereas with *ShapeSlide*, they needed to describe their requirements verbally or with their hand gestures. The words related to *Magnitude* were also more used with *ShapeSlide* when they instructed their partner acting on a modification (*e.g. "Can you increase it more?"*). According to Clark [12], the multiple alternatives of *ShapeCompare* create a common ground between pairs and minimize the communication costs, which can explain the smaller *TCT*.

The same iteration numbers in both VISUALIZATION could be due to the task design and the fact that these iterations may be necessary to reach the right design. In addition, we made two different sets of design criteria (*Small* and *Large*). Even if we tried to make these criteria as equivalent as possible, it seems that LARGE was easier than SMALL.

We were concerned that displaying lots of alternatives with *ShapeCompare* could increase the cognitive load. However, we did not find significant difference in the NASA TLX, which suggests *ShapeCompare* do not overload participants.

Although most participants preferred *ShapeCompare* for the ease of communication, some others prefer *ShapeSlide* in terms of interaction: e.g. "*ShapeSlide is interesting because it allows to have instant feedback and to cycle through all possible design while swiping*". It seems that these users prefer to have more initiatives on the design activity, instead of dealing with solutions given by the system. *ShapeCompare* is a first prototype, but additional functionalities would be required to allow the user to feel more in control of the alternative generation. In particular, it would be important that users can define the minimum, maximum and step size of the generated alternatives. This would allow them to achieve fine or coarse modifications. Moreover, with the current implementation, the generation of new alternatives can take up to 5 seconds. A solution to improve this could be to use several *CAD server*s to parallelize the computation of 3D meshes.

Finally, we believe that our system can be extended to any CAD objects as long as they require collaboration among multidisciplinary experts. For large objects, the number of displayed alternatives or the scaling could be adjusted.

DESIGN RECOMMENDATIONS

While we studied alternative exploration in a specific context, the approach of visualizing "small multiples" on a wall-sized display could be extended to other contexts as soon as parameter variations are involved. For example, it can be suitable for *generative design* in which users can specify preferred designs to the AI, or physical simulations such as weather predictions in which users can run several simulations with different parameter variations. Based on the observations of the usability study and the results of the controlled experiment, we draw some generic recommendations which can be applied to other contexts:

- A large number of alternatives can be displayed on the wall-sized display without overloading the users.
- Allowing users to generate/compare alternatives can help them to solve constraints and reach a trade-off.
- Users need to understand the effect of all possible modifications before generating new alternatives to facilitate exploration. One option could be to display previews of the most extreme modifications.

- Users need to keep track of the design history and the link between each alternative selection to understand the design evolution.
- The difference between side-by-side alternatives should be big enough to be perceived by users. An automatic solution to tune parameter steps can be valuable.

CONCLUSION AND FUTURE WORK

This paper investigates collaborative design exploration on a wall-sized display by proposing *ShapeCompare*, a system which enables users to generate and distribute design alternatives of a CAD model on large screens. *ShapeCompare* relies on "small multiple" representations of a CAD object to allow multidisciplinary teams to collaboratively explore, compare alternatives and reflect their ideas on a wall-sized display. It aims to reduce the iterations of industrial design review processes by avoiding miscommunication between experts and engineers in charge of the CAD data modification.

We ran a controlled experiment to assess how simultaneous visualization of multiple design alternatives affects the collaboration among experts during a constraint-solving task. We compared *ShapeCompare* with another technique named *ShapeSlide* which shows only one design alternative at a time, but enable users to quickly switch the displayed alternative. *ShapeSlide* could be used on any standard screens since it does not require a large screen space. The results showed that participants reach a consensus respecting the design constraints significantly faster with *ShapeCompare* than with *ShapeSlide*. We also found that participants used more deictic instructions and less verbal or gesture-based design expressions with *ShapeCompare* than with *ShapeSlide*. It suggests that the multiple alternative visualization helps collaborators during design exploration and negotiation by increasing the common grounds among them. To our knowledge, it is the first study to demonstrate the benefit of the "small multiple" concept on a wall-sized display.

The current system is still a research prototype. There is a lot of space to explore and improve the way to visualized alternatives and interact with them. For example, ways to classify or to merge relevant design alternatives should be investigated. In addition, the proposed approach could also be applied in many other contexts, such as generative design or physical simulation, for which many alternatives can be generated by intuitively varying parameters.

ACKNOWLEDGMENTS

This work was partially supported by European Research Council (ERC) grant n° 695464 "ONE: Unified Principles of Interaction"; and by Agence Nationale de la Recherche (ANR) grants ANR-10-EQPX-26-01 "EquipEx DIGISCOPE" and ANR-11-LABEX-0045-DIGICOSME "Labex Digicosme" as part of the program "Investissement d'Avenir" ANR-11-IDEX-0003-02 "Idex Paris-Saclay".

REFERENCES

[1] Bruno R. De Araújo, Géry Casiez, Joaquim A. Jorge, and Martin Hachet. 2013. Mockup Builder: 3D modeling on and above the surface. *Computers & Graphics* 37, 3 (2013), 165 – 178. DOI: http://dx.doi.org/10.1016/j.cag.2012.12.005

[2] Ravin Balakrishnan, George Fitzmaurice, Gordon Kurtenbach, and William Buxton. 1999. Digital Tape Drawing. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (UIST '99)*. ACM, New York, NY, USA, 161–169. DOI: http://dx.doi.org/10.1145/320719.322598

[3] Robert Ball and Chris North. 2008. The Effects of Peripheral Vision and Physical Navigation on Large Scale Visualization. In *Proceedings of Graphics Interface 2008 (GI '08)*. Canadian Information Processing Society, Toronto, Ont., Canada, 9–16. http://dl.acm.org/citation.cfm?id=1375714.1375717

[4] Robert Ball, Chris North, Chris North, and Doug A. Bowman. 2007. Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 191–200. DOI: http://dx.doi.org/10.1145/1240624.1240656

[5] Michel Beaudouin-Lafon. 2011. Lessons Learned from the WILD Room, a Multisurface Interactive Environment. In *Proceedings of the 23rd Conference on L'Interaction Homme-Machine (IHM '11)*. ACM, New York, NY, USA, Article 18, 8 pages. DOI: http://dx.doi.org/10.1145/2044354.2044376

[6] M. Beaudouin-Lafon, S. Huot, M. Nancel, W. Mackay, E. Pietriga, R. Primet, J. Wagner, O. Chapuis, C. Pillias, J. Eagan, T. Gjerlufsen, and C. Klokmose. 2012. Multisurface Interaction in the WILD Room. *Computer* 45, 4 (April 2012), 48–56. DOI: http://dx.doi.org/10.1109/MC.2012.110

[7] Julien Berta. 1999. Integrating VR and CAD. *IEEE Computer Graphics and Applications* 19, 5 (Sep. 1999), 14–19. DOI:http://dx.doi.org/10.1109/38.788793

[8] Xiaojun Bi and Ravin Balakrishnan. 2009. Comparing Usage of a Large High-resolution Display to Single or Dual Desktop Displays for Daily Work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1005–1014. DOI: http://dx.doi.org/10.1145/1518701.1518855

[9] Yannick Bodein, Bertrand Rose, and Emmanuel Caillaud. 2013. A roadmap for parametric CAD efficiency in the automotive industry. *Computer-Aided Design* 45, 10 (2013), 1198 – 1214. DOI: http://dx.doi.org/10.1016/j.cad.2013.05.006

[10] William Buxton, George Fitzmaurice, Ravin Balakrishnan, and Gordon Kurtenbach. 2000. Large displays in automotive design. *IEEE Computer Graphics and Applications* 20, 4 (July 2000), 68–75. DOI: http://dx.doi.org/10.1109/38.851753

[11] SangSu Choi, HyunJei Jo, Stefan Boehm, and Sang Do Noh. 2010. ONESVIEW: An Integrated System for One-Stop Virtual Design Review. *Concurrent Engineering* 18, 1 (2010), 75–91. DOI: http://dx.doi.org/10.1177/1063293X10361624

[12] Herbert H Clark, Susan E Brennan, and others. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991 (1991), 127–149.

[13] Dane Coffey, Chi-Lun Lin, Arthur G Erdman, and Daniel F Keefe. 2013. Design by Dragging: An Interface for Creative Forward and Inverse Design with Simulation Ensembles. *IEEE Trans. on Visualization and Computer Graphics* 19, 12 (Dec 2013), 2783–2791. DOI:http://dx.doi.org/10.1109/TVCG.2013.147

[14] Thomas Convard and Patrick Bourdot. 2004. History Based Reactive Objects for Immersive CAD. In *Proceedings of the Ninth ACM Symposium on Solid Modeling and Applications (SM '04)*. Eurographics Association, Aire-la-Ville, Switzerland, 291–296. http://dl.acm.org/citation.cfm?id=1217875.1217921

[15] Mary Czerwinski, Greg Smith, Tim Regan, Brian Meyers, George Robertson, and Gary Starkweather. 2003. Toward Characterizing the Productivity Benefits of Very Large Displays. In *Interact 2003*. IOS Press.

[16] Jakub Dostal, Uta Hinrichs, Per Ola Kristensson, and Aaron Quigley. 2014. SpiderEyes: Designing Attention- and Proximity-aware Collaborative Interfaces for Wall-sized Displays. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. ACM, New York, NY, USA, 143–152. DOI: http://dx.doi.org/10.1145/2557500.2557541

[17] Michele Fiorentino, Raffaele de Amicis, Giuseppe Monno, and Andre Stork. 2002. Spacedesign: a mixed reality workspace for aesthetic industrial design. In *Proceedings. International Symposium on Mixed and Augmented Reality*. 86–318. DOI: http://dx.doi.org/10.1109/ISMAR.2002.1115077

[18] T. Fleisch, G. Brunetti, P. Santos, and A. Stork. 2004. Stroke-input methods for immersive styling environments. In *Proceedings Shape Modeling Applications, 2004.* 275–283. DOI: http://dx.doi.org/10.1109/SMI.2004.1314514

[19] Clifton Forlines, Daniel Vogel, and Ravin Balakrishnan. 2006. HybridPointing: Fluid Switching Between Absolute and Relative Pointing with a Direct Input Device. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)*. ACM, New York, NY, USA, 211–220. DOI: http://dx.doi.org/10.1145/1166253.1166286

[20] Tovi Grossman, Ravin Balakrishnan, Gordon Kurtenbach, George Fitzmaurice, Azam Khan, and Bill Buxton. 2001. Interaction Techniques for 3D Modeling on Large Displays. In *Proceedings of the 2001 Symposium on Interactive 3D Graphics (I3D '01)*. ACM, New York, NY, USA, 17–23. DOI: http://dx.doi.org/10.1145/364338.364341

[21] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*. Advances in Psychology, Vol. 52. North-Holland, 139 – 183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[22] Takeo Igarashi, Takeo Igarashi, and John F. Hughes. 2007. A Suggestive Interface for 3D Drawing. In *ACM SIGGRAPH 2007 Courses (SIGGRAPH '07)*. ACM, New York, NY, USA, Article 20. DOI: http://dx.doi.org/10.1145/1281500.1281531

[23] J.H. Israel, E. Wiese, M. Mateescu, C. Zöllner, and R. Stark. 2009. Investigating three-dimensional sketching for early conceptual design—Results from expert discussions and user studies. *Computers & Graphics* 33, 4 (2009), 462 – 473. DOI: http://dx.doi.org/10.1016/j.cag.2009.05.005

[24] Mikkel R. Jakobsen and Kasper Hornbæk. 2014. Up Close and Personal: Collaborative Work on a High-resolution Multitouch Wall Display. *ACM Trans. Comput.-Hum. Interact.* 21, 2, Article 11 (Feb. 2014), 34 pages. DOI:http://dx.doi.org/10.1145/2576099

[25] Daniel F Keefe, Robert C Zeleznik, and David H Laidlaw. 2007. Drawing on Air: Input Techniques for Controlled 3D Line Illustration. *IEEE Transactions on Visualization and Computer Graphics* 13, 5 (Sep. 2007), 1067–1081. DOI: http://dx.doi.org/10.1109/TVCG.2007.1060

[26] Azam Khan, Justin Matejka, George Fitzmaurice, and Gordon Kurtenbach. 2005. Spotlight: Directing Users' Attention on Large Displays. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 791–798. DOI: http://dx.doi.org/10.1145/1054972.1055082

[27] Sebastien Kuntz. 2017. *improov*^{3}, The VR collaborative meeting room, MiddleVR company, Paris, FRANCE, http://www.improovr.com/home/. (2017). http://www.improovr.com/home/

[28] R. Langner, U. Kister, and R. Dachselt. 2019. Multiple Coordinated Views at Large Displays for Multiple Users: Empirical Findings on User Behavior, Movements, and Distances. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan 2019), 608–618. DOI: http://dx.doi.org/10.1109/TVCG.2018.2865235

[29] Glyn Lawson, Davide Salanitri, and Brian Waterfield. 2015. VR Processes in the Automotive Industry. In *Human-Computer Interaction: Users and Contexts*, Masaaki Kurosu (Ed.). Springer International Publishing, 208–217. DOI: http://dx.doi.org/10.1007/978-3-319-21006-3_21

[30] Valerie D Lehner and Thomas A DeFanti. 1997. Distributed virtual reality: supporting remote collaboration in vehicle design. *IEEE Computer Graphics and Applications* 17, 2 (March 1997), 13–17. DOI:http://dx.doi.org/10.1109/38.574654

[31] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. 2016. Shared Interaction on a Wall-Sized Display in a Data Manipulation Task. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2075–2086. DOI: http://dx.doi.org/10.1145/2858036.2858039

[32] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. 2017. CoReach: Cooperative Gestures for Data Manipulation on Wall-sized Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6730–6741. DOI: http://dx.doi.org/10.1145/3025453.3025594

[33] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E. Mackay. 2014. Effects of Display Size and Navigation Type on a Classification Task. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4147–4156. DOI:http://dx.doi.org/10.1145/2556288.2557020

[34] Jean-Baptiste Louvet and Cédric Fleury. 2016. Combining Bimanual Interaction and Teleportation for 3D Manipulation on Multi-touch Wall-sized Displays. In *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology (VRST '16)*. ACM, New York, NY, USA, 283–292. DOI: http://dx.doi.org/10.1145/2993369.2993390

[35] Morad Mahdjoub, Davy Monticolo, Samuel Gomes, and Jean-Claude Sagot. 2010. A collaborative Design for Usability approach supported by Virtual Reality and a Multi-Agent System embedded in a PLM environment. *Computer-Aided Design* 42, 5 (2010), 402 – 413. DOI: http://dx.doi.org/10.1016/j.cad.2009.02.009

[36] Pierre Martin, Stéphane Masfrand, Yujiro Okuya, and Patrick Bourdot. 2017. A VR-CAD Data Model for Immersive Design. In *Augmented Reality, Virtual Reality, and Computer Graphics*. Springer International Publishing, Cham, 222–241. DOI: http://dx.doi.org/10.1007/978-3-319-60922-5_17

[37] T.S. Mujber, T. Szecsi, and M.S.J. Hashmi. 2004. Virtual reality applications in manufacturing process simulation. *Journal of Materials Processing Technology* 155-156 (2004), 1834 – 1838. DOI: http://dx.doi.org/10.1016/j.jmatprotec.2004.04.401

[38] Mathieu Nancel, Olivier Chapuis, Emmanuel Pietriga, Xing-Dong Yang, Pourang P. Irani, and Michel Beaudouin-Lafon. 2013. High-precision Pointing on Large Wall Displays Using Small Handheld Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 831–840. DOI: http://dx.doi.org/10.1145/2470654.2470773

[39] Yujiro Okuya, Nicolas Ladeveze, Cédric Fleury, and Patrick Bourdot. 2018a. ShapeGuide: Shape-Based 3D Interaction for Parameter Modification of Native CAD Data. *Frontiers in Robotics and AI* 5 (2018), 118. DOI: http://dx.doi.org/10.3389/frobt.2018.00118

[40] Yujiro Okuya, Nicolas Ladeveze, Olivier Gladin, Cédric Fleury, and Patrick Bourdot. 2018b. Distributed Architecture for Remote Collaborative Modification of Parametric CAD Data. In *2018 IEEE Fourth VR International Workshop on Collaborative Virtual Environments (3DCVE)*. 1–4. DOI: http://dx.doi.org/10.1109/3DCVE.2018.8637112

[41] Dan R. Olsen. 2007. Evaluating User Interface Systems Research. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. Association for Computing Machinery, New York, NY, USA, 251–258. DOI: http://dx.doi.org/10.1145/1294211.1294256

[42] D. Rantzau, F. Maurer, C. Mayer, R. Löffler, O. Riedel, H. Scharm, and D. Banek. 2001. The integration of immersive Virtual Reality applications into Catia V5. In *Immersive Projection Technology and Virtual Environments 2001*. Springer Vienna, Vienna, 93–102. DOI:http://dx.doi.org/10.1007/978-3-7091-6221-7_10

[43] Alberto Raposo, Ismael Santos, Luciano Soares, Gustavo Wagner, Eduardo Corseuil, and Marcelo Gattass. 2009. Environ: Integrating VR and CAD in Engineering Projects. *IEEE Computer Graphics and Applications* 29, 6 (Nov 2009), 91–95. DOI: http://dx.doi.org/10.1109/MCG.2009.118

[44] Judy Robertson and Maurits Kaptein. 2016. *Modern Statistical Methods for HCI*. Springer. DOI: http://dx.doi.org/10.1007/978-3-319-26633-6

[45] Drew Skillman and Patrick Hackett. 2016. *Tilt Brush application, Google Inc., https://www.tiltbrush.com/*. https://www.tiltbrush.com/

[46] TechViz. 2004. TechViz XL. (2004). https://www.techviz.net/techviz-xl

[47] Daniel Vogel and Ravin Balakrishnan. 2005. Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays. In *Proc. of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. ACM, New York, NY, USA, 33–42. DOI: http://dx.doi.org/10.1145/1095034.1095041

[48] Beth Yost, Yonca Haciahmetoglu, Chris North, and Chris North. 2007. Beyond Visual Acuity: The Perceptual Scalability of Information Visualizations for Large Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 101–110. DOI: http://dx.doi.org/10.1145/1240624.1240639

[49] Peter Zimmermann. 2008. *Virtual Reality Aided Design. A survey of the use of VR in automotive industry*. Springer Netherlands, Dordrecht, 277–296. DOI: http://dx.doi.org/10.1007/978-1-4020-8200-9_13

Virtual Navigation considering User Workspace: Automatic and Manual Positioning before Teleportation

Yiran Zhang [1,2], Nicolas Ladevèze [1], Huyen Nguyen [1], Cédric Fleury [2] and Patrick Bourdot [1]
[1]Université Paris-Saclay, CNRS, LIMSI, VENISE team, Orsay, France
[2]Université Paris-Saclay, CNRS, Inria, LRI, Orsay, France
{yiran.zhang,nicolas.ladeveze,huyen.nguyen,patrick.bourdot}@limsi.fr,cedric.fleury@lri.fr

ABSTRACT

Teleportation is a navigation technique widely used in virtual reality applications using head-mounted displays. Basic teleportation usually moves a user's viewpoint to a new destination of the virtual environment without taking into account the physical space surrounding them. However, considering the user's real workspace is crucial for preventing them from reaching its limits and thus managing direct access to multiple virtual objects. In this paper, we propose to display a virtual representation of the user's real workspace before the teleportation, and compare manual and automatic techniques for positioning such a virtual workspace. For manual positioning, the user adjusts the position and orientation of their future virtual workspace. A first controlled experiment compared exocentric and egocentric manipulation techniques with different virtual workspace representations, including or not an avatar at the user's future destination. Although exocentric and egocentric techniques result in a similar level of performance, representations with an avatar help the user to understand better how they will land after teleportation. For automatic positioning, the user selects their future virtual workspace among relevant options generated at runtime. A second controlled experiment shows that the manual technique selected from the first experiment and the automatic technique are more efficient than the basic teleportation. Besides, the manual technique seems to be more suitable for crowded scenes than the automatic one.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; Interaction techniques; User studies.

KEYWORDS

Locomotion; teleportation; real workspace; virtual workspace; virtual object access; spatial awareness.

ACM Reference Format:
Yiran Zhang [1,2], Nicolas Ladevèze [1], Huyen Nguyen [1], Cédric Fleury [2] and Patrick Bourdot [1]. 2020. Virtual Navigation considering User Workspace: Automatic and Manual Positioning before Teleportation. In *26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), November 1–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3385956.3418949

VRST '20, November 1–4, 2020, Virtual Event, Canada

1 INTRODUCTION

Teleportation is a popular locomotion technique that allows a user to move beyond the limits of their available physical space while minimizing simulator sickness [26, 48]. Using such technique, the user can select a destination point and instantaneously appear at this new location in the virtual environment (VE). However, this technique usually does not consider the physical space surrounding the user and their position inside this space. Due to a lack of awareness of the real workspace boundaries, the user may quickly reach the limits of their workspace while performing a virtual task.

Common solutions, such as alerting the user when they approach the boundaries, increase the user's mistrust in virtual reality (VR) systems, and often break their immersion and sense of presence. These solutions also induce the user to stay still and perform many small teleportations, instead of using their real movements to reach objects of the VE, especially when they get stuck in a corner or against a boundary of their real workspace. On the other hand, real walking could be highly beneficial for improving immersion. Solutions like redirected walking [35] provide a compelling approach for the user to explore large VEs while overcoming the constraints of the real workspace. However, the redirected walking algorithms usually require physical spaces larger than 6m × 6m [6]. It may not be possible for common users with head-mounted displays (HMD) because their real workspace is also limited by the room size.

This work aims to help the user to gain a prior knowledge of the accessible area of the VE, which allows them to access multiple virtual objects with real walking and avoid reaching the real workspace limits. To achieve this goal, we propose to display and manipulate a virtual representation of the real workspace when using the teleportation technique. This virtual representation of the real workspace, also called *virtual workspace* in this paper, is similar to the concepts of *vehicle* [11] or *stage* [20].

Optimizing future access of the user to multiple objects within their physically accessible area could be useful in many scenarios. For example, in VR escape room games [1, 2], if the user can position their virtual workspace close to some area of interest where the clues are possibly hidden, they can fully explore this area by walking and thus avoid a lot of unnecessary teleportations. With first-person shooter games [5], by choosing an appropriate virtual workspace position, they can physically move to hide and attack enemies. Another example is some complex VR training system [4] that consists of several assembly tasks at different locations in the VE. Positioning the virtual workspace around an assembly area of each sub-task might help the user to focus better on knowledge acquisition and assembly procedure learning as they do not need to manage virtual objects' accessibility while completing this sub-task.

In order to help the user to define the position and orientation of this virtual workspace, we explore two strategies, named manual positioning and automatic positioning. The former allows the user to manually adjust the position and orientation of their future virtual workspace. The latter, using clustering techniques, automatically generates a series of possible virtual workspaces considering the interactive object layout in the VE. The user can select their future virtual workspace among the relevant options proposed by the system. We investigated these different positioning techniques in two controlled experiments. In the first one (pre-study), we assessed three manual positioning techniques and selected the most appropriate one. In the second one, we compared the selected manual positioning and the automatic positioning techniques to the basic teleportation technique. We evaluated the user performance in terms of efficiency, number of teleportations, and cognitive load, considering various virtual object layouts. From the results, we derived some usability guidelines for such techniques.

The paper is organized as follows. Section 2 reviews related work about teleportation and navigation techniques that take the user's real workspace into account. Section 3 details the two positioning strategies. Section 4 describes the two experiments and analyze the results. Finally, section 5 concludes by proposing some guidelines and discussing open problems of this contribution.

2 RELATED WORK

The mapping between the real and virtual world is a fundamental issue of every VR applications, and various previous works explore solutions to manage this relationship. To avoid the user's collisions with the real world and grant direct access to virtual objects, some applications choose to have a fixed one-to-one mapping between the real and virtual environments. For example, Cheng et al. [17] and Sra et al. [41] propose to procedurally generate the virtual environment based on a 3D scan of the real world with a depth camera. Consequently, the size and shape of the virtual environment are constrained. Redirected walking [35] or other view distortion techniques [45] can be used to map a large virtual environment to a small real workspace while allowing the user to walk freely. Impossible Space [44] also uses a self-overlapping architectural layout to allow the user to walk through multiple virtual rooms while staying in the same real room. However, these solutions are not suitable for all applications since they require a reasonably large real workspace, and physical walking could also be tiresome when the user has to travel long distances.

Virtual navigation is a generic solution which allows the user to go beyond the real workspace limits [12]. However, it breaks the one-to-one mapping between the real and virtual environment. Such mismatch may result in safety issues as the user could collide with physical obstacles that are invisible in the virtual world. The user may feel afraid of encountering real-world obstacles [18] and thus alter their movement behaviors [18, 39]. In such cases, it is crucial to provide the user with a way to visualize and understand the limits of their real workspace. For example, 3DM [14] not only allows the user to walk naturally on a magic carpet representing the tracking space, but also to move this magic carpet over a long-distance using a steering technique. Magic Barrier Tape [19] employs a virtual barrier tape to indicate the available walking area to the user. The user can go beyond the boundaries by "pushing" the tape. More recently, Chen et al. [15] present a human-joystick technique that takes the real workspace boundaries into account to prevent collisions during navigation.

Teleportation is another virtual-navigation technique widely used in VR applications. Basic teleportation allows the user to instantly appear at a remote target position using a pointing technique [12]. The instantaneous transition of viewpoint avoids the sensory conflict between the visual feedback and the user's vestibular systems, which reduces simulator sickness compared to other locomotion techniques [26, 48]. However, teleportation lacks optical flow, which limits the user's ability to perform path finding and leads to disorientation [7, 9]. Several existing approaches aim to improve teleportation. Point and Teleport technique [13, 23] allows the user to specify their orientation before the teleportation. Jumper [10] employs the user's eye gaze to specify the target destination and thus enables hand-free teleportation. Dash [9] quickly but continuously displaces the user's viewpoint to retain optical flow cues. The out-of-body locomotion technique [25] allows the user to seamlessly switch between a first-person and a third-person view to reduce the confusion caused by discontinuous avatar movements for multiple-user teleportation.

Some approaches combine teleportation with real walking to better use the available real workspace and facilitate real walking. For example, Redirected Teleportation [29] requires the user to step into a portal to activate teleportation, which unobtrusively reorients and re-positions the user away from the tracking space boundary. Interactive Portals [21] reorient the user to a safe position via portals sliding up from the ground in the center of the CAVE. Switch techniques [49] help the user to recover a one-to-one mapping between real and virtual workspace inside some areas in the VE. The user can access the virtual objects of the areas by walking. However, these areas need to be predefined according to the layout of VE and the shape of the real workspace, which makes it hard to apply in a generic context. Apart from that, the SteamVR plugin for Unity [3] can provide a virtual representation of the real workspace boundaries before the teleportation. However, the user cannot manipulate the orientation of this virtual representation.

In this paper, we overcome the limits of the existing approaches by designing two types of positioning techniques (manual vs. automatic) that allow the user to manage their virtual workspace before each teleportation, and to develop strategies to access multiple objects by real walking. These techniques do not rely on specific virtual object layouts or prior knowledge of the user's real workspace.

3 POSITIONING TECHNIQUES

In this section, we will present our considerations and design for the virtual workspace positioning techniques.

3.1 Manual Techniques

The manual positioning techniques use a 3D volume representing the user's real workspace, and the user can directly control its position and orientation in the VE to customize the teleportation.

Interacting with a predefined volume has been used for multiple-object selection (MOS) [31, 42] to select the enclosed objects at once. The user can use manipulation techniques, such as the go-go

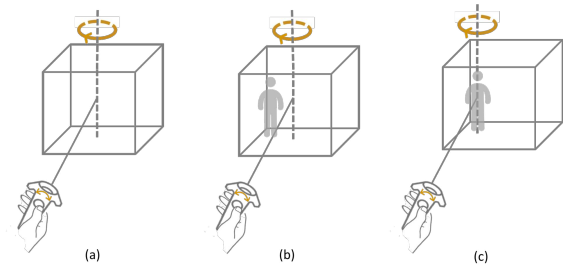

Figure 1: Three manual positioning techniques: the user (a) rotates the volume around its central vertical axis; (b) rotates the 3D volume and an avatar around the central axis; and (c) rotates the 3D volume around the avatar's vertical axis.

technique [31], to manipulate this volume remotely to select distant objects. In our approach, we extended the ray casting technique [33] to enable the virtual workspace manipulation.

In our first manual technique, the 3D volume of the virtual workspace appears when the user presses the touch pad of the controller. The intersection point between the virtual ray and the virtual ground determines the future position of this volume. The user can rotate the volume around its vertical axis (see Figure 1(a)) by sliding the finger in a circle on the touch pad with a one-to-one mapping. The objects fully or partially enclosed by the volume are selected and highlighted with a more intense colour. The user can release the touch pad to end the manipulation, and then they will be teleported into the newly specified virtual workspace with a correct matching to their real workspace.

To enhance the spatial awareness [12], we propose two other techniques, which add additional visual information to the virtual workspace representation: an avatar showing directly how the user will land after the teleportation. This avatar is a ghost representation of the user at their future location, which helps them to get self-related information.

These techniques differ in their rotation axis position. One technique uses an exocentric manipulation by rotating the 3D volume and the avatar around the volume's central vertical axis (see Figure 1(b)). The exocentric information can help people to see global trends [8] and enhances the size judgment [32]. The other technique uses an egocentric manipulation by rotating the 3D volume around the avatar's vertical axis (see Figure 1(c)). The egocentric cues can help people to gather the self-related information and result in more accurate distance estimation [30].

3.2 Automatic Technique

In many VR applications, the virtual objects that the user can directly interact with are usually predefined in the scenarios. Based on the layout of these objects and considering the user's actual real workspace, the system can compute possible virtual workspaces and propose them to the user. The user can then select their future virtual workspace among relevant options depending on the task requirements. Our approach to compute a set of suitable virtual workspaces is to: (i) organize objects into clusters by grouping or splitting them; (ii) compute bounding volumes for each cluster; and (iii) repeats the above steps until the size of the bounding volume and the size of user's real workspace become equivalent.

In the first step of the approach, bottom-up or top-down algorithms can be used to cluster virtual objects. The bottom-up algorithms treat each virtual object as a singleton cluster at the beginning, and then successively merge pairs of clusters until the user's real workspace can no longer enclose the bounding volume of a cluster. The top-down algorithms start with a cluster that includes all virtual objects and split the cluster recursively until each sub-cluster is smaller than the user's real workspace. Different criteria can be applied to organize objects into a cluster. For example, K-means is a fast and straightforward heuristic to group pairs of clusters based on their nearest mean [28]. Beyond those geometrical approaches, one can also use semantic knowledge about the scene to align objects of the same kind [43], or use collisions to define clusters when simulating the real-world behaviors [34].

Algorithm 1 Virtual workspaces positioning algorithm

Input: ObjList, UsrRWS
Output: PosList

```
1:  procedure TopDown(ObjList, UsrRWS)
2:      N ← ObjList.length
3:      if N = 1 then
4:          PosList.Add(AABB(ObjList).position)
5:          return PosList
6:      else
7:          bbox ← AABB(ObjList)
8:          if !EncapTest(bbox, UserRealWS) then
9:              lN, rN ← Kmeans(ObjList, 2)
10:             PosList.append(TopDown(lN, UsrRWS))
11:             PosList.append(topDown(rN, UsrRWS))
12:         else
13:             return PosList
```

As a first prototype, we implemented a simple top-down algorithm (see Algorithm 1) to compute the possible virtual workspace positions in sublinear time. It uses K-means (with k=2) to split the inputting objects (ObjList) into disjoint subsets based on their 2D positions (x, z), and generates a bounding volume around each subset using axis-aligned bounding box (AABB) approach [24]. The recursive splitting creates a binary tree, and processes until a subset either contains only one virtual object, or its bounding volumes can be encapsulated in the user's real workspace (UsrRWS). By traversing the binary tree, the algorithm creates and returns a list of bounding volume positions from the leaf nodes. Based on this list, the system can subsequently instantiate 3D volumes at each bounding volume position to represent the possible virtual workspaces. In the example of Figure 2, our algorithm provides five virtual workspaces for a given configuration of the VE and a 3m × 3m user's real workspace. This algorithm is a first implementation to test the related interaction technique and user acceptability of an automatic technique. It can be improved later by considering arbitrary-oriented bounding boxes (OBB) or other clustering methods.

The proposed virtual workspaces are normally invisible to the user. As soon as the user's virtual ray collides with a virtual workspace, it is displayed along with an avatar indicating the user's future destination (see Figure 3(a)). When the user walks into an overlapping

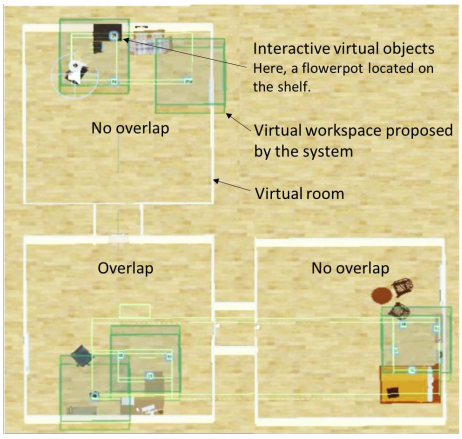

Figure 2: Results of the positioning algorithm: five virtual workspaces (green) are proposed to the user for 12 interactive objects (light blue) located in three virtual rooms (white). Three workspaces are disjointed (no overlap rooms), and two overlap each other (overlap room).

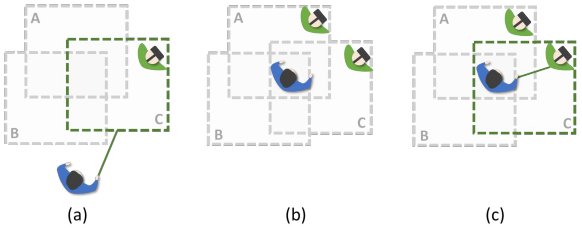

Figure 3: Top view of the user (a) selects a virtual workspace among relevant options; (b) enters an overlapping area and the system displays avatars to represent the connected virtual workspaces; and (c) selects a subsequent workspace by pointing the virtual ray to its corresponding avatar.

area between multiple virtual workspaces, the system allows the user to switch directly from their current virtual workspace to one of the connected workspaces by showing their related avatars (see Figure 3 (b)). The user can then select a new workspace by pointing the virtual ray to its associated avatar (see Figure 3 (c)). The virtual objects enclosed inside the selected workspace are highlighted with a more intense color.

4 EXPERIMENT 1

As a first step, we conducted a controlled experiment to evaluate the three manual techniques proposed in section 3.1. We aimed to assess the benefits of including an avatar at the user's future destination in the virtual workspace. We also wanted to compare egocentric and exocentric manipulation techniques in terms of user spatial awareness and performance. The experiment thus compared the following TECHNIQUES (see Figure 4):

- *Exo-without-avatar* is an exocentric technique allowing the user to move the virtual workspace representation and to rotate it around its central axis. No preview of the user's future position is offered.

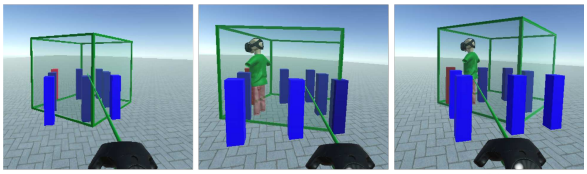

Figure 4: Conditions of the first experiment: (a) *Exo-without-avatar*, (b) *Exo-with-avatar* and (c) *Ego-with-avatar*.

- *Exo-with-avatar* is an exocentric technique for which the rotation axis is still at the center of the virtual workspace representation. An avatar is included to show the user's future position. This avatar is a simplified human body wearing a head-mounted display.
- *Ego-with-avatar* is an egocentric technique that uses the future position of the user (i.e. the avatar position) as the rotation axis of the virtual workspace representation. The same simplified avatar is used in this condition.

We did not include an egocentric technique without the avatar because the rotation axis is invisible, making it difficult for the user to understand the manipulation. The experiment was a within-subject design with TECHNIQUE as a factor. The order of the TECHNIQUES was counterbalanced across participants using a balanced Latin square.

4.1 Hypothesis

We expected the conditions with the avatar would help the user to anticipate their next position in the virtual scene. We also assumed that *Exo-with-avatar* would highlight the entire virtual workspace and would make it easier to enclose the targeted virtual objects, while *Ego-with-avatar* focuses on the user's future destination and causes less disorientation. Therefore we formulated the following hypotheses:

- **H1** *Exo-with-avatar* and *Ego-with-avatar* will reduce disorientation and the time required to find a target after the teleportation, compared to *Exo-without-avatar*.
- **H2** Less time will be required for positioning the virtual workspace representation with *Exo-with-avatar* than with *Ego-with-avatar*.
- **H3** Less time will be required to find a target after the teleportation with *Ego-with-avatar* than with *Exo-with-avatar*.

4.2 Participants

We recruited 12 participants, aged between 25 and 31 (6 men and 6 women). Only one person was left-handed. Three participants had VR experience. 11 out of 12 rated their everyday usage of head-mounted displays as very low.

4.3 Experiment setup

The VR setup consisted of an HTC Vive Pro Eye with both position and orientation tracking, as well as integrated eye-tracking technology. The virtual environment was rendered using Unity with a resolution of 1440 × 1600 pixels per eye at 90 Hz. The experiment room supported a 3m × 3m tracking area. User input was detected using a Vive handheld controller.

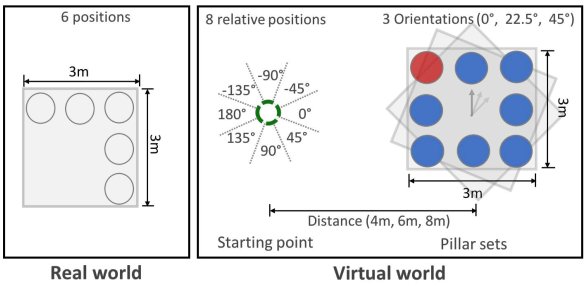

Figure 5: Each starting position in the virtual scene positioned the participant at one of the 6 positions in the real space (left). The pillar set was located with 3 different orientations, in one of the 8 directions around the participant, and at one of the 3 distances from the participant (right).

4.4 Experimental task

To assess spatial awareness, existing studies [12, 47] measure the time needed for participants to reorient themselves and find objects previously seen in the virtual scene. We used a similar task to evaluate the three TECHNIQUES in terms of spatial awareness and manipulation efficiency. Before each trial, the participant was asked to walk to a starting point presented by a green dotted circle on the floor. Then, a set of pillars were displayed, and the trial started. The pillar set consists of one red and seven blue pillars located on four sides of a 3m × 3m square. The participant had to adjust the virtual workspace position to enclose all the pillars as if they wanted to be able to access all of them without having to perform additional teleportations. Once all the pillars were enclosed, the participant could release the Vive controller touch pad to travel to the selected destination. Subsequently, the participant needed to touch the red pillar with the Vive controller to end the trial.

4.5 Procedure

Each participant was welcomed, received instructions on the task, and signed an informed consent form. After setting up the head-mounted display, we calibrated the eye tracker. For each TECHNIQUE, the participant first experienced training trials. Next, the participant completed 24 trials in randomized order resulting from a particular subset of the full combination of 6 starting points in the real space, 8 relative directions between the starting point and the pillar set (-135°, -90°, -45°, 0°, 45°, 90°, 135°, 180°), 3 distances between the starting point and the pillar set (4m, 6m, 8m) and 3 orientations of the pillar set (0°, 22.5°, 45°), as illustrated in Figure 5. After each TECHNIQUE, the participant filled out a questionnaire. At the end of the experiment, the participant also ranked the three TECHNIQUES according to their preference. The whole experiment lasted approximately 45 minutes.

4.6 Data collection

We registered 864 trials: 3 TECHNIQUES × 24 repetitions × 12 participants. For each trial, we logged the following measures:

- *Task Completion Time (TCT)*: the total duration of a trial. The measurement started when the participant arrived at the starting point and ended when the red pillar was touched.

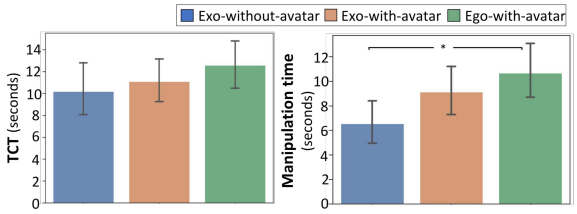

Figure 6: Mean *TCT* (left) and *manipulation time* (right) by TECHNIQUE. Error bars show 95% confidence intervals (CI).

- *Manipulation time*: the time used to enclose all the pillars. The measurement started when the virtual workspace representation collided with one of the pillars and ended when the participant triggered the teleportation.
- *Target identification time*: the time that the participant needed to reorient themself to find the red pillar. The measurement started just after the teleportation, and ended when the eye gaze of the participant collided with the red pillar (measured by the eye-tracking system of the HTC Vive).

We used the NASA-TLX questionnaire [27] and also added two more questions about *anticipation* (Were you able to anticipate where you would be after teleportation?) and *disorientation* (Did you feel disoriented after teleportation?). Criteria were graded on a 21-point scale and later converted to a 100-point score.

4.7 Statistical results

For each measure, we used normal QQ-plots and Shapiro-Wilk Tests to analyze data normality. *TCT*, *manipulation time* and *target identification time* were not normally distributed, so we applied a log-transformation to analyze them, as recommended by Robertson & Kaptein [37] (p. 316). To minimize the noise in our data, we averaged the 24 repetitions of each TECHNIQUE. We then ran a one-way ANOVA test and conducted post-hoc analysis with paired sample T-tests with Bonferroni corrections[1]. Means (M) are reported with standard deviations.

For *TCT* (see Figure 6, left), we did not find a significant effect of TECHNIQUE ($F_{2,22} = 1.291$, $p = 0.295$), and all conditions had close mean values: *Exo-without-avatar* ($M = 10.18 \pm 4.21$s), *Exo-with-avatar* ($M = 11.09 \pm 3.43$s) and *Ego-with-avatar* ($M = 12.56 \pm 3.85$s).

For *manipulation time* (see Figure 6, right), we observed a significant effect of TECHNIQUE ($F_{2,22} = 4.683$, $p = 0.0202$). Pairwise comparisons showed that participants spent less time with *Exo-without-avatar* ($M = 6.54 \pm 3.32$s) than with *Ego-with-avatar* ($M = 10.66 \pm 4.07$s, $p = 0.0042$). No significant differences were found between *Exo-without-avatar* and *Exo-with-avatar* ($M = 10.66 \pm 4.07$s, $p = 0.132$), and between *Exo-with-avatar* and *Ego-with-avatar* ($p = 0.64$).

For *target identification time* (see Figure 7, left), we detected a significant effect of TECHNIQUE ($F_{2,22} = 17.40$, $p < 0.0001$). Pairwise comparisons showed that target identification time was significantly shorter with *Exo-with-avatar* ($M = 0.81 \pm 0.20$s, $p = 0.0021$) and *Ego-with-avatar* ($M = 0.77 \pm 0.38$s, $p = 0.0015$) than with *Exo-without-avatar* ($M = 1.79 \pm 1.06$s). No significant differences were found between *Exo-with-avatar* and *Ego-with-avatar* ($p = 0.57$).

[1] All statistical analyses were performed with R and we used a significance level of $\alpha = 0.05$ for all tests.

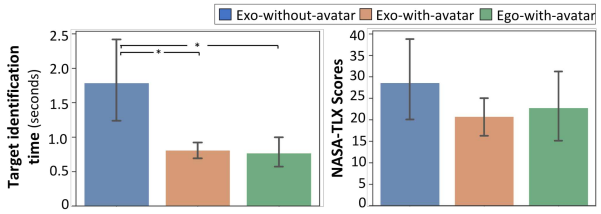

Figure 7: Mean *target identification time* (left) and NASA-TLX score (right) by TECHNIQUE. Error bars show 95% CI.

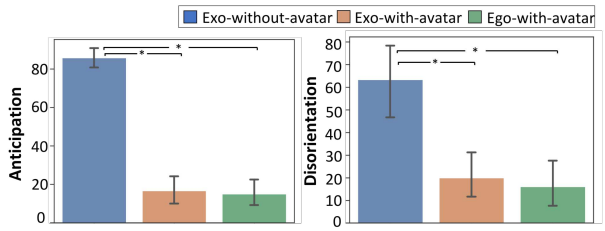

Figure 8: Mean *anticipation* (left) and *disorientation* (right) score by TECHNIQUE. Error bars show 95% CI.

For the subjective questionnaire, we used Friedman's tests and Wilcoxon signed-rank tests for post-hoc analysis in conformity with such non-parametric data. For cognitive load, we did not find a significant effect of TECHNIQUE ($\chi^2(2) = 1.721$, $p = 0.423$) on the NASA-TLX score (see Figure 7, right). However, we detected a significant effect of TECHNIQUE on *anticipation* ($\chi^2(2) = 19.19$, $p < 0.0001$) and *disorientation* ($\chi^2(2) = 21.80$, $p < 0.0001$). Post-hoc analysis shows that *Exo-with-avatar* ($M = 16.67 \pm 12.67$, $p = 0.0057$) and *Ego-with-avatar* ($M = 15.00 \pm 12.06$, $p = 0.0061$) resulted in a significantly better anticipation (see Figure 8, left) compared to *Exo-without-avatar* ($M = 85.83 \pm 9.00$). It shows that significantly less disorientation (see Figure 8, right) was perceived by the participants with *Exo-with-avatar* ($M = 20.00 \pm 19.19$, $p = 0.015$) and *Ego-with-avatar* ($M = 16.08 \pm 19.08$, $p = 0.0099$) than with *Exo-without-avatar* ($M = 63.33 \pm 30.70$). In addition, 11 out of 12 participants preferred *Exo-with-avatar* and *Ego-with-avatar* over *Exo-without-avatar*, and 8 out of 12 participants ranked *Ego-with-avatar* as their favorite condition.

4.8 Discussion

Despite the non-significant difference in task completion time, we found that the use of an avatar has a significant impact on manipulation time and target identification time. On the one hand, participants performed faster the manipulation task of the virtual workspace with *Exo-without-avatar* than with *Ego-with-avatar*. Manipulation with *Exo-without-avatar* seems also slightly faster than with *Exo-with-avatar*, but the difference is not significant. On the other hand, the two conditions with avatar resulted in significantly less disorientation and thus a shorter target identification time, which supports **H1**. Even if the results could be predictable since the visual feedback directly shows how the user will "land" in the VE, it is interesting to measure its actual impact. While the user spends slightly more time positioning the avatar, they can plan and better understand the upcoming teleportation, decreasing disorientation and the time needed to complete the task after the teleportation. In the conditions with avatar, positioning the avatar can increase the cognitive load, but finding the target requires less cognitive effort. This can explain why the overall difference in cognitive load is not significant.

Contrary to our expectations, we were not able to find significant differences between exocentric and egocentric techniques in term of user performance. For manipulation time, the difference is not significant, which does not support **H2**. For target identification time, no significant difference was found between *Exo-with-avatar* and *Ego-with-avatar*, which rejects **H3**. We also did not detect significant differences between *Exo-with-avatar* and *Ego-with-avatar* for cognitive load, anticipation and disorientation. However, participants preferred *Ego-with-avatar* to *Exo-with-avatar* according to the questionnaires. In particular, participants reported that *Ego-with-avatar* allowed them to "focus more on themselves" (P6) during the manipulation step, was "easier for positioning themselves" (P9), and was "easier for finding" (P3) the target objects after teleportation.

The two conditions with avatar seem to reach close performance levels. However, we wanted to select one of them to compare it with the automatic positioning technique in the second experiment. Consequently, we decided to choose the *Ego-with-avatar* based on the user preference.

5 EXPERIMENT 2

The goal of this experiment is to compare the manual technique selected from the first experiment (i.e., *Ego-with-avatar*) and the automatic technique to a basic teleportation. In this experiment, we set up a more realistic task similar to an escape room game. The VE consisted of a series of virtual rooms. Participants needed to select multiple objects to escape each room and continue the exploration. The experiment followed a [3×2] within-subject design with the following factors:

- TECHNIQUE: *Basic*, *Manual* and *Automatic*,
- LAYOUT: *Overlap* and *No-overlap*.

For TECHNIQUE (see Figure 9), the three variations are:

- *Basic* is the basic teleportation technique used as a baseline. A virtual ray appeared when participants pressed the controller's touchpad. The teleportation position was determined by the collision point between the ray and the virtual floor. It was represented by a green dotted circle. Participants activated teleportation by releasing the touch pad.
- *Manual* is the manual *Ego-with-avatar* technique described in section 4. Participants used the virtual ray to manipulate the virtual workspace representation instead of the dotted circle used in the *Basic* technique.
- *Automatic* is the automatic technique described in Section 3. Participants used the virtual ray to select a virtual workspace among the relevant options proposed by Algorithm 1.

For LAYOUT, two different object layouts were used:

- *No-overlap*: objects were laid out in two separate areas, which could be included in the participant's real workspace. The *Automatic* technique thus proposed one virtual workspace for each area.

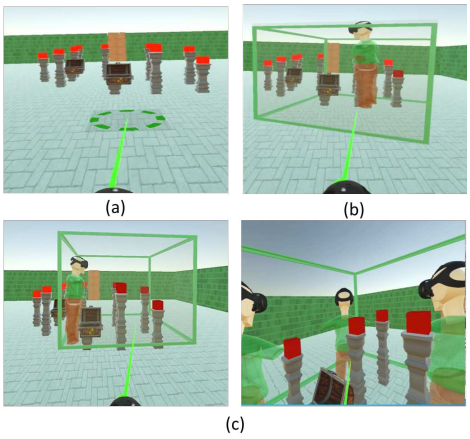

Figure 9: Three conditions for TECHNIQUE: (a) *Basic* teleportation; (b) *Manual* technique selected from EXPERIMENT 1; (c) *Automatic* technique: the user selects their future virtual workspace among relevant options (left), or selecting avatars in overlapping conditions (right).

- *Overlap*: objects were spread in an area larger than the participant's real workspace. The *Automatic* technique proposed a set of virtual workspaces which enclosed only a subset of the objects.

The order of the TECHNIQUES was counter-balanced across participants using a balanced Latin square, and the order of the LAYOUT was also counter-balanced for each TECHNIQUE.

5.1 Hypothesis

In comparison to the basic teleportation, we expected that the manual and the automatic positioning techniques would allow the user to access more easily multiple objects using physical walking. With the automatic technique, the user could select their future virtual workspace among the proposed ones and thus would be able to avoid the manipulation step required for positioning it. However, in a crowded virtual environment, the large number of proposed virtual workspaces could be confusing for the user. We, therefore, formulated the following hypotheses:

- **H1** *Automatic* and *Manual* will result in better user performance, compared to the *Basic* teleportation.
- **H2** *Automatic* and *Manual* will result in better sense of presence, compared to the *Basic* teleportation.
- **H3** In *No-overlap*, *Automatic* performs better than *Manual*.
- **H4** In *Overlap*, *Manual* performs better than *Automatic*.

5.2 Participants

We recruited 12 participants, aged between 25 and 32 (7 men, 5 women). 6 participants had VR experience. 8 out of 12 rated their everyday usage of head-mounted displays as very low.

5.3 Experiment setup

The VR setup was the same as in Experiment 1.

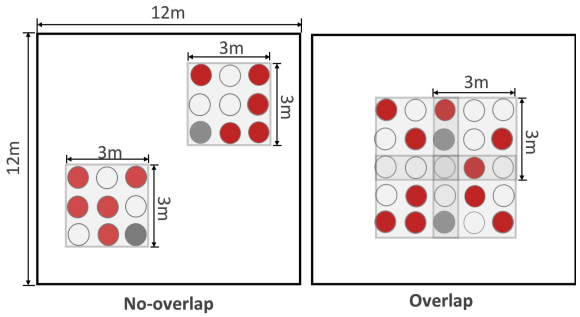

Figure 10: Two object layouts were used: (left) *No-overlap* in which objects (red) and treasure boxes (gray) were randomly located in two disjointed 3m × 3m areas; and (right) *Overlap* in which objects (red) and treasure boxes (gray) were randomly placed in a single 5m × 5m area. The light gray rectangles represent the virtual workspaces computed in the *Automatic* condition.

5.4 Experimental task

Participants traveled in a large virtual environment composed of nine rooms. In each room, they had to select multiple objects. When participants reached their real workspace limits, a warning sign appeared in their field of view with an alarm sound to ensure participants' safety. The first room was used for the training task, and the other eight were set up for the evaluation: half with the *No-overlap* layout and half with the *Overlap* layout. Both types of layouts required participants to access ten target objects, grab them with the controller, and bring them back to one of two treasure boxes one by one (see Figure 10). With the *No-overlap* layout, the objects and the treasure boxes were located within two disjointed 3m × 3m areas. Each area contained five targets and one treasure box located randomly on 6 of 9 positions. With the *Overlap* layout, the objects and the treasure boxes were placed randomly on 12 of 25 positions located in a single 5m × 5m area. In *Automatic* condition of this layout, the algorithm computed four overlapping virtual workspace positions to cover the full area.

5.5 Procedure

Each participant was welcomed, received instructions on the task, and signed an informed consent form. For each TECHNIQUE, the participant first completed the training task with eight targets and two treasure boxes located inside four 3m × 3m areas (three overlapped and one not-overlapped). During this step, the experimenter was allowed to answer their questions, if any. Next, the participant completed eight trials (4 with *No-overlap* and 4 with *Overlap*, or vice versa). After each TECHNIQUE, they filled out an Igroup Presence Questionnaire (IPQ) [36, 38] to measure the sense of presence and two NASA-TLX questionnaires [27] to assess the cognitive load of each LAYOUT. We used the color of the rooms to help the participant to differentiate the two LAYOUTS. At the end of the experiment, the participant also ranked the three techniques according to their preference. The whole experiment lasted around 60 min.

5.6 Data collection

We registered 288 trials: 3 TECHNIQUES × 2 LAYOUTS × 4 repetitions × 12 participants. For each trial, we collected the following measures:

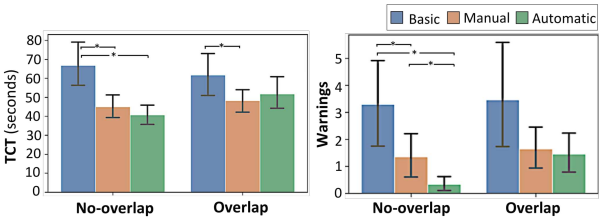

Figure 11: Mean *TCT* (left) and *warnings* (right) by TECHNIQUE × LAYOUT. Error bars show 95% CI.

- *Task Completion Time (TCT)*: the total duration of a trial. The measurement started when the participant entered a room and ended when all objects were put in the treasure boxes.
- *Warnings*: the number of times the participant triggered the warning sign.
- *Teleportations*: the number of teleportations performed.

Each question of the NASA-TLX was graded on a 21-point scale and converted to a 100-point score. Each question of the IPQ was graded on a 7-point Likert scale (from 0 to 6).

5.7 Statistical results

To minimize the noise in our data, we averaged the 4 repetitions of each TECHNIQUE × LAYOUT. Means (M) are reported with standard deviations.

For *TCT*, we used normal QQ-plots and Shapiro-Wilk Tests to analyze data normality. The data was not normally distributed, so we applied a log-transformation to analyze it following statistical recommendations [37]. A two-way repeated measures ANOVA with the model TECHNIQUE × LAYOUT revealed a significant effect of TECHNIQUE ($F_{2,22} = 11.08$, $p = 0.0005$) and interaction effect ($F_{2,22} = 4.79$, $p = 0.019$), but no significant effect of LAYOUT ($F_{1,11} = 2.13$, $p = 0.17$) was found. For TECHNIQUE, post-hoc Tukey HSD tests indicated that performing the task with *Manual* ($M = 45.70 \pm 9.65$s, $p = 0.0019$) and *Automatic* ($M = 45.40 \pm 9.46$s, $p = 0.0011$) was significant faster than with *Basic* ($M = 61.70 \pm 20.19$s). For TECHNIQUE × LAYOUT (see Figure 11, left), post-hoc Tukey HSD tests shown the task with the *No-overlap* layout was significant faster to achieve with *Manual* ($M = 43.75 \pm 10.63$s, $p = 0.0005$) and *Automatic* ($M = 39.80 \pm 8.75$s, $p < 0.0001$) than with *Basic* ($M = 64.02 \pm 20.60$s). For the task with the *Overlap* layout, *Basic* ($M = 58.99 \pm 19.23$s) was significantly different from *Manual* ($M = 46.97 \pm 10.86$s, $p = 0.044$), but not from *Automatic* ($M = 49.87 \pm 13.75$s, $p = 0.17$). In all cases, no significant differences were found between *Manual* and *Automatic*.

For *Warnings*, we used non-parametric tests in conformity with the nature of count data. We first aggregated the data by TECHNIQUE and a Friedman test revealed a significant effect of TECHNIQUE ($\chi^2(2) = 11.51$, $p = 0.0032$). Wilcoxon Signed Rank tests[2] shown that participants triggered significantly more *Warnings* with *Basic* ($M = 3.38 \pm 3.32$) than with *Manual* ($M = 1.50 \pm 1.11$, $p = 0.041$) and *Automatic* ($M = 0.90 \pm 0.67$, $p = 0.034$). No significant differences were found between *Manual* and *Automatic*. We then split the data by LAYOUT and ran a Friedman test for each LAYOUT (see Figure 11, right). For *No-overlap*, it indicated a significant effect of TECHNIQUE ($\chi^2(2) = 14.28$, $p = 0.0008$). Wilcoxon Signed Rank tests shown that participants triggered significant more *Warnings* for the *No-overlap* layout with *Basic* ($M = 3.29 \pm 2.91$) than with *Manual* ($M = 1.35 \pm 1.41$, $p = 0.029$) and *Automatic* ($M = 0.33 \pm 0.44$, $p = 0.011$). *Manual* was also significantly different than *Automatic* ($p = 0.032$) for *No-overlap*. For *Overlap*, no significant effect of TECHNIQUE was found ($\chi^2(2) = 4.95$, $p = 0.084$).

[2]In this experiment, all Wilcoxon Signed Rank tests were performed with Holm-Bonferroni corrections.

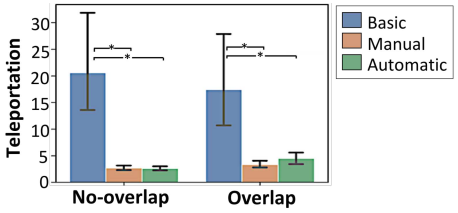

Figure 12: Mean *Teleportations* by TECHNIQUE × LAYOUT. Error bars show 95% CI.

For *Teleportations*, we also used non-parametric tests. We first aggregated the data by TECHNIQUE and a Friedman test revealed a significant effect of TECHNIQUE ($\chi^2(2) = 19.50$, $p < 0.0001$). Wilcoxon Signed Rank tests shown that participants teleported significantly more with *Basic* ($M = 19.00 \pm 17.49$) than with *Manual* ($M = 3.00 \pm 0.85$, $p = 0.0050$) and *Automatic* ($M = 3.53 \pm 0.90$, $p = 0.0015$). No significant difference was found between *Manual* and *Automatic*. We then split the data by LAYOUT and ran a Friedman test for each LAYOUT (see Figure 12). For both LAYOUTS, we had similar results to that of the aggregated data, pointing out that there was probably no interaction effect of TECHNIQUE×LAYOUT.

For the NASA-TLX questionnaires, we aggregated the data by TECHNIQUE and did not detect a significant effect of TECHNIQUE ($\chi^2(2) = 2.426$, $p = 0.2974$). We also aggregated the data by LAYOUT and observed that the *Overlap* layout ($M = 42.22 \pm 21.72$) induced a significantly higher cognitive load than the *No-overlap* layout ($M = 38.75 \pm 19.76$, $p = 0.0050$). Further analysis on the data split by LAYOUT revealed that a significant higher cognitive load was required with *Automatic* ($M = 32.98 \pm 18.76$) than with *Manual* ($M = 28.96 \pm 17.95$, $p = 0.031$) for the *Overlap* layout.

For the IPQ questionnaire (see Figure 13), a Friedman test revealed a significant effect of TECHNIQUE ($\chi^2(2) = 19.63$, $p < 0.0001$). Wilcoxon Signed Rank tests reported a significantly better presence with *Manual* ($M = 3.89 \pm 0.47$, $p = 0.0025$) and *Automatic* ($M = 4.80 \pm 0.53$, $p = 0.0025$) than with *Basic* ($M = 4.75 \pm 0.69$).

Finally, 11 out of 12 participants preferred *Manual* and *Automatic* for both tasks over the *Basic* condition. For the *No-overlap* layout, 6 out of 12 participants ranked *Automatic* as their favorite, and 5 participants preferred *Manual*. For the *Overlap* layout, 6 out of 12 participants preferred *Manual*, and 5 participants preferred *Automatic*.

5.8 Discussion

The results provide evidence that the *Manual* and *Automatic* techniques outperformed the *Basic* teleportation. In particular, participants completed the task significantly faster when they were able to choose the position of the future virtual workspace, compared to the *Basic* teleportation. This supports **H1**. It can be explained by the fact that the participants could reach multiple virtual objects easily with physical walking, avoiding unnecessary teleportations. It is

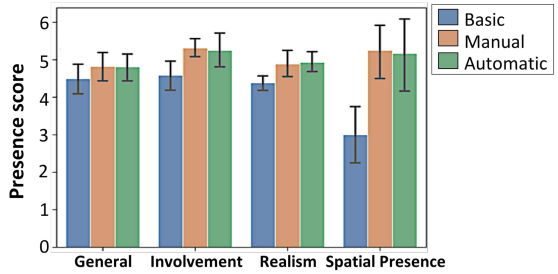

Figure 13: Results from the IPQ questionnaire by TECHNIQUE × LAYOUT. Error bars show 95% CI.

also confirmed by the significantly smaller number of teleportations executed with *Manual* and *Automatic* compared to *Basic*.

In addition to the better performance, the *Manual* and *Automatic* techniques also resulted in higher sense of presence compared to the *Basic* teleportation, according to the IPQ questionnaire. This supports **H2**. With *Basic* teleportation, since the user "cannot imagine the accessible area" (P3) and "cannot determine if an object is accessible" (P7), a larger number of warnings were triggered while performing the task, compared to *Manual* and *Automatic*. Excessive warnings often lead the user to distrust the VR system and break the immersion. For example, participants "feel fear" (P4) and were "afraid to move" (P5) with *Basic* teleportation. The fact that the user performs a more significant number of teleportations and walks less with the *Basic* teleportation is also detrimental to their immersion.

For the task with the *No-overlap* layout, no significant differences in *TCT* were found between *Manual* and *Automatic*, which does not support **H3**. However, significantly more warnings were detected with *Manual* condition than with *Automatic*. The *Manual* technique requires the user to position the virtual workspace manually, and they may sometimes make mistakes, e.g., not including all the target objects or putting the objects too close to the real workspace's limits. As a result, they trigger warnings when they try to access these objects. But the user seems to be able to quickly reposition the virtual workspace when facing issues with the *Manual* technique, which explains the non-significant differences for *TCT*. Consequently, both techniques are suitable in a virtual environment with no-overlapping interaction areas. However, the *Automatic* technique seems more appropriate in such a context since fewer warnings are triggered, thus avoiding immersion breaks.

For the task with the *Overlap* layout, no significant differences in *TCT* were found between *Manual* and *Automatic*, which does not support **H4**. However, while the task was significantly faster to achieve with *Manual* compared to *Basic*, similar results were not reported for *Automatic*, suggesting that a small difference could exist between *Manual* and *Automatic*. In addition, the score from the NASA-TLX shown that *Manual* significantly reduced the cognitive load, compared to *Automatic*. As the virtual workspaces proposed by the system can be numerous and overlapped each other, the user sometimes has to pass through an intermediate virtual workspace to reach the one behind, which can be time consuming and increase the cognitive load. Users also felt "constrained" (P2, P7) as they needed "to adapt to a previously defined position" (P5). Consequently, the *Manual* technique seems more appropriate than the *Automatic* one when the VE is crowded with many objects in the same area.

6 CONCLUSION

In this paper, we proposed and evaluated several techniques helping a user to be aware of their future virtual workspace and manage its position and orientation before the teleportation. Such techniques are interesting to facilitate access to multiple virtual objects through the user's physical movements in their real workspace without reaching its limits. To this aim, we investigated manual and automatic techniques for positioning this virtual workspace.

A first experiment focused on manual positioning techniques. It demonstrates that using an avatar to represent the user's future position in the virtual workspace reduces disorientation and thus the time needed to locate targeted objects after teleportation. It also shows that exocentric and egocentric techniques with an avatar result in a close performance levels, but the egocentric technique seems to be preferred by users. A second experiment compared the egocentric manual technique with an avatar and an automatic technique to a basic teleportation. The manual and automatic positioning techniques outperform the basic teleportation in terms of efficiency and immersion. Although these two positioning techniques reach equivalent performances, each one seems to have its advantages depending on the layout of the virtual environment. Compared to the manual technique, the automatic one causes fewer collisions with the real workspace limits in sparse virtual object layouts, but it induces a higher cognitive load for crowded scenes. In conclusion, this study demonstrates the benefits of the virtual workspace positioning approaches over the basic teleportation.

The future work consists of enhancing the automatic technique by adapting the clustering algorithm to obtain a good balance between the number of proposed virtual workspaces according to the virtual scenario and the real workspace configuration. Moreover, further investigations are required to evaluate these techniques considering different shapes and sizes of real workspaces, the density of the interactive objects in the virtual environment, and the needs of the VR applications. In addition, it remains unclear how the techniques would perform in other VR scenarios or more cognitively challenging situations, for example, in an exploration task where the main goal is not purely to interact with objects. Depending on user expectations, skills and experiences, the virtual workspace's visual feedback could be automatically adjusted based on the amount of the information contained in the scene [22] to avoid overloading the user's field of view and to fit virtual scenario needs. Finally, the suitability of automatic techniques could be studied for specific scenarios that require a perfect match between the real and virtual environments, for example, to provide tangibility to virtual objects [16, 40, 46] or in collaborative co-located applications.

ACKNOWLEDGMENTS

This work was partially supported by European Research Council (ERC) grant n° 695464 "ONE: Unified Principles of Interaction"; and by Agence Nationale de la Recherche (ANR) grants ANR-10-EQPX-26-01 "EquipEx DIGISCOPE" as part of the program "Investissement d'Avenir" ANR-11-IDEX-0003-02 "Idex Paris-Saclay".

REFERENCES

[1] 2017. Tales of escape. https://store.steampowered.com/app/587860/Tales_of_Escape/. [Online; accessed].

[2] 2017. We were here. https://store.steampowered.com/app/582500/We_Were_Here/. [Online; accessed].

[3] 2018. Essential introduction to Steam VR and Unity part 2: Teleporting and navigating around your scene. https://www.youtube.com/watch?v=GozB7zlD1wQ. [Online; accessed].

[4] 2018. Self Guided VR Training. https://pixovr.com/vr-training-center/. [Online; accessed].

[5] 2020. HalfLife Alyx. https://store.steampowered.com/app/546560/HalfLife_Alyx/. [Online; accessed].

[6] Mahdi Azmandian, Timofey Grechkin, Mark T Bolas, and Evan A Suma. 2015. Physical Space Requirements for Redirected Walking: How Size and Shape Affect Performance.. In *ICAT-EGVE*. 93–100.

[7] Niels H Bakker, Peter O Passenier, and Peter J Werkhoven. 2003. Effects of head-slaved navigation and the use of teleports on spatial orientation in virtual environments. *Human factors* 45, 1 (2003), 160–169.

[8] Woodrow Barfield, Craig Rosenberg, and Thomass A. Furness III. 1995. Situation awareness as a function of frame of reference, computer-graphics eyepoint elevation, and geometric field of view. *The International Journal of Aviation Psychology* 5, 3, 233–256.

[9] Jiwan Bhandari, Paul MacNeilage, and Eelke Folmer. 2018. Teleportation without spatial disorientation using optical flow cues. In *Proceedings of Graphics Interface*, Vol. 2018.

[10] Benjamin Bolte, Frank Steinicke, and Gerd Bruder. 2011. The jumper metaphor: an effective navigation technique for immersive display setups. In *Proceedings of Virtual Reality International Conference*.

[11] Patrick Bourdot and Damien Touraine. 2002. Polyvalent Display Framework to Control Virtual Navigations by 6DOF Tracking. In *Proceedings of IEEE Virtual Reality Conference, VR '02*. IEEE Computer Society, 277–278.

[12] Doug A Bowman, David Koller, and Larry F Hodges. 1997. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*. IEEE, 45–52.

[13] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. ACM, 205–216.

[14] Jeff Butterworth. 1992. 3DM: a three-dimensional modeler using a head-mounted display. (1992).

[15] Weiya Chen, Nicolas Ladeveze, Céline Clavel, and Patrick Bourdot. 2016. Refined experiment of the altered human joystick for user cohabitation in multi-stereocopic immersive CVEs. In *IEEE International Workshop on Collaborative Virtual Environments, 3DCVE@IEEEVR*. IEEE Computer Society, 1–8.

[16] Lung-Pan Cheng, Li Chang, Sebastian Marwecki, and Patrick Baudisch. 2018. iTurk: Turning Passive Haptics into Active Haptics by Making Users Reconfigure Props in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 89.

[17] Lung-Pan Cheng, Eyal Ofek, Christian Holz, and Andrew D Wilson. 2019. VRoamer: Generating On-The-Fly VR Experiences While Walking inside Large, Unknown Real-World Building Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 359–366.

[18] Alexandros Koilias Christos Mousas, Dominic Kao and Banafsheh Rekabdar. 2020. Real and Virtual Environment Mismatching Induces Arousal and Alters. In *Virtual Reality Conference, 2020. VR 2020. IEEE*. IEEE.

[19] Gabriel Cirio, Maud Marchal, Tony Regia-Corte, and Anatole Lécuyer. 2009. The magic barrier tape: a novel metaphor for infinite navigation in virtual worlds with a restricted walking workspace. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 155–162.

[20] Cédric Fleury, Alain Chauffaut, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. 2010. A generic model for embedding users' physical workspaces into multi-scale collaborative virtual environments. In *ICAT 2010 (20th International Conference on Artificial Reality and Telexistence)*.

[21] Sebastian Freitag, Dominik Rausch, and Torsten Kuhlen. 2014. Reorientation in virtual environments using interactive portals. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 119–122.

[22] Sebastian Freitag, Benjamin Weyers, Andrea Bönsch, and Torsten W Kuhlen. 2015. Comparison and Evaluation of Viewpoint Quality Estimation Algorithms for Immersive Virtual Environments. *ICAT-EGVE* 15 (2015), 53–60.

[23] Markus Funk, Florian Müller, Marco Fendrich, Megan Shene, Moritz Kolvenbach, Niclas Dobbertin, Sebastian Günther, and Max Mühlhäuser. 2019. Assessing the Accuracy of Point & Teleport Locomotion with Orientation Indication for Virtual Reality using Curved Trajectories. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[24] Jeffrey Goldsmith and John Salmon. 1987. Automatic creation of object hierarchies for ray tracing. *IEEE Computer Graphics and Applications* 7, 5 (1987), 14–20.

[25] Nathan Navarro Griffin and Eelke Folmer. 2019. Out-of-body locomotion: Vectionless navigation with a continuous avatar representation. In *25th ACM Symposium on Virtual Reality Software and Technology*. 1–8.

[26] MP Jacob Habgood, David Moore, David Wilson, and Sergio Alapont. 2018. Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 371–378.

[27] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183.

[28] K Krishna and M Narasimha Murty. 1999. Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29, 3 (1999), 433–439.

[29] James Liu, Hirav Parekh, Majed Al-Zayer, and Eelke Folmer. 2018. Increasing walking in VR using redirected teleportation. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 521–529.

[30] Jack M Loomis, José A Da Silva, Naofumi Fujita, and Sergio S Fukusima. 1992. Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance* 18, 4 (1992), 906.

[31] John Finley Lucas. 2005. *Design and evaluation of 3D multiple object selection techniques*. Ph.D. Dissertation. Virginia Tech.

[32] Suzanne P McKee and Harvey S Smallman. 1998. 14 Size and speed constancy. *Perceptual constancy: Why things look as they do* (1998), 373.

[33] Mark R Mine. 1995. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept* (1995).

[34] Ji-Young Oh, Wolfgang Stuerzlinger, and Darius Dadgari. 2006. Group selection techniques for efficient 3D modeling. In *3D User Interfaces (3DUI'06)*. IEEE, 95–102.

[35] Sharif Razzaque, Zachariah Kohn, and Mary C. Whitton. 2001. Redirected Walking. In *Eurographics 2001 - Short Presentations*. Eurographics Association.

[36] Holger Regenbrecht and Thomas Schubert. 2002. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 11, 4 (2002), 425–434.

[37] Judy Robertson and Maurits Kaptein. 2016. *Modern Statistical Methods for HCI*. Springer.

[38] Thomas W Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie* 15, 2 (2003), 69–71.

[39] Adalberto L Simeone, Ifigeneia Mavridou, and Wendy Powell. 2017. Altering user movement behaviour in virtual environments. *IEEE transactions on visualization and computer graphics* 23, 4 (2017), 1312–1321.

[40] Adalberto L Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional reality: Using the physical environment to design virtual reality experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3307–3316.

[41] Misha Sra, Sergio Garrido-Jurado, Chris Schmandt, and Pattie Maes. 2016. Procedurally generated virtual reality from 3D reconstructed physical space. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. 191–200.

[42] Rasmus Stenholt. 2012. Efficient selection of multiple objects on a large scale. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*. 105–112.

[43] Wolfgang Stuerzlinger and Graham Smith. 2002. Efficient manipulation of object groups in virtual environments. In *Proceedings IEEE Virtual Reality 2002*. IEEE, 251–258.

[44] Evan A Suma, Zachary Lipps, Samantha Finkelstein, David M Krum, and Mark Bolas. 2012. Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (2012), 555–564.

[45] Qi Sun, Li-Yi Wei, and Arie Kaufman. 2016. Mapping virtual and physical reality. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–12.

[46] Keisuke Suzuki, Sohei Wakisaka, and Naotaka Fujii. 2012. Substitutional reality system: a novel experimental platform for experiencing alternative reality. *Scientific reports* 2 (2012), 459.

[47] Bernd Froehlich Tim Weissker, Alexander Kulik. 2019. Multi-Ray Jumping: Comprehensible Group Navigation for Collocated Users in Immersive Virtual Reality. IEEE.

[48] Tim Weißker, André Kunert, Bernd Frohlich, and Alexander Kulik. 2018. Spatial updating and simulator sickness during steering and jumping in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 97–104.

[49] Yiran Zhang, Nicolas Ladeveze, Cédric Fleury, and Patrick Bourdot. 2019. Switch Techniques to Recover Spatial Consistency Between Virtual and Real World for Navigation with Teleportation. In *International Conference on Virtual Reality and Augmented Reality*. Springer, 3–23.

Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs

Yiran Zhang*
University Paris-Saclay, CNRS, LISN, VENISE team, Orsay, France
Huyen Nguyen†
University Paris-Saclay, LISN, VENISE team, Orsay, France
Nicolas Ladevèze‡
University Paris-Saclay, CNRS, LISN, VENISE team, Orsay, France
Cédric Fleury§
IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest, France
Patrick Bourdot¶
University Paris-Saclay, CNRS, LISN, VENISE team, Orsay, France

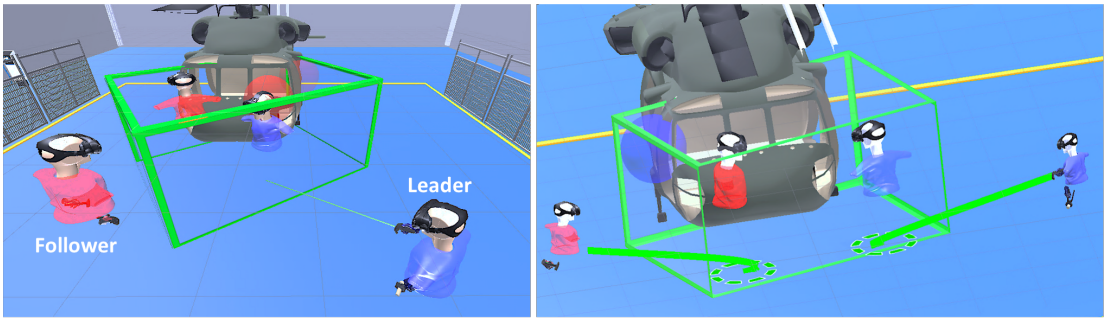

Figure 1: Two virtual workspace positioning techniques to help co-located users to recover their spatial consistency after a teleportation. The 3D volume framed in green represents the virtual representation in the VE of the user shared physical workspace. While positioning this virtual workspace, the users can predict their future position and orientation after the teleportation by observing their preview avatars (each user is represented by a distinct color). *Left*: **Leader-and-Follower** technique allows one of the users to fully manipulate the position and orientation of the virtual workspace. The other user can only communicates their own requirements for this manipulation. *Right*: **Co-manipulation** technique integrates the inputs from both of the users, allowing concurrent positioning.

Abstract

In many collaborative virtual reality applications, co-located users often have their relative position in the virtual environment matching the one in the real world. The resulting spatial consistency facilitates the co-manipulation of shared tangible props and enables the users to have direct physical contact with each other. However, these applications usually exclude their individual virtual navigation capability, such as teleportation, as it may break the spatial configuration between the real and virtual world. As a result, the users can only explore the virtual environment of approximately similar size and shape compared to their physical workspace. Moreover, their individual tasks with unlimited virtual navigation capability, which often take part in a continuous workflow of a complex collaborative scenario, have to be removed due to this constraint. This work aims to help overcome these limits by allowing users to recover spatial consistency after individual teleportation in order to re-establish their position in the current context of the collaborative task. We use a virtual representation of the user's shared physical workspace and develop two different techniques to position it in the virtual environment. The first technique allows one user to fully position the virtual workspace, and the second approach enables concurrent positioning by equally integrating the input from all the users. We compared these two techniques in a controlled experiment in a virtual assembly task. The results show that allowing two users to manipulate the workspace significantly reduced the time they spent negotiating the position of the future workspace. However, the inevitable conflicts in simultaneous co-manipulation were also a little confusing to them.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Collaborative interaction

1 Introduction

Computer-supported collaborative work (CSCW) tools have been widely deployed in diverse fields such as industrial training, data exploration, product design, and entertainment, to name a few, to allow a group of users to communicate, interact with each other, and coordinate their activities to solve collaborative tasks. According to the collaborators' geographical location and whether the collaboration is performing simultaneously, the group interaction can be categorized according to a time/location matrix [13]. In this matrix, different group interactions can be distinguished as same-time (synchronous) and/or different-time (asynchronous) interactions, as well as same-location (co-located) and/or different-location (remote) interactions. This paper investigates co-located synchronous teamwork using head-mounted displays (HMDs), where multiple users share the same physical tracked space while immersing in a collaborative virtual environment (CVE).

In many collaborative virtual reality (VR) applications, the users'

*e-mail: yiran.zhang@universite-paris-saclay.fr
†e-mail: huyen.nguyen@universite-paris-saclay.fr
‡e-mail: nicolas.ladeveze@universite-paris-saclay.fr
§e-mail: cedric.fleury@imt-atlantique.fr
¶e-mail: patrick.bourdot@universite-paris-saclay.fr

relative position in the virtual environment (VE) matches their position in the real world. In this situation, the spatial consistency between the real (physical) and virtual environment enriches the VR experience by allowing, for example, a direct physical interaction between users [25] as well as the integration of shared tangible props in the scene [3,11,29]. However, this one-to-one mapping also limits the users' accessible area in the VE, and they can only explore the VE whose size is similar to the size of the physical workspace.

Teleportation is a widely used navigation technique that allows the users to explore a virtual space that is larger than their physical workspace while minimizing simulator sickness [15, 19, 36]. However, for the co-located users equipped with HMDs, after individual teleportation, the spatial relationship of their avatars in the virtual environment often differs from their counterpart in the real world. This position offset, also referred to as spatial desynchronization [23], makes the interaction that relies on the one-to-one mapping of the real and virtual workspace impossible. The lack of awareness of the position of other users in the real world increases the risk of collisions between them. In addition, they can only hear each other's voice from their position in the real world rather than their avatar's position. The dual-presence of the real and virtual audio stimuli generates perceptual conflicts and can greatly impact the user's task performance [10].

In this context, our work aims to overcome these limits by helping the users to recover the spatial consistency after individual teleportation. Such spatial consistency recovery techniques are useful in many scenarios, especially for complex workflows involving individual sub-tasks as well as collaborative sub-tasks (which specifically require spatial consistency). This is typically the case for complex virtual assembly simulations, where the users often have to perform a continuous virtual activity including individual and collaborative sub-tasks at different times, depending on the actual operation at hand. For example, during the assembly task, they can first navigate individually to different warehouses to obtain mechanical pieces. If the spatial consistency can be restored in the following collaborative phase when they come back to a shared space, they can then walk freely within this area and interact directly with each other without having any perceptual conflicts. Besides, shared tangible objects can also be integrated into the assembly task to coordinate the users' movement and provide them with additional passive haptic feedback.

In our approach, we deploy a virtual representation of the users' shared physical workspace in the VE. They can position such the virtual workspace in the VE while taking into consideration the requirements of their subsequent collaborative task. This step thus facilitates the recovery of the spatial consistency after teleporting inside it. We develop two techniques to allow the users to define the virtual workspace's position and orientation. In the scenario of two co-located users, the first technique enables one of the users to control the virtual workspace in the VE, while the other have to communicate their needs with verbal suggestions or other communication cues. The second technique integrates all of the users' inputs equally, thus enabling simultaneous positioning of the virtual workspace. Inspired by the virtual assembly task given as an example above, we envisioned a collaborative virtual riveting task to investigate the performance of these two techniques. The recovered spatial consistency allows the users to have direct physical contact to perform riveting, providing passive haptic feedback during the collaborative task. From the results of the controlled experiment, we derived some usability guidelines for such techniques.

The contributions of this work are:

1. The design of two interactive techniques that allow the users to recover a shared spatial consistency after individual navigation tasks to facilitate collaborative and tangible interaction between them. We intentionally developed two techniques that involve opposite types of collaboration in order to compare them: *Co-manipulation* is symmetric, while *Leader & Follower* is asymmetric.
2. Empirical results on participants' performance and preference when using these two techniques on a task alternating individual and collaborative sub-tasks.
3. An actual scenario demonstrating how recovering the spatial consistency can be useful for a collaborative VR task.

2 Related work

Many collaborative VR application designs rely on a one-to-one mapping between the users' relative positions and rotations in the real and virtual environment. The spatial consistency provided by such mapping enables the possibility of introducing a tangible interface to the co-located users. Indeed, the blended real and virtual environment enriches the users' virtual experience and overcomes the lack of tactile feedback of VR, contributing to a higher sense of presence [21]. For example, a shared prop can be integrated in a virtual windshield [30] or a virtual car hood [3] assembly task to coordinate the users' co-manipulation. In addition, the spatial consistency between the real and virtual workspace allows the co-located users to interact directly with each other without going through an intermediary step, such as a handshake between them [25]. Finally, under the spatial consistency condition, the users' position in the real world is the same as in the VE, which helps to prevent possible collisions during real walking. Moreover, since the sound coming for the users matches with their virtual avatars' location, there is no perceptual conflict regarding the 3D spacial sound and thus the spatial information can be implicitly communicated as if it was in the real world.

One of the major drawbacks in the one-to-one mapping required by such applications is that it limits the size of the virtual environment to the same size of the users' physical workspace. It also constricts the use of virtual navigation since their individual navigation capabilities can break spatial consistency. One possible solution to avoid spatial desynchronization is to consider the co-located users as a group and allow them to virtually navigate as a single entity.

The physical workspace shared by the users can be embedded in the CVE by incorporating a virtual representation of the real environment into the virtual world. For example, 3DM [9] deploys a magic carpet to represent the tracking space. In addition, the user's physical workspace can be incorporated into the virtual environment by being imagined as a virtual vehicle [7] or a virtual cabin [14]. By manipulating such a virtual representation, the users can therefore navigate in the VE while preserving their spatial relationship. For example, C1x6 [22] allows co-located users to navigate inside a virtual vehicle as a group. The users can pilot the vehicle using a shared stationary 3D tracking sphere within the physical workspace. More recently, Multi-Ray jumping [35] allows co-located users to teleport as a group while maintaining their spatial offset during the navigation. When the navigator specifies a target teleportation position using a ray, the corresponding position of the passenger is computed and communicated using a second ray.

Group navigation during a continuous VR experience has its limits and is sometimes unnecessary. In cluttered or confined virtual environments, such as corridors, users often collide with or find themselves inside virtual objects in order to maintain spatial relationships among the group. To solve this problem, the previous study of Beck et al. [5] propose automatically moving users close to each other when they go through a narrow place and recovering the spatial consistency configuration after reaching a collision-free state. However, this approach causes short-term spatial desynchronization and induces users discomfort depending on their shifted offset to the open passage. Moreover, group navigation limits users' individual activities for loosely coupled collaboration stages, for example, individual object searching before the collaborative assembly task described in [10]. Therefore, it is crucial to preserve users' indi-

vidual virtual navigation capabilities and help them recover spatial consistency in some areas of the VE when the need arises.

To meet these criteria, Min et al. [25] propose a recovery algorithm to adjust users' relative position and orientation in the virtual and physical world until they become aligned. The co-located users can use redirected walking technology independently to explore a VE larger than their physical workspace. When the spatial consistency is required to perform direct physical interaction in the VE, the users can trigger the recovery algorithm to achieve the recovered state. As the most natural method for traveling in a VE, walking can help the users better understand the size of VE by providing them with vestibular cues [8]. However, redirected walking technology usually requires a physical workspace larger than 6m $\times$ 6m [4], which is difficult to meet for many VR systems, especially the ones using HMDs. In addition, long physical walk may tire users after a certain time. Consequently, such solutions are not always suitable for large scale individual navigation in any VR systems.

In a single-user context, some previous works explore how to recover a spatial consistency between the real workspace and its virtual counterpart after large scale virtual navigation. They allow a user to teleport themself in a predefined virtual workspace maximizing their usable real space [39] or to choose the position and orientation of their future virtual workspace when they need to adapt the placement of their real space in the virtual environment [40]. However, these solutions are designed for a single user and cannot manage the spatial constraints of multiple users.

Inspired by these approaches, we extend them to the collaborative context by allowing users to customize their shared workspace position and orientation before teleportation while recovering a spatial consistency between them. We proposed in this paper two recovery techniques that enable individual virtual navigation and recover spatial consistency within user-defined areas when necessary for the subsequent collaborative interactions.

3 Spatial Consistency Recovery Techniques

In this section, we will detail the design and implementation of the two interactive techniques to recover spatial consistency in a co-located collaboration using HMDs. These recovery techniques incorporate a virtual representation of the physical workspace of co-located users in the VE. This representation is comparable to the one used for a single user in previous studies [39, 40], except that it handles multiple users. This virtual workspace representation includes a 3D volume with the same shape and size as the real workspace shared by the co-located users. In addition, we present the current group configuration by adding preview avatars [34, 35] in the 3D volume to directly show how each user will be positioned after teleportation. These avatars have different colors that correspond to the users' avatar colors in the VE.

By positioning the virtual workspace in the VE, users can define an area to recover spatial consistency regarding different scenarios and tasks. As collaborative interaction in the VE can be symmetric or asymmetric, we propose two different strategies to allow users to control the position and orientation of the virtual workspace: *Leader and Follower* and *Co-manipulation* techniques. Both of the techniques will be described for a pair of co-located collaborators but they can be extended to a bigger teamwork.

3.1 Leader and Follower Technique

Our first technique allows one of the users (named the *leader*) to position the virtual workspace, while the other user (named the *follower*) can only communicate verbally to the leader their requirements (see Figure 2). When approaching an area that requires spatial consistency for performing collaborative tasks, the leader can press the HTC Vive controller's touch-pad to trigger the control of the virtual workspace. It displays the virtual workspace for both users.

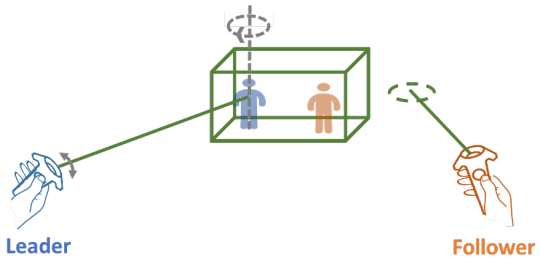

Figure 2: Leader and Follower technique: the leader controls the position and orientation of the virtual workspace and once it is configured, the follower can teleport directly into this space to recover the spatial consistency. The users can observe their future position within this workspace by looking at their preview avatars.

The intersection point between a virtual ray and the virtual ground determines the future position of the leader.

Based on this position and the actual spatial relationship between both users in the real world, the future position of the virtual workspace with the follower's position inside is computed. The leader can also rotate the virtual workspace around the vertical axis of their preview avatar by sliding their finger in a circle on the touch-pad with a one-to-one mapping. This design aims to help users better anticipate the virtual workspace configuration after teleportation by providing them with egocentric cues, as it rotates the virtual workspace around the user's future position. Our motivation is to leverage some generic study results which emphasise that egocentric cues can help users gather self-relevant information and estimate distances more accurately [24]. This choice is also encouraged by the result of a previous single-user virtual workspace positioning study, where we found that users preferred the egocentric design to the situation where the virtual workspace and the user's avatar are rotated as a whole unit [40].

The virtual workspace and the virtual ray controlled by the leader are always visible to the follower during the manipulation (see Figure 1 left). Consequently, even if the follower cannot take action, they can still communicate with their leader about the positioning of the virtual workspace. Once the future position of the virtual workspace is satisfactory to both of them, the leader can release the touch-pad to end the manipulation.

The leader is then teleported to this newly positioned workspace. The follower can use their ray to select the leader-defined virtual workspace and releases the touch-pad to teleport inside that workspace. The spatial consistency between users is thus restored. Instead of having the leader teleporting the two users at the same time, this process is split into two steps to avoid unwanted teleportation of the follower which can create frustration and disorientation.

3.2 Co-manipulation Technique

Unlike the first technique in which only one user (the *leader*) can position the virtual workspace, our second approach allows both users to simultaneously control the virtual workspace representation in the VE (see Figure 3). To achieve this goal, we considered different interaction techniques which have been proposed in the literature to enable multiple users to simultaneously manipulate a shared object. Indeed, collaborative object manipulation is one of the most important interaction tasks in CVE. One plausible solution is to average the translation and rotation of the user movements to obtain the final movement of the shared object [2, 16, 17, 29, 31]. In addition, the user input can be asymmetrically integrated by assigning different degree-of-freedom (DOF) control of the shared object to the users [3, 26].

In the context of teleportation, the users define a remote targeted

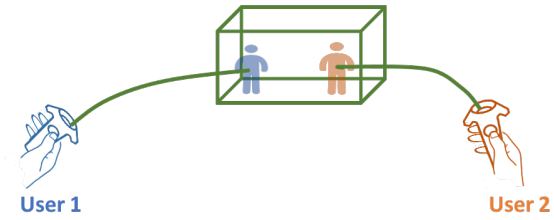

Figure 3: Co-manipulate technique from the front view: the two users will concurrently manipulate the workspace using a bending ray.

point using a pointing technique. To avoid introducing additional inputs, we propose a novel interaction technique that allows users to move and rotate the virtual workspace representation together based on the translation motion of targeted points on the virtual ground. In order to pull the virtual workspace towards users' desired configuration, we consider that the user-defined targeted positions and the users' preview avatars are connected by a mass-spring-damper system (see Figure 4). Similar physically-based approaches have been used to produce realistic virtual object grasping [6] or simulate collision during object manipulation [18].

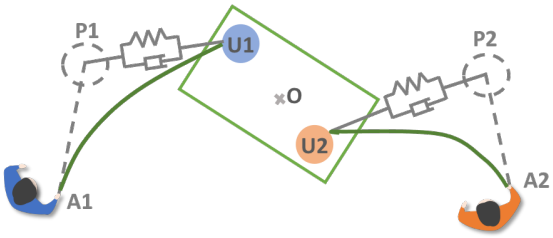

Figure 4: Co-manipulate technique from the top view: the position and orientation of the workspace are calculated from a physically-based approach using a mass-spring-damper system.

At each time step, the translational force coming from each user is first computed. If we assume that U_1 and U_2 are users' preview avatar positions inside the virtual workspace, P_1 and P_2 are the targeted positions defined by the users, then the force coming from *User 1* and *User 2* is computed as follows:

$$\vec{F_1} = k \cdot (P_1 - U_1) + b \cdot \vec{V_{p1}} \quad (1)$$

$$\vec{F_2} = k \cdot (P_2 - U_2) + b \cdot \vec{V_{p2}} \quad (2)$$

$$\vec{F} = \vec{F_1} + \vec{F_2} \quad (3)$$

where k and b are respectively the spring and damper coefficients, $\vec{V_{p1}}$ and $\vec{V_{p2}}$ are the velocity of the targeted point for *User 1* and *User 2*, respectively. The spring and damper coefficients were empirically set to 3.14N/m and 9.85N.s/m to maintain the critical damping of the system. Finally, we symmetrically integrate the user input by adding up the forces that come from both of them (Equation 3) and applying the total force to the point O, the center of gravity of the virtual workspace.

To allow the users to simultaneously control the virtual workspace rotation, we sum the torque coming from each user using the following formula:

$$\vec{T} = \vec{U_{1O}} \times \vec{F_1} + \vec{U_{2O}} \times \vec{F_2} \quad (4)$$

where the $\vec{U_{1O}}$ and $\vec{U_{2O}}$ are the vectors from the point U_1 and U_2 to the center of gravity of the virtual workspace, respectively.

Providing the users with appropriate feedback to show the current state of the shared object's position and orientation is critical in a collaborative manipulation task. Inspired by the Bent Pick Ray technique [27], we use a similar bending ray to continuously inform users about their mutual actions during the whole co-manipulation of the virtual workspace (see Figure 4). The curved ray starts at each user's virtual hand position (A_1 or A_2) and ends at the position of their preview avatar (U_1 or U_2). The user-defined targeted destinations (P_1 or P_2) serve as an additional control point to define a Bézier curve of the 2nd degree (i.e., parabolic curve segment). The deformation of the curved ray indicates the direction and intensity of the user's drag on the virtual workspace.

The users can press the HTC Vive controller's touch-pad to display the virtual ray. The system computes the distance between two targeted points defined by the users. The co-manipulation of the virtual workspace is triggered when this distance is smaller than a specific value (e.g., the length of the diagonal of the virtual workspace). This threshold is set to avoid the situation when one user wants to perform simple individual teleportation and accidentally activates the co-manipulation mode of the virtual workspace. As a first prototype, we implemented a simple approach to end the users' co-manipulation. When reaching an agreement on the configuration of the virtual workspace, either of them can end the co-manipulation by releasing the touch-pad. They will then be teleported into the newly defined workspace, and the spatial relationship between them will thus be recovered. However, by giving both users the ability to end the co-manipulation on their own, there is a risk that users can accidentally trigger this process due to misunderstandings in communication. In future studies, this design could be improved using some alternative solutions, such as locking the virtual workspace manipulation when one of the users is satisfied with the current configuration and waiting for the confirmation from the other.

Finally, if users want to stop the co-manipulation and return to the basic individual teleportation state, they can intentionally increase the distance between the two targeted points and exceed the predefined threshold. The threshold values used for starting and stopping the co-manipulation can be parameterized. For example, the stopping value can be greater than the starting one to allow users to manipulate the virtual workspace over a broader range. In addition, an additional threshold can be set between these two values. For example, when the distance between the two separate target points is about to reach this threshold, the curved rays will turn red, informing users in advance that the co-manipulation is about to end and that the virtual workspace is about to disappear.

4 User Study

We conducted a controlled experiment to compare the two spatial consistency recovery techniques presented in the previous section. We did not compare these two techniques with a baseline condition with which the users have to perform individual teleportation without any assistance to recover the spatial consistency. This is because, for such a baseline condition, it is nearly impossible for the users to recover their spatial consistency without removing their HMDs from time to time, verifying their spatial relationship in the real world, and applying it to the virtual world.

In this experiment, we set up a virtual riveting task in which the co-located participants were asked first to complete an individual task and then return to a designated area to achieve together the riveting of a helicopter shell. Participants had to position the virtual workspace representation according to the current riveting task location. The restored spatial consistency inside the newly defined workspace allows participants to interact directly with each other during the riveting process.

The experiment followed a within-subject design and assessed

two TECHNIQUEs, i.e., *Leader-and-Follower* and *Co-manipulation* described in Section 3 (see Figure 1). The experimental protocol has been approved by the ethics committee of the university.

4.1 Hypotheses

We assumed that *Co-manipulation* would help users to better position the workspace as their intention can be directly conveyed by the manipulation of the virtual workspace, and thus avoiding possible misunderstandings that can occur during the verbal communication in *Leader-and-Follower*. Moreover, we expected to find a power imbalance in the negotiation between the participants and differences in their respective contribution to the task in the *Leader-and-Follower* as it implies an asymmetric role assignment in the virtual workspace positioning process. Therefore, the following hypothesises are formulated:

H1 *Leader-and-Follower* will require more time to discuss and negotiate the future workspace position compared to *Co-manipulation*.

H2 *Co-manipulation* will induce better workspace positioning resulting in a better performance for the riveting task.

H3 *Leader-and-Follower* will be more challenging for the leader.

4.2 Participants

We recruited 24 participants, aged between 21 and 50 ($M = 27.27 \pm 5.51$), with normal or corrected-to-normal vision. 12 pairs were formed at the time of recruitment resulting in 8 male-male and 4 female-male groups. 18 out of 24 participants had previous VR experience and 10 of them rated their everyday usage of HMDs as very low.

4.3 Experiment setup

The VR setup includes two HTC Vive Pro Eye headsets [1]. The outside-in tracking supported by the HTC Vive Lighthouse Tracking System enables tracking of co-located users when one user is out of sight of another (for example, when one user is behind another), ensuring the safety of the user. The two workstations controlling the headsets are connected via a local network, and the tracking spaces of the two users are aligned to a common coordinate system by a calibration procedure [37]. User inputs were obtained through the two Vive handheld controllers that were used for each user. The experiment room supported a 3m $\times$ 4m tracking area. The virtual environment was rendered using Unity (released 2019.4.21) with a resolution of 1440 $\times$ 1600 pixels per eye at 90 Hz. The average time from sending a message from one headset until the reception was measured at 5ms.

4.4 Experiment task

Before the experiment, each participant of a pair was asked to walk to their respective starting points presented by dashed circles on the virtual floor. Each participant was equipped with two controllers, one for teleportation and another for the riveting task. The latter was presented in the VE as a hammer or a riveting pliers, depending on the participant's role. Each participant first performed an individual task to prepare the team riveting task. As illustrated in Figure 5, the participant with the hammer had to teleport close to the charging station to charge the hammer, while the other participant needed to teleport near a shelf to grab the rivets with the riveting pliers. To ensure the safety of the participants during the individual navigation phase towards the charging station and the shelf, warning signs appeared in their field of view with alarm sounds when they reached the limit of their physical working space, or stayed too close to each other and were about to collide [12, 23, 33]. After the participants finished their own individual tasks, they returned to the team riveting area indicated by the yellow frame on the ground. Then, participants had to position their virtual workspace to enclose three riveting positions required by the collaborative task.

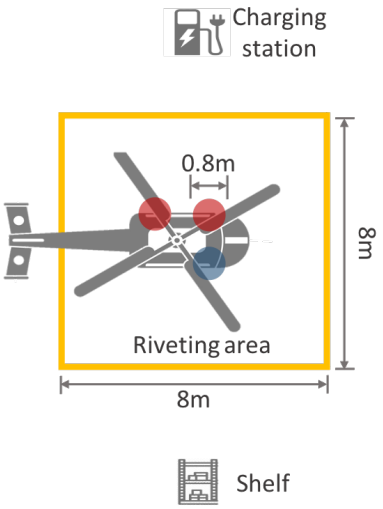

Figure 5: Top view of the VE implemented for the user study: a riveting area (including semi-transparent spheres presenting three predetermined riveting positions) as the shared working area, and two separate areas (including the charging station and the shelf) for individuals tasks.

Individual criteria for the future virtual workspace position were provided differently to the participants: one participant was informed of a part of the required riveting positions, while the other was informed of the remaining positions. Such design is based on the fact that each worker may have a different sequence workflow in an actual riveting process according to their expertise and preference. In addition, the negotiation is encouraged during the virtual workspace positioning stage to mitigate the imbalance of control in *Leader-and-Follower* condition.

The predetermined riveting positions were enclosed in a semi-transparent sphere (d = 0.8m) and displayed to the participants. There were three spheres in total in each operation and each sphere includes two riveting points. Each user was randomly assigned to see one or two of the three spheres either in red or blue. For example, in the case shown in Figure 5, one user can see the two spheres in red, while the other can see the remaining one in blue. To help the participants better determine the position and orientation of the virtual workspace, the color of the sphere is darkened when it is completely enclosed by the workspace.

After the participants configured the position and orientation of the virtual workspace and arrived at this newly positioned workspace with a recovered spatial consistency, the team riveting task began. The participant with the riveting pliers had to place the rivet in the drilled hole of the helicopter shelf, while the other completed the riveting by tapping the end of the rivet with the hammer. The rivet end and the hammer were respectively mapped to the upper area of the controllers held by the participants. As a result, they could feel the hit as they hammered the rivet, which provided them with passive haptic feedback for the collaborative riveting (see Figure 6). When the two riveting points inside one sphere were filled, the color of this sphere faded to gray, prompting the participants to walk to the next riveting position. During the team riveting, since the users' positions in the real world are the same as their avatars' in the VE, the warning signs were only triggered when the users reached the limit of their physical workspace.

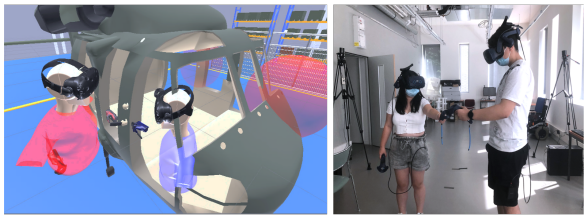

Figure 6: The collaborative riveting task requires that the two controllers of the users (which are displayed as a hammer or a riveting pliers depending on the role of the user) come into direct contact to perform the riveting. The users can feel the hit when hammering the rivet, which provides them passive haptic feedback for the collaborative task. *Left*: bird-eye view of the virtual environment. *Right*: view of the shared real environment.

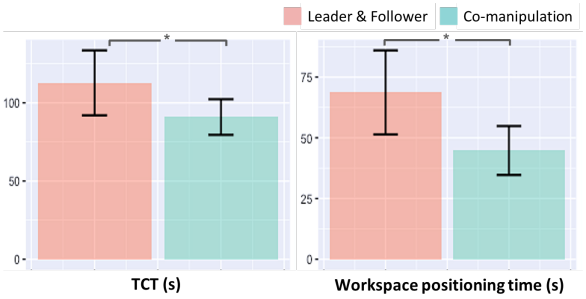

Figure 8: Mean *TCT* (left) and *workspace positioning time* (right) by technique. Error bars show 95% confidence intervals (CI)

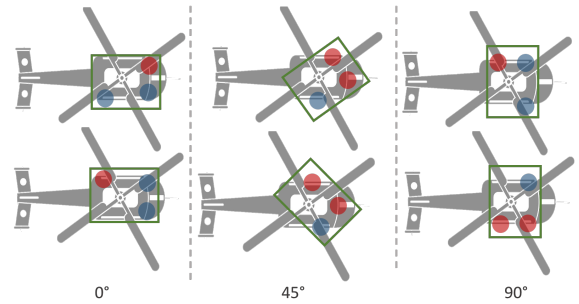

Figure 7: The six optimal virtual workspace positions for the six variations the riveting task. These variations included three types of rotations (0°, 45°, 90°) and different riveting positions.

Based on the size of the physical workspace and the riveting area, the co-manipulation starting and stopping values are set to 3m and 10m, respectively, and the stopping warning threshold was set to 8m.

4.5 Procedure

After arriving at the laboratory, each pair of participants received instructions on the task, signed an informed consent form, and filled out a demographic questionnaire. In each session, participants tried two conditions (techniques), a first one then the other as two sub-sessions, each sub-session including a set of trials. The order of presentation of the conditions was counterbalanced among the participants. At the beginning of each condition, the participants received a training trial. During this training, the experimenter was allowed to answer the participants' questions, if any. Then, the participants completed six trials in a randomized order with different targeted riveting positions, resulting in six optimal virtual workspace positions with three different rotations (0°, 45°, 90°), as illustrated in Figure 7. Participants filled out a questionnaire after each technique. At the end of the experiment, participants ranked the two techniques according to their preferences. The whole experiment lasted about 45 minutes, and the total VR exposure time was about 30 minutes on average.

4.6 Data collection

We registered 144 trials: 2 TECHNIQUEs $\times$ 6 repetitions $\times$ 12 pairs. For each trial, we logged the following measures:

- *Task Completion Time (TCT)* is the total time spent by the participants completing one trial. Measurements began at the start of the individual task and ended when the six rivets were installed.
- *Workspace positioning time* is the time spent by the participants positioning the virtual workspace. The measurement started when both participants entered the riveting area and ended when they teleported inside the newly defined workspace (by releasing the touch-pad of their controller). For each of these measurements, we summed the values if more than one workspace positioning operation was required during the trial.
- *Riveting time* is the time spent by the participants completing the riveting task. The measurement began when the participants were first teleported into the shared workspace and continued until all six rivet placements were completed.
- *Number of positions* is the number of times the participants positioned the virtual workspace in one trial to complete the installation of the six rivets.
- *Number of warnings* is the number of the warnings triggered by the participants due to a collision with the workspace borders during the team riveting.

We used the NASA-TLX questionnaire [20] to assess the cognitive task load. Participants were also asked to evaluate their leadership ("Who was the leader, you or your partner?"), contribution ("To what extent did you and your partner contribute to positioning the workspace?"), and talkativeness ("Who talked the most, you or your partner?"). Several previous studies have used similar questions to investigate leadership in collaborative tasks [32, 38]. Criteria were graded on a 21-point scale and later converted to a 100-point score.

4.7 Statistical results

We averaged the 6 repetitions of each technique to minimize the noise in the data. All statistical analyses were performed in R with a significance level of $\alpha = 0.05$ for all tests. Means (M) are reported with standard deviations.

For *TCT* (see Figure 8, left), we used Shapiro-Wilk test and QQ plots to analyze data normality. The data did not conform to normal distribution, so we applied a log transformation to it following the statistical recommendations [28] (p.316). The Kolmogorov's D-test then showed its goodness-of-fit to the log-normal distribution. We thus ran the analysis using the log-transform of *TCT*. The paired sample t-test revealed that the participants achieved the task significantly faster with *Co-manipulation* ($M = 90.90s \pm 17.95s$) than with *Leader-and-Follower* ($M = 112.81s \pm 34.00s$, p = 0.037) with an effect size of 0.57.

Regarding *workspace positioning time* (see Figure 8, right), we followed the same analysis procedure as applied to *TCT* and observed a significant difference of TECHNIQUES in the paired samples t-test (p = 0.008) with a large effect size of 0.81. It was shown that the participants spent significantly longer time positioning the workspace with *Leader-and-Follower* ($M = 68.70s \pm 28.46s$) than with *Co-manipulation* ($M = 44.74s \pm 15.87s$)).

Concerning *riveting time* (see Figure 9, left), a paired sample t-test was used as the data was normally distributed. We did not find any significant difference between *Co-manipulation* ($M = 29.48s \pm 5.58s$)) and *Leader-and-Follower* ($M = 35.66s \pm 11.59s$, p = 0.075).

For *number of positions* (see Figure 9, right), we used a non-parametric test in conformity with the nature of count data. Wilcoxon signed-rank test revealed that the participants positioned the workspace significantly more often with *Leader-and-Follower* ($M = 1.22 \pm 0.32$) than with *Co-manipulation* ($M = 1.04 \pm 0.10$), p = 0.025) with an effect size of 0.74.

For *number of warnings* (see Figure 10, left), we used a non-parametric test for post-hoc analysis in conformity with this type of data. Wilcoxon signed-rank test showed that the participants detected significantly more the warnings with *Leader-and-Follower* ($M = 0.85 \pm 0.50$) than with *Co-manipulation* ($M = 0.58 \pm 0.46$), p = 0.034) with an effect size of 0.55.

Regarding the subjective questionnaire, we used non-parametric Wilcoxon signed-rank tests for post-hoc analysis. For NASA-TLX score, we did not find any significant difference on cognitive task load between *Leader-and-Follower* ($M = 43.09 \pm 12.32$) and *Co-manipulation* ($M = 42.47 \pm 11.75$), p = 0.99). In *Leader-and-Follower*, no significant differences in talkativeness were found among participants in the leader role ($M = 42.92 \pm 16.71$) and follower role ($M = 48.33 \pm 18.50$), p = 0.506). Leaders ($M = 50.00 \pm 11.48$) were also found to have no significant differences in the level of contribution as followers ($M = 52.50 \pm 5.84$, p = 0.792). However, we detected an imbalance of leadership in *Leader-and-Follower* condition, with higher value for leaders ($M = 40.00 \pm 14.61$) compared to followers ($M = 57.92 \pm 17.64$), p = 0.015) with a large effect size of 1.11. Finally, 14 out of 24 participants preferred *Co-manipulation* over *Leader-and-Follower*. However, this result was not statistically significant according to the Binomial test (p = 0.541). Six participants out of 10 who preferred *Leader-and-Follower* took a leadership role in the experiment.

4.8 Discussion

The results provide evidence that *Co-manipulation* was more efficient than the *Leader-and-Follower* for workspace positioning. In *Co-manipulation*, the participants spent significantly less time on completing the task. In particular, the time used for negotiation and positioning the workspace was significantly decreased when both

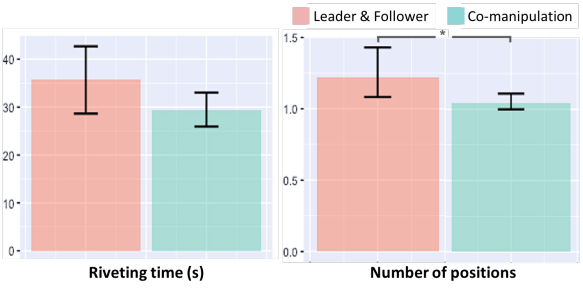

Figure 9: Mean *riveting time* (left) and *number of positions* (right) by technique. Error bars show 95% CI.

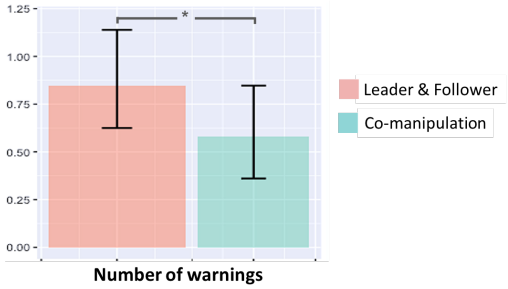

Figure 10: Mean *number of warnings* by technique. Error bars show 95% CI.

participants were able to manipulate the workspace, compared to the *Leader-and-Follower*. It therefore supports **H1**. This can be explained by the fact that each participant could adjust the position of the workspace according to the individual criteria provided to them. Besides, participants' desired positions and intentions could be communicated implicitly through the manipulation of the workspace. In the *Leader-and-Follower*, communicating the desired positions could be difficult, and we observed three different approaches that the participants used to achieve this. The participants could either (i) teleport themselves close to the required riveting point, (ii) use a ray to point to the target position, or (iii) use a virtual object (e.g., helicopter or avatar) as a reference to verbally describe its position. However, the use of additional teleportation or pointing, as well as the potential misunderstandings that could arise from the verbal communication, resulted in the need to use more time to exchange the information on riveting positions between the participants.

Contrary to our expectations, we were not able to find any significant difference between *Co-manipulation* and *Leader-and-Follower* in terms of *riveting time*, which does not support **H2**. However, we found that the participants triggered significantly more warnings and executed a significantly larger number of workspace positioning operations in *Leader-and-Follower* during the riveting process. Indeed, the accuracy of the workspace positioning affects the performance of the subsequent riveting tasks. If the user-defined workspace does not include all the required rivet positions, warnings will be triggered when users attempt to access a rivet outside the workspace. The virtual workspace then needs to be re-positioned to complete the task. Therefore, some minor accuracy differences may still exist between these two techniques.

Participants agreed that it was the leader who directed the workspace positioning in *Leader-and-Follower*. However, no significant difference in levels of verbal activity and contribution was found, which rejects **H3**. This can be explained by the fact that the individual position criteria were provided to the follower, which required them to actively join the workspace positioning task. We also did not detect any significant difference between *Leader-and-Follower* and *Co-manipulation* for cognitive task load. Although the *Co-manipulation* can be more efficient, a few participants also felt "out of control" (P4) in such a condition as the simultaneous manipulation inevitably produced "conflicts" (P11) during the virtual workspace positioning. This may also explain why a large number of participants, especially those in the leader role, preferred to use the *Leader-and-Follower* technique.

According to the participants' self-evaluation of their VR experience, the 12 pairs consisted of three novice-novice groups, four expert-novice groups and five expert-expert groups. In the *Co-manipulation* condition, the expert-expert groups ($M = 80.39s \pm 11.81s$) outperformed the expert-novice groups ($M = 98.92s \pm$

24.79s) and the novice-novice groups ($M = 97.94s \pm 9.77s$) with less average task completion time. However, the same result was not found in the *Leader-and-Follower* condition, as the expert-expert groups ($M = 121.88s \pm 45.68s$) needed more time to complete the task than the two other groups ($M = 106.80s \pm 29.34s$ for expert-novice groups and $M = 105.71s \pm 7.56s$ for novice-novice groups). It is difficult to draw formal conclusions based on such a limited number of pairs by groups, and further studies are needed. However, several explanations are still worth discussing. One is that the *Leader-and-Follower* technique could rely more on the approach the users use to achieve spatial information exchange than on their level of VR experience. Another is that when both users are experts, it is possible that the follower may not accept their status and therefore may challenge the leader's decisions more frequently, thus increasing the execution time (i.e., the follower has their own understanding of workspace management due to their VR expertise and would like to act as a leader).

5 Conclusion

This paper proposes and evaluates two interactive techniques that allow two co-located users to recover spatial consistency between their real and virtual workspace after individual teleportation. These recovery techniques use a virtual representation of the users' shared physical workspace in the VE. The *Leader-and-Follower* technique allows only one user to position the virtual workspace, while the *Co-manipulation* enables both of them to manipulate the virtual workspace at the same time. The recovered spatial consistency allows the users to access surrounding virtual objects through physical movements, enables direct physical interaction between them, and avoids the perceptual conflicts of dual presence of the users and their avatars in the virtual scene.

We conducted a controlled experiment to evaluate the performance of the two techniques in a collaborative virtual riveting task. The results showed that the positioning of the virtual workspace can be significantly faster with *Co-manipulation* than with *Leader-and-Follower*. In *Co-manipulation*, users' intents can be directly communicated by manipulating their future virtual workspace, shortening the time it takes to communicate its correct position and allowing an efficient subsequent collaborative task. In addition, significantly more attempts to reposition the virtual workspace and more warnings given when the users collided with the workspace borders during the collaborative task were measured in *Leader-and-Follower*. However, despite a better performance of the *Co-manipulation*, it also introduces conflicts in the way to position the virtual workspace. Moreover, it is sometimes difficult for users to understand their impact on controlling the final position and orientation of the workspace during the co-manipulation.

Nevertheless, *Leader-and-Follower* may be a relevant technique to reach spatial consistency when one of the users is well aware of all the requirements of the subsequent collaborative tasks. For example, in training or education application, the trainer may be responsible for placing the workspace to include all the required training contents for the following tasks. Moreover, this technique can also be applied for asynchronous collaborative interaction. In a collaborative system that uses tangible interaction, a user can define a workspace and leave a tangible prop in it. When other users arrive at the scene, they can continue to use that same prop within the workspace configured by the previous user. However, these two techniques should be further improved when a physical prop is used in the workspace. When recovering the spatial consistency between the users' real workspace and the virtual workspaces, the physical object must be paired with its virtual counterpart to allow the users to touch it when reaching. The physical object thus adds more constraints to the possible positioning of virtual workspace in the VE, even imposing a unique positioning solution.

Both of these techniques can be extended to collaborative tasks involving more than two users. However, the increasing number of users brings more conflicts and difficulties to concurrent manipulation. Therefore, further investigation is necessary to determine which technique is more appropriate for different numbers of users and in various collaboration scenarios. Moreover, different spring-damper values can be tested in the co-manipulation condition, giving users unbalanced control over the positioning of the virtual workspace and generating an alternative in-between the co-manipulation and the leader-follower approach. It could be thus interesting to investigate the impact of this asymmetric strategy on users during the collaborative virtual workspace positioning in some future studies.

Finally, these spatial consistency recovery techniques can also be used in some AR training applications. When using these AR setups, the user can see and hear other users in the real world from a different position and orientation than their avatars after individual virtual navigation, which can be confusing. Our techniques can solve this spatial desynchronization issue and correct the position and orientation mapping of the users for the collaboration phase.

Acknowledgments

This work was supported by French government funding managed by the National Research Agency under the Investments for the Future program (PIA) with the grant ANR-21-ESRE-0030 (CONTINUUM project).

References

[1] HTC Vive Pro Eye. `https://business.vive.com/fr/product/vive-pro-eye-office/`. Accessed: 2022-01-05.

[2] L. Aguerreche, T. Duval, and A. Lécuyer. Short paper: 3-hand manipulation of virtual objects. In *JVRC 2009*, pp. 4–p, 2009.

[3] L. Aguerreche, T. Duval, and A. Lécuyer. Comparison of three interactive techniques for collaborative manipulation of objects in virtual reality. In *CGI 2010 (Computer Graphics International)*, 2010.

[4] M. Azmandian, T. Grechkin, M. T. Bolas, and E. A. Suma. Physical space requirements for redirected walking: How size and shape affect performance. In *ICAT-EGVE*, pp. 93–100, 2015.

[5] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, 2013.

[6] C. W. Borst and A. P. Indugula. Realistic virtual grasping. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 91–98. IEEE, 2005.

[7] P. Bourdot and D. Touraine. Polyvalent display framework to control virtual navigations by 6dof tracking. In *Virtual Reality, 2002. Proceedings. IEEE*, pp. 277–278. IEEE, 2002.

[8] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3D User interfaces: theory and practice, CourseSmart eTextbook*. Addison-Wesley, 2004.

[9] J. Butterworth, A. Davidson, S. Hench, and M. T. Olano. 3dm: A three dimensional modeler using a head-mounted display. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics*, I3D '92, pp. 135—-138. Association for Computing Machinery, 1992.

[10] W. Chen, N. Ladeveze, C. Clavel, D. Mestre, and P. Bourdot. User cohabitation in multi-stereoscopic immersive virtual environment for individual navigation tasks. In *Virtual Reality (VR), 2015 IEEE*, pp. 47–54. IEEE, 2015.

[11] L.-P. Cheng, S. Marwecki, and P. Baudisch. Mutual human actuation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 797–805, 2017.

[12] G. Cirio, M. Marchal, T. Regia-Corte, and A. Lécuyer. The magic barrier tape: a novel metaphor for infinite navigation in virtual worlds with a restricted walking workspace. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp. 155–162. ACM, 2009.

[13] C. A. Ellis, S. J. Gibbs, and G. Rein. Groupware: some issues and experiences. *Communications of the ACM*, 34(1):39–58, 1991.

[14] C. Fleury, A. Chauffaut, T. Duval, V. Gouranton, and B. Arnaldi. A generic model for embedding users' physical workspaces into multi-

scale collaborative virtual environments. In *ICAT 2010 (20th International Conference on Artificial Reality and Telexistence)*, 2010.

[15] J. Frommel, S. Sonntag, and M. Weber. Effects of controller-based locomotion on player experience in a virtual reality exploration game. In *Proceedings of the 12th international conference on the foundations of digital games*, pp. 1–6, 2017.

[16] A. S. García, J. P. Molina, P. González, D. Martínez, and J. Martínez. An experimental study of collaborative interaction tasks supported by awareness and multimodal feedback. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry*, pp. 77–82. ACM, 2009.

[17] A. S. García, J. P. Molina, D. Martínez, and P. González. Enhancing collaborative manipulation through the use of feedback and awareness in CVEs. In *Proceedings of the 7th ACM SIGGRAPH international Conference on Virtual-Reality Continuum and Its Applications in industry*, p. 32. ACM, 2008.

[18] G. Gonzalez-Badillo, H. I. Medellin-Castillo, and T. Lim. Development of a haptic virtual reality system for assembly planning and evaluation. *Procedia Technology*, 7:265–272, 2013.

[19] M. J. Habgood, D. Moore, D. Wilson, and S. Alapont. Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 371–378. IEEE, 2018.

[20] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139 – 183. North-Holland, 1988.

[21] H. G. Hoffman. Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180)*, pp. 59–63. IEEE, 1998.

[22] A. Kulik, A. Kunert, S. Beck, R. Reichel, R. Blach, A. Zink, and B. Froehlich. C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments. *ACM Transactions on Graphics (TOG)*, 30(6):1–12, 2011.

[23] J. Lacoche, N. Pallamin, T. Boggini, and J. Royan. Collaborators awareness for user cohabitation in co-located collaborative virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, p. 15. ACM, 2017.

[24] J. M. Loomis, J. A. Da Silva, N. Fujita, and S. S. Fukusima. Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):906, 1992.

[25] D.-H. Min, D.-Y. Lee, Y.-H. Cho, and I.-K. Lee. Shaking hands in virtual space: Recovery in redirected walking for direct interaction between two users. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 164–173. IEEE, 2020.

[26] M. S. Pinho, D. A. Bowman, and C. M. Freitas. Cooperative object manipulation in immersive virtual environments: framework and techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 171–178. ACM, 2002.

[27] K. Riege, T. Holtkamper, G. Wesche, and B. Frohlich. The bent pick ray: An extended pointing technique for multi-user interaction. In *3D User Interfaces (3DUI'06)*, pp. 62–65. IEEE, 2006.

[28] J. Robertson and M. Kaptein. *Modern Statistical Methods for HCI*. Springer, 2016.

[29] R. A. Ruddle, J. C. Savage, and D. M. Jones. Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(4):285–308, 2002.

[30] H. Salzmann, J. Jacobs, and B. Froehlich. Collaborative interaction in co-located two-user scenarios. In *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pp. 85–92. Eurographics Association, 2009.

[31] H. Salzmann, M. Moehring, and B. Froehlich. Virtual vs. real-world pointing in two-user scenarios. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pp. 127–130. IEEE, 2009.

[32] A. Steed, M. Slater, A. Sadagic, A. Bullock, and J. Tromp. Leadership and collaboration in shared virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pp. 112–115. IEEE, 1999.

[33] M. Wang, S. L. Lyckvi, C. Chen, P. Dahlstedt, and F. Chen. Using advisory 3d sound cues to improve drivers' performance and situation awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2814–2825, 2017.

[34] T. Weissker and B. Froehlich. Group navigation for guided tours in distributed virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2524–2534, 2021.

[35] T. Weissker, A. Kulik, and B. Froehlich. Multi-ray jumping: comprehensible group navigation for collocated users in immersive virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 136–144. IEEE, 2019.

[36] T. Weissker, A. Kunert, B. Fröhlich, and A. Kulik. Spatial updating and simulator sickness during steering and jumping in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 97–104. IEEE, 2018.

[37] T. Weissker, P. Tornow, and B. Froehlich. Tracking multiple collocated htc vive setups in a common coordinate system. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 593–594. IEEE, 2020.

[38] J. Wideström, A.-S. Axelsson, R. Schroeder, A. Nilsson, I. Heldal, and Å. Abelin. The collaborative cube puzzle: A comparison of virtual and real environments. In *Proceedings of the third international conference on Collaborative virtual environments*, pp. 165–171, 2000.

[39] Y. Zhang, N. Ladeveze, C. Fleury, and P. Bourdot. Switch techniques to recover spatial consistency between virtual and real world for navigation with teleportation. In *International Conference on Virtual Reality and Augmented Reality*, pp. 3–23. Springer, 2019.

[40] Y. Zhang, N. Ladevèze, H. Nguyen, C. Fleury, and P. Bourdot. Virtual navigation considering user workspace: Automatic and manual positioning before teleportation. In *26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2020.

Accuracy of Deictic Gestures to Support Telepresence on Wall-sized Displays

Ignacio Avellino
Inria, Univ Paris-Sud & CNRS
avellino@lri.fr

Cédric Fleury
Univ Paris-Sud & CNRS, Inria
cfleury@lri.fr

Michel Beaudouin-Lafon
Univ Paris-Sud & CNRS, Inria
mbl@lri.fr

ABSTRACT

We present a controlled experiment assessing how accurately a user can interpret the video feed of a remote user showing a shared object on a large wall-sized display by looking at it or by looking and pointing at it. We analyze distance and angle errors and how sensitive they are to the relative position between the remote viewer and the video feed. We show that users can accurately determine the target, that eye gaze alone is more accurate than when combined with the hand, and that the relative position between the viewer and the video feed has little effect on accuracy. These findings can inform the design of future telepresence systems for wall-sized displays.

Author Keywords

Pointing; Telepresence; Wall-sized display; Remote collaboration

ACM Classification Keywords

H.5.3 [Group and Organization Interfaces]: Collaborative Computing; Computer-supported cooperative work

INTRODUCTION

Large interactive rooms with wall-sized displays help users manage the increasing size and complexity of data in science, industry and business. They naturally support co-located collaboration among small groups but can also be interconnected to support remote collaborative work. Video is critical to such remote collaboration as it supports non-verbal cues, turn-taking and shared understanding of the situation [3]. However, current telepresence systems are designed for meetings where users sit around a conference table and do not support spaces where users move around and work on shared data.

Our goal is to study telepresence systems that support remote collaboration across wall-sized displays by combining the shared task space with the shared person space [2]. The former refers to the task at hand and involves actions such as making changes, annotating and referencing objects; the latter refers to the collective sense of co-presence and involves facial expressions, voice, gaze and body language. Buxton [1] defines the overlap between these spaces as the *reference space*, where *"the remote party can use body language to reference the work"*. Our goal is therefore to study the reference space in the context of wall-sized displays.

We focus on telepresence systems linking two distant rooms with wall-sized displays showing the same content (Figure 1). Live video feeds of the users are captured by an array of cameras at eye level and displayed on the remote display at the corresponding position. Users can thus see the face of remote users and interact in a consistent way with the shared content.

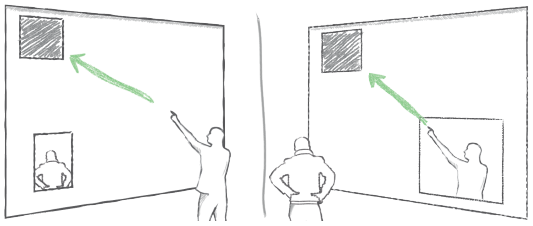

Figure 1. Users working with shared objects using a wall-sized display.

Referencing shared objects to support collaboration and mutual understanding is common when working together [6]. This paper investigates users' ability to accurately determine which shared object is referenced by a remote user without the need for dedicated technology such as telepointers.

We conducted a controlled experiment to study (a) how accurately a local user perceives a reference to a shared object performed by a remote user either by looking at it or by pointing their hand at it, and (b) whether the position of the local user in front of the wall-sized display affects this accuracy.

RELATED WORK

A number of systems, from the early VideoWhiteboard [9] and ClearBoard [4] to the more recent Connectboard [8] and Holoport [5], have explored how to combine person and task spaces with vertical or slanted displays, based on a glass metaphor where both spaces are overlaid. All these systems provide gaze awareness, the ability to notice direct eye contact and to notice what the remote person is looking at, and perception of which object the remote person is drawing or manipulating. However they are meant to be used by participants who do not move much in front of the display, and the accuracy of remote pointing has not been evaluated.

Nguyen & Canny [7] developed a telepresence system that uses a camera and projector per participant to create spatial faithfulness with multiparty conferencing. This system avoids the so-called *Mona Lisa effect*, where the image of

Ignacio Avellino, Cédric Fleury and Michel Beaudouin-Lafon. Accuracy of Deictic Gestures to Support Telepresence on Wall-sized Displays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI 2015)*, ACM, April 2015.

ISBN 9781450331456
https://doi.org/10.1145/2702123.2702448

a subject looking into the camera is seen by remote participants as looking at them, irrespective of their position. Wong & Gutwin [10] assessed pointing accuracy but used a Collaborative Virtual Environment, where users are represented by avatars instead of live video feeds.

In summary, while a number of telepresence systems have been proposed that enable remote collaboration on shared objects, very few studies have assessed the accuracy of designating such objects remotely through pointing or gazing.

EXPERIMENT

We conducted a controlled experiment to assess how accurately an observer can determine which object a remote user is showing on a wall-sized display. The remote user shows a target and the observer must determine which one it is. We use pre-recorded videos of the remote user and display them on the wall-sized display in front of the observer at the same position where the recording camera was placed.

We control three factors: how the remote user specifies the target (by turning the head or by turning the head and pointing at it), the position of the target relative to the video displayed on the wall (19 positions) and the position of the observer in front of the wall-sized display (5 positions).

Experimental Setup

The 19 targets are displayed on a $5.5m \times 1.8m$ wall-sized display made of a grid of 8×4 30" monitors. Each target is a black letter on a white background surrounded by a blue circle. We exclude letters that could be confused, such as O and Q. We use a concentric radial distribution of targets in order to control for both distance and angle to the video. Three rings of targets surround a central one (*ring0*), where the video is displayed (Figure 2). The first two rings (*ring1* and *ring2*) have 8 targets, one for each cardinal and diagonal direction. Due to the aspect ratio of the wall-sized display, the third ring (*ring3*) has only two targets. *Ring0, ring1* and *ring2* are 11.5° apart when measured from the viewing position, while *ring2* and *ring3* are 23° appart.

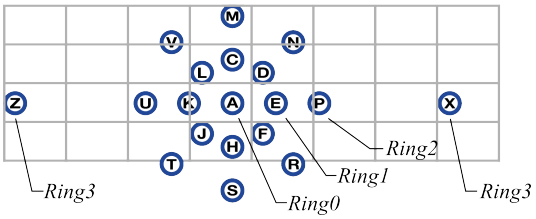

Figure 2. The 19 targets laid out in 4 rings on the wall-sized display.

Video Recording

We recorded 114 10-second videos of three different actors showing the 19 targets on a wall-sized display in both conditions (*head* and *pointing*): a 29 year-old woman with pulled back hair and brown eyes, a 29 year-old man with brown medium-length hair and brown eyes and a 27 year-old man with short brown hair and hazel eyes. The camera was set in front of the wall-sized display, at the position of the central target. The actors were placed $230cm$ away from the camera so that the pointing hand was within the recorded frame in all pointing directions. The actors were instructed to point or look successively at each target.

Video Playback

The participants in the experiment sat in front of the same display used for video recording, $230cm$ away from the display. The recorded videos were displayed on top of the central target, at the same position as the recording camera. Based on the focal length used at recording time ($43mm$ in $35mm$ equivalent focal length), we adjusted the size of the video so that the remote users appeared life-size, as if they were sitting $230cm$ behind the display, i.e. $460cm$ away from the participants. The height of the chair was adjusted for each participant so that the video was at eye-level.

Participants

12 right-handed participants (8 male), aged 21 to 33 (median 27), all computer science graduates, participated in the study. All participants had normal or corrected to normal vision. 10 had a right dominant eye, 2 a left one. 2 participants used conferencing systems every day, 6 more than once a week, 3 once a week, and 2 almost never. All participants received sweets as compensation for their time.

Task

Participants watch each video playing in an infinite loop. When they are ready to answer, they tap a large "Stop" button on an iPad 3 tablet and answer which target was being shown, by tapping the target on a replica of the display layout.

Procedure

The within-subject design has the following factors:

- TECHNIQUE used to indicate the targets, with 2 conditions: *head*, the natural combination of head turning and gazing, and *pointing*, the combination of head turning, gazing and pointing the target with the arm and finger;
- POSITION of the participant in front of the display, with 5 conditions: *center*, located in front of the videos, *left* (resp. *farLeft*), located 1m (resp. 2m) to the left, and *right* (resp. *farRight*), located 1m (resp. 2m) to the right;
- ACTOR: we recorded 3 sets of videos with 3 different actors to ensure that the choice of the remote person does not have an effect;
- TARGETS: 19 targets were used, a central one surrounded along 8 directions by 3 rings of targets with only the left and right targets on the last ring ($19 = 1 + 8 \times 2 + 2$).

For each participant, the conditions were grouped by TECHNIQUE, then by ACTOR and then by POSITION. The order of presentation was counterbalanced across conditions by using Latin squares for the first three factors and a randomized order for TARGET. Each Latin square was mirrored and the result was repeated as necessary. For each TECHNIQUE×ACTOR×POSITION condition, the order of the 19 targets was randomized so that successive videos never showed targets in adjacent rings from the same direction.

For training, we used the same subset of 12 videos covering all directions and distances for each participant to practice the task and the entry of answers. 4 different random positions were used (2 for the *head* condition and 2 for the *pointing* condition) with 3 videos each. Then, the 570 videos were presented: 2 TECHNIQUEs x 3 ACTORs x 5 POSITIONs x 19 TARGETs. A mandatory break was held every 190 trials, corresponding to 2 sets of 5 positions, and a reminder of an optional break was provided every 95 trials.

For each trial, we collected the answer from the participants, i.e. which target they thought was being shown. At the end of the experiment, participants filled out a short questionnaire.

Results

We measure the accuracy of the participants in determining the target shown by the remote users by two types of errors: Distance error, the number of rings between the actual target and the target chosen by the participant; angle error, the difference between the angles of the two targets, as a multiple of 45°. For angle errors, we remove trials where the indicated target is the central one, since it has no meaning in this case.

We find a small learning effect[1] of TECHNIQUE on distance error: it significantly decreases from 0.36 for the first technique to 0.29 for the second one ($F(1, 6827) = 39.19, p < 0.0001$). Many participants learned which videos correspond to the farthest targets and used that information to assign intermediate targets to subsequent videos, based on the angle of the head or arm being smaller than that of the farthest targets. We did not find a learning effect on the angle error ($F(1, 6467) = 3.58, p = 0.059$).

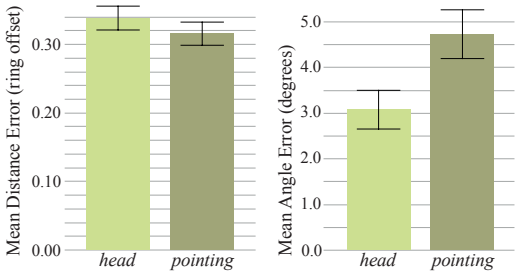

Figure 3. Distance and angle error by TECHNIQUE (bars indicate CI)

Figure 3 shows the mean and 95% Confidence Intervals (CI) of distance and angle error by TECHNIQUE. The multiway ANOVA with REML shows no significant difference in distance error between the two TECHNIQUEs ($F(1, 6817) = 0.87, p = 0.35$), but the difference in angle error is significant ($F(1, 6459) = 39.37, p < 0.0001$), with a mean of 3.10 for *head* vs. 4.72 for *pointing*. Surprisingly, using the arm led to higher angle errors than using only the head.

Figure 4 shows the mean and CI of distance and angle error by POSITION. The multiway ANOVA with REML shows no

[1]Statistical analyses were performed with SAS JMP. We performed Mixed Model REML (Restricted Maximum Likelihood) analyses with "participant" as a random factor.

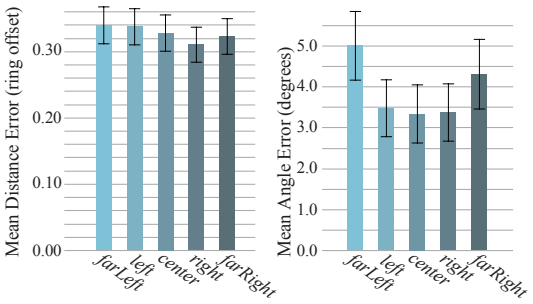

Figure 4. Distance and angle error by POSITION (bars indicate CI)

significant differences in distance error among the five POSITIONs ($F(4, 6817) = 0.95, p = 0.44$), but a significant difference for angle error ($F(4, 6459) = 3.84, p = 0.0040$), with means (from left to right) 5.00, 3.47, 3.33, 3.37, 4.31. A Tukey HSD post-hoc test reveals two groups of positions significantly different from each other: {*farLeft*, *farRight*}, {*center*, *left*, *right*, *farRight*}.

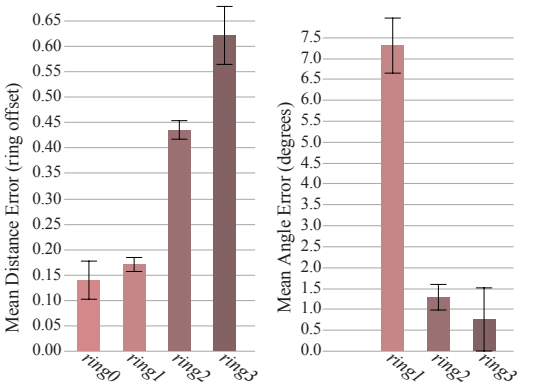

Figure 5. Distance and angle error by target ring (bars indicate CI)

Figure 5 shows the mean and CI of distance and angle error for each ring of targets. The multiway ANOVA with REML shows significant differences for distance ($F(3, 6817) = 294.14, p < 0.0001$) and angle ($F(2, 6459) = 165.01, p < 0.0001$) error (as mentioned before, *ring0* was removed for the angle error analysis). Distance error means (from *ring0* to *ring3*) are: 0.14, 0.17, 0.44, 0.62. Angle error means (from *ring1* to *ring3*) are: 7.30, 1.28, 0.75. A Tukey HSD post-hoc test reveals three significantly different groups of rings for distance error: {*ring0*, *ring1*}, {*ring2*}, {*ring3*}. For angle error, {*ring1*} is significantly different from {*ring2*, *ring3*}.

Discussion

According to our results, distance error is not significantly different when using the head or head+arm, but the angle error is larger when using the head+arm. While the difference is small, this is a surprising result. By analyzing the videos, we noticed that most of the time the direction of the arm does not indicate the target. This is because users place the tip of their finger on the line of sight between their eye and the target. In the post-hoc questionnaires, we found that 4 participants used

only the arm direction when determining the pointed target; 6 used first the arm and when in doubt looked at the eyes and head direction; only one participant determined the correct location by connecting the eyes with the tip of the finger. Because the arm is the most salient cue in the video, it is likely that users use it as the primary source for determining the pointed target, ignoring the geometrical interpretation that we perform in a face-to-face environment.

The fact that the position of the participant relative to the video feed has little or no effect was also surprising. We did not expect this effect, analogous to the *Mona-Lisa effect*, to be so strong. The extreme positions, *farLeft* and *farRight*, had a higher angle error, which can be explained by the fact that the observers are looking at the image with an angle of 49°, making the task harder.

We found that distance error increases when the targets are further away from the video feed. This may be due to the fact that the videos are shot from the front. In this setting, a small change of angle, e.g. 10°, when pointing near the center produces a noticeable change in distance between the shoulder and the finger in the 2D projection of the arm. The same change in angle when the arm is pointing at the last ring of targets results in a much smaller distance between the finger and the shoulder in the 2D projection of the arm, making it harder to notice. Angle error, on the other hand, decreases when the targets are further away from the video feed. This is due to the radial layout: the distance between targets on an inner ring is smaller than on an outer ring, resulting in larger variations in the direction of the hand and arm. It is interesting to note that on the farthest targets, for which it was only possible to express distance, users still made angle errors because they thought that the target laid on *ring2*.

CONCLUSION AND FUTURE WORK

We investigated the ability to accurately determine which shared object a remote user is referencing when sharing data on a wall-sized display. We found that showing objects only with the head leads to smaller angle error than with the head and arm. We also found no effect of the observer's position on accuracy, except at the farthest positions for angle error.

However, while distance accuracy decreases when the object is further away from the video of the remote user, angle accuracy increases with object distance. We attribute this effect to the fact that when capturing the remote user's image from the front, some body movements are more salient than others, conveying cues that help users determine more accurately the distance of close targets and the direction of far targets.

This study has three main implications for design:

1. Users can accurately estimate which object is being indicated only by looking at the head on the remote video, without requiring explicit pointing actions nor telepointers;
2. Therefore even when their hands are busy, users can point using their head without losing accuracy, supporting, e.g., the use of deictic instructions when holding tools;
3. The position of the video feed relative to the observer is not critical for accuracy when indicating remote objects, thus it can be moved around without loss of accuracy; this allows the system to match the video with the camera position on the remote side, maintaining spatial relationships with shared objects on both sides.

In future work we plan to investigate the body cues used to indicate an object and how to convey them over a telepresence system. We also plan to apply a similar methodology to situations that involve eye contact and understand how video distorts the perception of simple communicative acts. Finally we want to apply these findings to a functional system that supports video communication in large interactive rooms.

ACKNOWLEDGMENTS

We thank Wendy Mackay for discussions about the experimental design and statistical analysis. This work is supported by French National Research Agency grant ANR-10-EQPX-26-01 "Digiscope".

REFERENCES

1. Buxton, W. Mediaspace–meaningspace–meetingspace. In *Media Space 20 + Years of Mediated Life*, Computer Supported Cooperative Work. Springer, 2009, 217–231.
2. Buxton, W. A. S. Telepresence: Integrating shared task and person spaces. In *Proc. Graphics Interface, GI'92* (1992), 123–129.
3. Isaacs, E. A., and Tang, J. C. What video can and cannot do for collaboration: A case study. *Multimedia Systems 2*, 2 (1994), 63–73.
4. Ishii, H., and Kobayashi, M. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proc. Human Factors in Computing Systems, CHI'92*, ACM (1992), 525–532.
5. Kuechler, M., and Kunz, A. HoloPort - a device for simultaneous video and data conferencing featuring gaze awareness. In *Proc. Virtual Reality, VR'06*, IEEE (2006), 81–88.
6. Mackay, W. E. Media spaces: environments for informal multimedia interaction. In *Computer Supported Co-operative Work*, M. Beaudouin-Lafon, Ed., John Wiley & Sons (1999), 55–82.
7. Nguyen, D., and Canny, J. MultiView: Spatially faithful group video conferencing. In *Proc. Human Factors in Computing Systems, CHI'05*, ACM (2005), 799–808.
8. Tan, K.-H., Robinson, I., Samadani, R., Lee, B., Gelb, D., Vorbau, A., Culbertson, B., and Apostolopoulos, J. Connectboard: A remote collaboration system that supports gaze-aware interaction and sharing. In *Workshop on MMSP '2009*, IEEE (2009), 1–6.
9. Tang, J. C., and Minneman, S. VideoWhiteboard: Video shadows to support remote collaboration. In *Proc. Human Factors in Computing Systems, CHI'91*, ACM (1991), 315–322.
10. Wong, N., and Gutwin, C. Where are you pointing?: The accuracy of deictic pointing in CVEs. In *Proc. Human Factors in Computing Systems, CHI'10*, ACM (2010), 1029–1038.

CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays

Ignacio Avellino **Cédric Fleury** **Wendy E. Mackay** **Michel Beaudouin-Lafon**
LRI, Univ. Paris-Sud, CNRS,
Inria, Université Paris-Saclay
F-91400 Orsay, France
{avellino, cfleury, mackay, mbl}@lri.fr

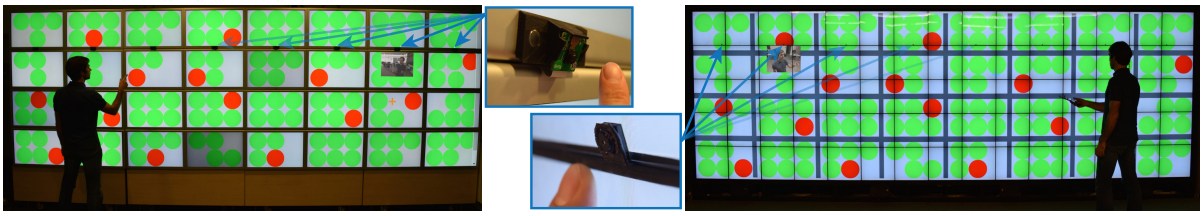

Figure 1: The WILD (left) and WILDER (right) wall-sized displays running *CamRay*, and close-ups on the cameras (center).

ABSTRACT

Remote collaboration across wall-sized displays creates a key challenge: how to support audio-video communication among users as they move in front of the display. We present *CamRay*, a platform that uses camera arrays embedded in wall-sized displays to capture video of users and present it on remote displays according to the users' positions. We investigate two settings: in *Follow-Remote*, the position of the video window follows the position of the remote user; in *Follow-Local*, the video window always appears in front of the local user. We report the results of a controlled experiment showing that with *Follow-Remote*, participants are faster, use more deictic instructions, interpret them more accurately, and use fewer words. However, some participants preferred the virtual face-to-face created by *Follow-Local* when checking for their partners' understanding. We conclude with design recommendations to support remote collaboration across wall-sized displays.

ACM Classification Keywords

H.5.3 [Information Interfaces and Presentation] (e.g. HCI): Group and Organization Interfaces - Computer-supported cooperative work; H.4.3 [Information Systems Applications]: Communications Applications - Computer conferencing, teleconferencing, and videoconferencing

Author Keywords

Telepresence, Remote collaboration, Wall-sized displays, Camera array

CHI 2017, May 06-11, 2017, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/3025453.3025604

INTRODUCTION

Compared with desktop displays, high-resolution wall-sized displays let users physically navigate large amounts of data [2], scan and find objects more easily [18], and use spatial memory to move and classify objects more efficiently [20]. Because of their large size, such displays also support *colocated* group work: users can understand and be aware of what their colleagues are doing, enabling tightly-coupled collaboration [19]. But how can we support such collaboration in *remote* settings? Current tools for remote collaboration are designed for users sitting at a desk or a video conferencing table, and do not scale to large interactive spaces where users can move.

Our challenge is to create a system in which remote users of wall-sized displays can interact easily with each other, as well as with data on the wall. Most remote collaboration systems use video as an effective surrogate for face-to-face conversation. However, in wall-sized display environments where users move, it is unclear where to position cameras and where to display the captured video.

We introduce *CamRay*, a platform that captures video of users as they move in front of wall-sized displays using camera arrays, and presents this video on remote walls. Using available hardware and open-source platforms, we add cameras to existing large wall-sized displays, stream video to a remote site in real time and display it according to users' position.

We first examine related work and describe an observational study that informed the design of *CamRay*, which we then present in detail. We report on an experiment where pairs of users worked on a data manipulation task, while we manipulate the position of each other's video. We conclude with implications for the design of systems for remote collaboration on wall-sized displays and discuss directions for future work.

Our contribution is twofold: 1) *CamRay*, a platform that captures and displays users as they move in front of wall-sized displays; 2) an experiment that shows how the position of the video feed affects collaboration and communication.

RELATED WORK

We first review Clark's work on communication as it provides a theoretical basis for our work. We then review previous work on the use of video for remote collaboration and systems that support remote collaboration across large interactive spaces.

Technology-mediated Communication

According to Clark [9], communication is characterized by a series of messages between parties which, once understood, become part of their *common ground*: the mutual knowledge, beliefs and assumptions shared by partners in communication [9, 10]. Common ground is updated through *grounding*, the collective process by which participants try to reach mutual belief that what has been said has been understood [8]. This process has a cost in technology-mediated communication, which Clark characterizes and defines [8], e.g., *start-up* cost to establish communication, *delay* and *asynchrony* when it is not purely real-time. Telepresence systems must take these costs into account and attempt to reduce them in order to support effective video-mediated communication.

Personal Video for Remote Collaboration

Previous work has shown the benefits of personal video for remote collaboration. According to Isaacs & Tang [15], a *"video channel adds or improves the ability to show understanding, forecast responses, give non-verbal information, enhance verbal descriptions, manage pauses and express attitudes."* Veinott et al. [31] have shown that seeing each other's faces in collaborative tasks improved the negotiation of common ground, as opposed to using non-video media. Monk & Gale [23] observed that having mutual gaze awareness provides an alternative to non-linguistic channels for awareness of a remote person's understanding.

Previous work on Media Spaces [4, 22] has leveraged the power of video-mediated communication by creating systems that support peripheral awareness, chance encounters, locating colleagues and other social activities. While Media Spaces have also been used to support focused remote collaboration, they have not emphasized settings where distributed groups work on shared data in large interactive spaces. Our work extends the concept of Media Spaces to such settings.

A number of remote collaboration systems have used video to convey more than just people's faces. Hydras [27] keep spatial relations among remote participants consistent in a multi-party conversation; VideoDraw [30] and VideoWhiteboard [29] show the shadow of the remote participant overlaid with the shared space they can draw on. Clearboard [16] expands on this idea by overlapping personal video with a shared task space. More recently, Nguyen & Canny [24] and Bos et al. [5] have explored how trust formation in video conferencing is affected by spatial distortions and communication channels. Nguyen & Canny [25] also showed how video framing affects empathy. Although previous systems have explored video as a tool to support remote collaboration, they depend on a static user sitting in front of a display. This does not scale to large interactive spaces that support physical navigation.

Personal video has also been used to provide remote awareness in what Buxton calls the *reference* space [6], by integrating the shared *person* and *task* spaces. Three's Company [28] implements this space by positioning the shadows of the users' arms on top of shared content, while ImmerseBoard [14] deforms the user's arm to place it on top of the content. Although the benefits of reference spaces for video-mediated communication have been demonstrated, research has not focused on wall-sized displays. We believe that creating reference spaces can greatly enhance collaboration across large displays, since they provide ample real-estate to integrate content and people.

Remote Collaboration in Large Interactive Spaces

To support remote collaboration across 3D virtual environments, Beck et al. [3] capture users through depth cameras and present them using a realistic 3D reconstruction in an immersive virtual environment. Willert et al. [32] use a 2D array of cameras mounted on the bezels of the screens of a wall-sized display to capture video. They provide an extended window metaphor between two sites, but do not study communication. Dou et al. [11] place RGB and depth cameras on wall-sized displays to capture two remote sites and create a room-sized telepresence system. The goal of this type of systems is to display remote video using the available screen space. They let people see each other and engage in conversation, but make it hard to collaborate on shared objects.

Luff et al. [21] proposed a high-fidelity telepresence system that supports remote collaboration on shared digital objects. Participants engaged in different formations, allowing them to meet the requirements of the task at hand, such as pointing to objects or talking to collaborators. This was possible because physical relations between video and digital objects was kept intact, such that, when users looked or pointed at objects, others knew what they were referring to. Although this system allows to collaborate on digital shared content, our focus is on large interactive spaces where people can walk.

We believe that remote communication on wall-sized displays can benefit from keeping physical relations between people and objects faithful as in a remote site. Avellino et al. [1] used this strategy in a large interactive space and showed that video on wall-sized displays can be used for accurately interpreting deictic gestures: placing video relative to content allows to accurately understand remote indications of shared digital objects. Nonetheless, it does not ensure that people will be able to see each other's face when collaborating, since they move and might be far from the video.

In summary, previous research has shown that video supports remote collaboration in various settings, but wall-sized displays have received little attention.

OBSERVATIONAL STUDIES

Our goal is to create a system that supports remote collaboration across wall-sized displays, so that users are able to communicate easily with each other over audio-video links

while working on shared content. To inform the design of this system, we created low-fidelity prototypes and conducted several observations. We simulated two remote sites by dividing a wall-sized display with a curtain, to simplify the setup.

In the first prototype, we asked two collaborators to put together a slide show presentation based on text and images from a presentation they had recently worked on. On each side, blank sheets of paper (for blank slides), text clippings, and images were laid out on the display. Two helpers held tablets running a videoconferencing software, enabling collaborators to see each other. We simulated shared content by manually syncing changes between the two sides (Fig. 2-top).

In this session, participants looked at the content on the wall-sized display much more than at each other's face on the tablets. They only looked at each other when they disagreed, and when they discussed the meaning of objects or where they should be placed. Based on the observation and debriefing, we hypothesized that the participants would have looked at the video feeds more often if they had been located on the wall-sized display, close to the content they were working on.

In the second low-fidelity prototype, we displayed the video feeds on the wall-sized display. We set up two cameras on each side: a front camera attached to the bezels of the display, and a back camera placed at the back of the room, facing the display. In this way, we could capture both the face and the back of participants. We asked two collaborators working on a publication to sort their related work using a Wizard-of-Oz prototype application built for this purpose. PDFs of scientific papers were laid out on the two sides of the display; their position and current page were synchronized. Each user had three video feeds (Fig. 2-bottom): on their left, the remote person's front camera feed; right below, a smaller feed of their own front camera; and, on their right, a feed of the remote back camera. These video feeds had fixed position and size, making it easy to determine when participants looked at them.

In this second session, participants physically moved to a specific video feed according to the task they were working on. They used the front-facing camera feed for discussions and arguments about the content of a particular paper, or how to cluster it. They used the back-facing camera feed when interpreting references to objects and locations and to maintain *common ground*—mostly through deictic instructions, e.g., *"this one should go there"*. In other words, *conversational communication* was best supported by the front-facing video and *gestural communication* by the back-facing video. However, participants had to stop what they were doing to move in front of the fixed video feeds. This interrupted their work and was perceived as annoying.

Based on these observations, it became clear that we needed to capture the users' faces as they moved in front of the display and present the video feeds in a flexible way. We identified two approaches to place the video feeds:

- close to each user, to support face-to-face conversation; or,
- matching the remote user's position, to understand where the user is looking and pointing.

Figure 2. Early observations: assembling a slide-show on a paper prototype with tablets streaming video (top). Sorting related work with video on the wall-sized display (bottom).

Based on our observations, we believe that the second solution, where the video feed is placed in the context of the shared content, will:

- support the use of deictic instructions in manipulation tasks;
- increase efficiency when manipulating content; and,
- be preferred to other video placements.

However, since we also observed the value of face-to-face communication supported by the first approach, we wanted to create a system that supports both approaches in order to compare them. In addition, while we only observed pairs of users with one user per site, this approach should scale to more than one user at each site, as well as to more than two sites.

CAMRAY: A CAMERA ARRAY FOR TILED DISPLAYS

We created *CamRay*, a system that adds telepresence capabilities to wall-sized displays. Our prototype links two interactive rooms with large tiled displays on our campus:

- *WILD* consists of an 8×4 grid of 30" LCD screens with 22mm top, left and right bezels and 30mm bottom bezels. It measures $5.5\text{m} \times 1.8\text{m}$ for a resolution of $20\,480 \times 6\,400$ pixels. It is controlled by a cluster of 16 Apple Mac Pros running Linux, each managing two screens.
- *WILDER* consists of a 15×5 grid of 21.6" LCD screens with 3mm bezels. It measures $5.9\text{m} \times 2\text{m}$ for a resolution of $14\,400 \times 4\,800$ pixels. It is controlled by a cluster of 10 PCs running Linux, each managing a row of 7 or 8 screens.

Both wall-sized displays are equipped with a *VICON* infrared tracking system used to track users with 6 degrees of freedom.

We mounted 8 cameras on each display to capture the users' faces as they move: one camera per column of monitors in *WILD* (8 in total) and 8 equally-spaced cameras in *WILDER*. For consistency, we placed the cameras proportionally to the overall horizontal size of each display. On *WILD*, the cameras are standard Raspberry Pi *Camera Modules*, placed on the bezels (Fig. 1-left). On *WILDER*, because of the thinner screen bezels, the cameras are smaller *Spy Cameras for Raspberry*

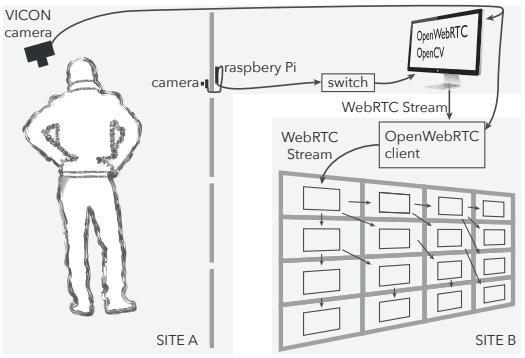

Figure 3. *CamRay* architecture to communicate from site A to site B. A similar setup is used to communicate from site B to site A.

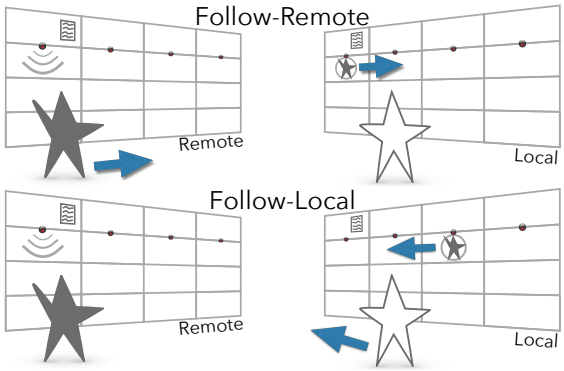

Figure 4. The two video modes. The arrows show which participant controls the position of the video displayed at the local site.

Pi[1] (8.5mm × 11.3mm) (Fig. 1-right). The cameras are placed on the nearest bezel above eye level on both displays. Each camera is directly connected to a *Raspberry Pi*[2] board mounted onto the back of the displays, and the cables are slipped though the gap between two adjacent screens (Fig. 3). The boards are connected using an Ethernet network to a dedicated computer (desktop Mac Pro) that processes the 8 video streams.

Each *Raspberry Pi* captures and encodes video in H.264[3] and streams it to the dedicated computer over UDP using *GStreamer*[4], an open source multimedia framework. The videos are captured at 30 frames per second with a resolution of 800 × 600 pixels to avoid overloading the main computer.

The dedicated computer runs a custom C++ application based on *OpenCV*[5] and *OpenWebRTC*[6]. It uses *OpenCV* to receive all the video streams and automatically select the one that corresponds to the camera in front of the user. To achieve this selection, the custom C++ application receives the user position data sent by the *VICON* infrared tracking system using *Open Sound Control* messages. This application uses the *OpenWebRTC* library to stream the selected video to the remote wall-sized display using the WebRTC protocol, which supports video over firewalls and large-area networks.

At the other location, an *OpenWebRTC* relay server receives the remote video stream and transmits it to the node of the visualization cluster that runs the top-left screen of the wall-sized display. A web application based on *NW.js*[7] runs on each node of the cluster. These applications can display the WebRTC video stream that they receive, either from the relay server or from another node, and they can also forward the stream to 2 or 3 others nodes. Using a tree pattern (Fig. 3), all the nodes of the cluster can receive the video stream with low latency. In our experience, this approach is much better than overloading the main server by having it send the video feeds to all the cluster nodes at the same time.

Once all the nodes receive the stream, the video of the remote user can be displayed and moved all over the tiled wall-sized display, spanning several screens if necessary. The video window can be displayed on top of any application that may be running on the wall. The relay server notifies all the nodes of the cluster about the position of the video window so that each node can decide whether to display the video or part of it on their associated screens. In addition, the relay server receives the position of the local and remote users through *WebSockets*, and it uses this information to compute the position of the video window according to the video mode (see below).

CamRay can easily be adapted to a variety of tiled wall-sized displays: the number of cameras can be adapted to the size of the display, and the tree pattern used to distribute the stream can scale to larger clusters. If the main computer becomes overloaded because of the larger number of cameras or higher-quality video, the load can be distributed over several computers, each one connected to a subset of the cameras. The WebRTC protocol traverses almost any network, making it possible to connect to diverse sites. *CamRay* can support more than one user per site since users are identified by the *VICON* system. *CamRay* also scales to multiple remote locations because the relay server can simultaneously receive several video streams from different WebRTC connections. In such multi-user, multi-site configurations, *CamRay* would send one video stream per user and per site, which is still much less than the total number of cameras per site.

Positioning Video Feeds

We implemented the two video modes described in the previous section (Fig. 4):

- *Follow-Local*: the video window follows the horizontal position of the *local* user, i.e. the local user has the video window always in front of her. This mode supports, e.g., a face-to-face conversation even if the users are standing at different positions relative to their display.
- *Follow-Remote*: the video window follows the horizontal position of the *remote* user. This mode makes it easy to interpret deictic references made by the remote user since

[1] https://www.adafruit.com/products/1937
[2] https://www.raspberrypi.org/
[3] We use the *raspivid* command included with *Raspbian*
[4] https://gstreamer.freedesktop.org/
[5] http://opencv.org/
[6] http://www.openwebrtc.org/
[7] http://nwjs.io/

the video feed has the same position relative to the shared content as the remote user.

In both modes the video of the remote user is mirrored horizontally to ensure spatial consistency: when a user looks to the left, she is displayed as looking to the left in the video window at the remote location. In other words, the remote user is seen as standing behind the wall-sized display, as in Clearboard [16]. *CamRay* mirrors the video when capturing it.

Face-to-face conversation greatly benefits from eye contact. In video-mediated conversation, eye contact requires that the video feed be close to the camera. *CamRay* supports arbitrary positioning of the video window, but as recommended by Chen [7], by default it positions the video window right below the closest camera in the camera array. As a result, the video window jumps among 8 discrete positions.

Finally we do not show feedback of the users' own video, unlike most desktop videoconferencing systems. Surprisingly, nobody asked for it in our observations. Some participants reported that they trust the system to capture them properly, since they do not have to adjust a webcam position as in standard desktop videoconferencing systems.

The two video modes can scale to multiple users and multiple sites. *Follow-Remote* can simply display the remote users at their remote positions. In case of overlap, the system can either use transparency or a simple physics engine to avoid collisions. For *Follow-Local*, the system can lay out remote users side by side or in a half-circle, consistently across sites, in the same way as in some multiparty videoconferencing systems.

EXPERIMENT

In order to assess the effects of video position on communication and the trade-offs it incurs, we ran a controlled experiment comparing three ways to display a remote collaborator video on wall-sized displays:

- *Follow-Remote*: the video windows appears on the wall at the same position as the remote collaborator;
- *Follow-Local*: the video windows appears on the wall in front of the local user; and
- *Side-by-Side* (control condition): the fixed video window appears on a separate screen, perpendicular to the wall.

In our observations, deictic gestures were better supported when the video was placed in the context of the shared content. Therefore we formulate the following hypotheses:

- H1: participants use more deictic instructions in *Follow-Remote* than *Follow-Local* and *Side-by-Side*;
- H2: participants manipulate data more efficiently in *Follow-Remote* than in *Follow-Local* and *Side-by-Side*; and,
- H3: *Follow-Remote* is preferred for manipulation tasks when giving and receiving instructions.

Method

The $[3 \times 2]$ within-participant design has two primary factors and a secondary factor:

- VIDEO (*Follow-Local*, *Follow-Remote*, *Side-by-Side*);
- LAYOUT (*Medium* and *Hard*); and
- ROLE (*Instructor* and *Performer*).

LAYOUT controls the difficulty of the task, while ROLE accounts for the asymetry of the task, as described below. The order of VIDEO conditions is counterbalanced across pairs using a balanced Latin Square; the order of LAYOUT and ROLE are counterbalanced for each VIDEO condition.

Participants

We recruited 12 pairs of participants, aged between 23 and 40, all with normal or corrected-to-normal vision, none color blind. Couples were formed as participants were recruited, leading to 9 male-male, 2 female-male and 1 female-female couples. 1 participant used video conferencing systems on a daily basis, 8 on a weekly basis, 6 on a monthly basis, 5 on a yearly basis and 4 almost never.

Hardware and Software

The setup of the experiment is composed of the two wall-sized displays, *WILD* and *WILDER*. *Follow-Local* and *Follow-Remote* conditions are implemented with *CamRay* (Fig. 1). The video windows of the remote users move horizontally at a fixed height of 1.75m (center of the window) at both sites. In *Side-by-Side*, video is displayed on an LCD screen on the left side of the room, at approximately the same height as the window in the other conditions. In all three VIDEO conditions, the video windows have the same size ($34.7\text{cm} \times 26\text{cm}$).

Although *WILD* and *WILDER* have different sizes and resolutions, we scale the content so that it spans the entire display. We use *Webstrates* [17] to create and synchronize content. Participants interact with each wall-sized display with a cursor controlled by a handheld pointer through raycasting. The pointer is mounted on a smartphone that displays a virtual button for picking and dropping. The orientation and position of the pointers and of the participants' heads are tracked using the VICON tracking systems in each room.

Procedure

Task Description

During our observations, participants often referred to on-screen objects by pointing and looking at them. To assess whether our setup enables the interpretation of such deictic gestures by remote participants, we need a task that required the production and interpretation of such gestures.

We use a version of Liu et al.'s [20] task, which consists of classifying discs into containers based on their label. In one condition of their experiment, one participant had to tell the other which disc to move into which container. They naturally used deictic instructions, such as *"take this one and put it here"*. We adapted this abstract data manipulation task to a remote setting: at one site, the *Instructor* sees the labels and gives instructions, while at the other site, the *Performer* manipulates the discs. This forces each dyad to produce and interpret deictic instructions.

We divided each wall-sized display into 32 (8 rows × 4 columns) virtual containers holding up to 6 discs each (Fig. 1).

Discs belong to one of 8 classes, represented by the letters *C, D, H, N, K, R, X, Z*. When more than two discs of the same class are in the same container, they are properly classified and turn green. Misclassified discs are red. On the *Instructor* side, the disc classes are displayed in a small font (2mm × 2.5mm), forcing the *Instructor* to move to read the labels.

Layout: when the task begins, the layout features 160 discs, five per container. 12 discs are misclassified, distributed randomly across containers. The goal is to classify all the red discs by picking, moving and dropping them into a correct container. We assign a ROLE to each participant: the *Instructor* sees the disc labels but cannot interact with them; the *Performer* sees green and red discs without labels but can manipulate them with a pointing device. The *Instructor* must therefore guide the *Performer* to classify the discs.

We created two types of LAYOUTs by varying the euclidean distance between a red disc and its closest solution[8]. This distance is between 1.5 and 2.6 for *Medium* layouts, vs. between 2.7 and 3.5 for *Hard* layouts. The further away a solution is from a disc, the more navigation is required, making the task harder. We generate random LAYOUTs for both *Medium* and *Hard* and pick one at random when starting a new session.

Trial description: a trial is the correct classification of a disc. A trial begins when the last disc of the previous trial is dropped, and ends when the disc is correctly classified (which may take several pick and drops).

Participants were welcomed and given paper instructions on how to perform the task. They were instructed to solve the task as quickly and accurately as possible. For each VIDEO condition, each participant performed 2 practice conditions (one for each ROLE), followed by 4 experimental conditions (2 LAYOUT × 2 ROLE). Participants filled out 5 questionnaires: one for collecting demographic data, one after each VIDEO condition, and one at the end of the experiment. Participants could take a break after each LAYOUT and were encouraged to do so at the end of a VIDEO condition block. Sessions lasted about 70 minutes including the time to fill out the questionnaires.

Data Collection

We logged each pick and drop event with the time, position of the disc on the screen and number of discs left to classify. Using each room's *VICON* tracking system, we recorded kinematic data of (a) user position, (b) user head direction and (c) cursor movement. Sessions were video recorded.

Pairs assessed their understanding of each other's actions and use of video in the questionnaire at the end of each VIDEO condition. The final questionnaire assessed the strategies and participant's preference when acting as *Instructor* and *Performer*. We used 5-point Likert scales and open-ended comments.

Analysis Procedure

We analyze three different measures: task performance, movement data from the kinematic logs, and conversations.

[8]The unit is the size of a container and the distance between two adjacent containers is 1.

Task Performance

We measure performance as Task Completion Time (*TCT*). The number of pick-and-drops for classifying one disc is a less useful indicator of performance than time, since all layouts were successfully solved with low error rates. *TCT* is the time required to correctly classify a disc. Since this may require several attempts, *TCT* starts when the *Performer* drops the previous disc and ends on the first drop in the correct container. We observed that some dyads picked one disc immediately and waited for an instruction, whereas others waited for an instruction and then picked a disc. To ensure a fair comparison and account for the time taken to find a container and produce the instruction in all trials, we include the time elapsed from the previous drop until a disc is picked.

Kinematic Analysis

To account for the slightly different sizes of the two wall-sized display, we normalize user position, cursor position and head direction between −1 and 1. After normalization, two users standing at the same relative position, e.g., the center of each room, have the same value, e.g., 0 on the X axis.

Conversational Analysis

Using the sessions' video recordings, we tagged each pick and drop and coded (I) the *Instructor strategy* to indicate containers/discs; (II) the *Performer error* when performing instructions; and, for both roles, (III) the *word count*, including the amount of deictic instructions.

I. Instructor strategy to indicate containers or discs used the following coding scheme:

- pointing: using the finger to point, no verbal instruction;
- pure deictic: using only deictic instructions (*"..goes there"*);
- relative to own position: relative to the *Instructor*'s position (*"here, one up"*);
- relative to video: relative to the *Instructor*'s video (*"where I am, second row"*);
- relative to disc: relative to where the *Performer* is moving the disc (*"there, one up"*);
- relative to container: relative to where the disc is picked (*"two to the right, one down"*);
- absolute: relative to the display grid (*"column 3, row 4"*);
- based on previous pick/drop: using the location where the previous disc was picked or dropped (*"put it in the same place as the last one"*).

II. Performer error when performing instructions used the following coding scheme:

- understanding error: error when interpreting an instruction;
- instruction error: the *Instructor* provides a wrong instruction (the container is not of the same class as the disc); and,
- interaction error: the *Performer* accidentally drops a disc while moving it.

III. Word count serves as a measure of the efficiency when producing and understanding instructions. We used a coding scheme based on Gergle et al. [12]. We only coded utterances relevant to instructions, i.e. references to a specific disc and position. We counted words related to acknowledgment of behavior only when discs were not already dropped and changed

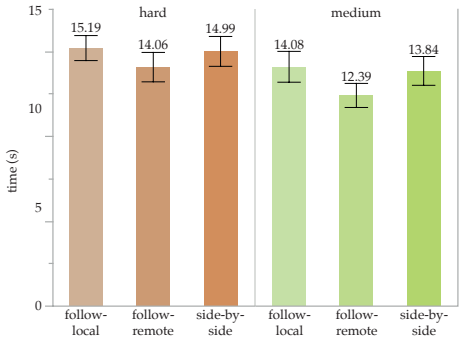

Figure 5. *TCT* **in seconds for each** VIDEO × LAYOUT **condition. Bars show the 95% confidence intervals.**

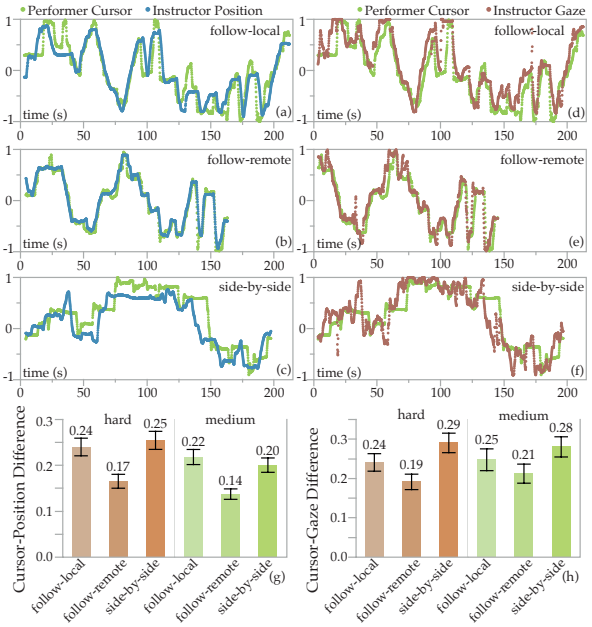

Figure 6. Kinematic data for dyad 9. The paths show a bird's eye view of the normalized horizontal positions of the participant, cursor and gaze over time. Left: *Performer* **cursor and** *Instructor* **position for** *Follow-Local* **(a),** *Follow-Remote* **(b) and** *Side-by-Side* **(c). Right:** *Performer* **cursor and** *Instructor* **gaze for** *Follow-Local* **(d),** *Follow-Remote* **(e) and** *Side-by-Side* **(f). The histograms show the cursor-position difference (g) and cursor-gaze difference (h) for each** VIDEO **condition, with 95% confidence intervals.**

their color to green; once this happened, words were considered redundant for the classification and ignored. We ignored context information not relevant to an instruction (such as discussing the task itself) and back channel responses such as *"hmmm"* or *"so"*. Hauber et al. [13] also used this approach for counting words. Politeness forms were not coded, e.g., *"could you please"*. Finally, repeated terms were counted once, since we identified that many participants repeated utterances, e.g., *"that one, yes, yes, yes"*).

RESULTS

We registered 4330 pick and drop events (excluding practice trials) and aggregated them into 1728 disc classifications (12 discs * 2 ROLE * 2 LAYOUT * 3 VIDEO * 12 pairs).

Task Performance

We tested *TCT* for normality in each VIDEO condition using a Shapiro-Wilk W test[9] and found that it was not normally distributed. We tested for goodness-of-fit with a lognormal distribution using Kolmogorov's D test, which showed a non-significant result for all three VIDEO conditions. Therefore, we ran the analyses using the log-transform of *TCT*, as recommended by Robertson & Kaptein [26] (p. 316). We also ran all the analyses using the original time data an found the same effects with very similar p values.

We ran an analysis of variance for the model *TCT* ~ VIDEO × LAYOUT × Rand(PARTICIPANT) with a REsidual Maximum Likelihood (REML) analysis [10]. The result of the full factorial analysis (Fig. 5) yields a significant effect on VIDEO ($F_{2,1699}$ = 7.69, p = 0.0005) and LAYOUT ($F_{1,1714}$ = 22.41, $p < 0.0001$); and a non-significant VIDEO × LAYOUT interaction ($F_{2,1713}$ = 0.50, p = 0.61).

A post-hoc analysis[11] reveals that in *Follow-Remote* (*MT* = 13.23±6.88 s), participants classified discs significantly faster than in *Follow-Local* (*MT* = 14.63±7.15 s, p=0.0004) and *Side-by-Side* (*MT* = 14.41±7.47 s, p=0.0173). There was no difference between *Follow-Local* and *Side-by-Side*. Data shows an improvement of *Follow-Remote* over *Side-by-Side* of 8.2% (1.19s); and over *Follow-Local* of 9.6% (1.41s).

Kinematic Analysis

In *Follow-Remote*, we observed that after picking a disc, the *Performer* would try to predict the target container by looking at the *Instructor*'s cursor and head direction. Some *Performer*s where even able to interpret target containers with minimal instructions, often following the *Instructor* and dropping the disc into the container that the *Instructor* was looking at.

We computed two measures to investigate this observation. *Cursor-position difference*: the horizontal distance between the *Performer*'s cursor and the *Instructor*'s position (Fig. 6a-c); and, *cursor-gaze difference*: the horizontal distance between the *Performer*'s cursor and the estimated point the Instructor is looking at (based on the direction of the head) (Fig. 6d-f). To get a single value per trial, we average these measures for all kinematic data points between a pick and a drop.

Cursor-Position Difference

We find a significant difference in *cursor-position difference* for VIDEO ($F_{2,1697}$ = 64.09, $p < 0.0001$) and for LAYOUT ($F_{1,1711}$ = 32.42, $p < 0.0001$), but not for VIDEO × LAYOUT ($F_{2,1711}$ = 0.023, p = 0.98) (Fig. 6g). A post-hoc analysis shows that *Follow-Remote* (X = 0.151±0.116) has a significantly smaller *cursor-position difference* than *Follow-Local* (X = 0.228±0.154) and *Side-by-Side* (X = 0.227±0.155).

[9] All analyses are performed with SAS JMP

[10] Unless otherwise specified, all analyses used this method.

[11] All post-hoc analysis are performed using a Tukey-Kramer "Honestly Significant Difference" (HSD) test

	Follow-Local	*Follow-Remote*	*Side-by-Side*
pure deictic	8	92	43
relative to video	3	318	0
relative to own position	0	0	7
relative to disc	116	61	148
relative to container	221	63	196
established pick order	284	302	257
based on previous drop	58	27	19
based on previous pick	10	5	12
absolute	447	254	435
arbitrary by *Performer*	80	83	111
none	248	241	247
total	1475	1446	1475
		4396	

Table 1. *Instructor* **strategies for indicating objects.**

Cursor-Gaze Difference

We also find a significant difference in *cursor-gaze difference* for VIDEO ($F_{2,1697} = 30.64$, $p < 0.0001$), but not for LAYOUT ($F_{1,1704} = 2.28$, $p = 0.13$) nor VIDEO × LAYOUT ($F_{2,1704} = 0.9021$, $p = 0.41$) (Fig. 6h). A post-hoc analysis shows that all VIDEO conditions are significantly different from each other. *Follow-Remote* has the smallest *cursor-gaze difference* ($X = 0.201 \pm 0.190$), followed by *Follow-Local* ($X = 0.244 \pm 0.216$) and *Side-by-Side* ($X = 0.284 \pm 0.217$).

Conversational Analysis

We analyze the strategies used by the *Instructor* and the errors produced by the *Performer* when picking and dropping discs. For this analysis, we use the tagged data for each pick and drop, not the aggregated data for correctly classified discs. While tagging video data two new categories emerged: *arbitrary* (no instruction), when the *Performer* picks any disc; *established pick order*, when the *Performer* picks in an order defined at the beginning of the session.

4396 events were tagged (Table 1), evenly distributed among the three VIDEO conditions. Note that events that have no strategy come from the *Performer* correcting errors (most often due to a failed interaction, such as releasing the disc too soon while moving it), which required no instruction: she would re-pick the disc and drop it in the planned destination.

Instructor Strategies

We investigate the role of video on the use of deictic instructions by the *Instructor* (*Instructor* strategies or *IS*). We consider as deictic instructions the following categories: pure deictic, relative to video and relative to own position. For the last two, we always observed that the *Instructor* used a deictic pronoun to make a reference relative to the position of the video or to herself. We counted 410 deictic instructions for *Follow-Remote* (318 relative to video; 92 pure deictic), 50 for *Side-by-Side* (7 relative to own position; 43 pure deictic) and 11 for *Follow-Local* (3 relative to video, 8 pure deictic). No deictic relative to own were used.

As expected, participants were able to use more deictic instructions in *Follow-Remote* (28% of total) than *Side-by-Side* (3.4%) and *Follow-Local* (only 0.7%). If we take a closer look at the strategies for disc drop events only, the use of deictic instructions in *Follow-Remote* goes up to 45% (265 relative to video, 65 pure deictic; 729 total). These findings support *H1*.

We were surprised to see some participants using deictic instructions in *Side-by-Side*. We believe that they tried, failed, and switched to less unambiguous strategies such as using coordinates relative to the container where the disc was picked. We were also surprised that almost all participants pointed with their hands in all VIDEO conditions, even though they clearly knew that pointing would not be correctly understood in *Follow-Local* and *Side-by-Side*.

Performer Errors

We investigate the role of video on *Performer*s producing errors (*PE*) when interpreting instructions, especially deictic instructions. We remove instruction and interaction errors from the analysis, leaving 246 misunderstanding errors. Overall, participants produced fewer errors in *Follow-Remote* (66, 27% of total), followed by *Follow-Local* (82, 33% of total) and *Side-by-Side* (98, 40% of total).

Follow-Remote accounted for fewer errors if we consider the total number of deictic instructions produced in each VIDEO condition. 36% (4/11) of deictic instructions led to an error in *Follow-Local* and 40% (20/50) in *Side-by-Side*, but only 5% (21/410) in *Follow-Remote*. Deictic instructions were better interpreted in *Follow-Remote* than in the other VIDEO conditions, supporting *H2*.

Word Count

We measure word count (*WC*) as a measure of communication efficiency. Using fewer words to communicate the same information suggests that the communication is more efficient, because the information is transmitted through other non-linguistic channels—video in our case. Participants used significantly different number of words in each VIDEO condition ($F_{2,3739} = 50.0747$, $p < 0.0001$). As expected, in *Follow-Remote*, *Instructor*s used significantly fewer words ($WC = 2.98 \pm 2.66$ words) per instruction than in *Follow-Local* ($WC = 3.80 \pm 3.42$ words) and *Side-by-Side* ($WC = 4.07 \pm 3.52$ words).

We tagged the number of deictic pronouns used by *Instructor*s. In *Follow-Remote*, 272 deictic pronouns were used, 110 in *Side-by-Side* and only 70 in *Follow-Local*.

In summary, when providing instructions in *Follow-Remote*, *Instructor*s used fewer words but more deictic pronouns than in other VIDEO conditions. To illustrate this point, *Instructor*s in *Follow-Local* typically used more verbose instructions, e.g., *"two to the left, then top"*, whereas in *Follow-Remote* they used short instructions with a deictic pronoun, e.g., *"top"* once they were in the correct column or simply *"there"* while pointing.

Qualitative Feedback

We asked participants to answer a short questionnaire at the end of each VIDEO condition, and a final questionnaire at the end of the experiment. Questionnaires had both Likert scales and open questions.[12]

The questionnaire for the different VIDEO conditions had two identical parts, one for each ROLE. Most questions were about

[12] We used a Wilcoxon Signed rank test for Likert scale data analysis

perceived attention to each other: (Q1) I paid attention to my partner; (Q2) My partner paid attention to me; (Q3) It was easy to understand my partner; (Q4) My partner found it easy to understand me; (Q5) My behavior was in direct response to my partner's behavior and (Q6) The behavior of my partner was in direct response to my behavior. (Q7) asked to estimate how much time they spent looking at the video when classifying objects (on a scale from 1 to 100), and (Q8) asked to assess how useful was the video of their partner for solving the task.

We found a significant effect of VIDEO on how useful the video was when acting both as *Performer* ($F_{2,22} = 26.96$, $p < 0.0001$) and *Instructor* ($F_{2,22} = 11.34$, $p = 0.0004$). For *Performer*s, the video of the remote partner was significantly more useful in *Follow-Remote* (mean 4.42) than in *Follow-Local* ($p < 0.0001$, mean 2.67) and *Side-by-Side* ($p < 0.0001$, mean 2.25). For *Instructor*s, the video was also more useful in *Follow-Remote* (mean 2.83) than in *Follow-Local* ($p = 0.00003$, mean 1.92) and *Side-by-Side* ($p = 0.0011$, mean 1.50). Also, *Instructor*s had the impression that their partner paid significantly more attention to them ($F_{2,22} = 7.50$, $p = 0.0033$) in *Follow-Remote* (mean 4.58) than in *Follow-Local* ($p = 0.0051$, mean 3.92) and *Side-by-Side* ($p = 0.013$, mean 3.83).

We also found an effect of VIDEO on how much participants looked at video both as *Performer* ($F_{2,22} = 13.24$, $p = 0.0002$) and *Instructor* ($F_{2,22} = 11.44$, $p = 0.0004$). *Performer*s used the video significantly more in *Follow-Remote* ($\%\ of\ time = 87 \pm 17$) than in *Follow-Local* ($p = 0.0047$, $\%\ of\ time = 53 \pm 35$) and *Side-by-Side* ($p = 0.0002$, $\%\ of\ time = 40 \pm 33$). *Instructor*s used the video significantly more in *Follow-Remote* ($\%\ of\ time = 55 \pm 36$) than in *Follow-Local* ($p = 0.023$, $\%\ of\ time = 30 \pm 25$) and *Side-by-Side* ($p = 0.0003$, $\%\ of\ time = 15 \pm 22$).

The final questionnaire asked participants (Q1) if they understood their partner's instructions when acting as *Performer* and (Q2) if their partner understood their instructions when acting as *Instructor*. It also asked (Q3-4-5) how often they used each Instructor Strategy (*IS*) in each VIDEO condition, (Q6-7) their preferred VIDEO condition as *Instructor* and as *Performer*, and (Q8-9-10) a description of how they used the video in each VIDEO condition.

Participants reported that as *Performer*s in *Follow-Remote*, they understood their partner's instructions significantly better ($p = 0.0188$), and that the most used strategy to indicate objects was to use the video. We found no other significant effects. The vast majority of *Performer*s preferred *Follow-Remote* (22/24), while *Side-by-Side* was ranked first twice. *Follow-Local* and *Side-by-Side* were ranked as the second preferred strategy by roughly half the participants (12 and 10 respectively) and as least preferred by the other half (12 each).

Video in *Follow-Remote* was used by *Performer*s to *"see where [the Instructor] was and then get the column where the object should be"* (P5) and *"to know on which column I have to place my object"* (P6). It also allowed them to *"follow [the Instructor's] position around the wall"* (P7). Many *Performer*s used the video *"to know what column [the Instructor] wants to pick and sometimes even the row"* (P9). We also observed that they used the video to determine gaze and predict the destination container more quickly: *"to get [the Instructor's] position, even the gaze helped me"* (P21). Some *Performer*s cleverly used the video in *Follow-Local* to estimate their partner's position. As people move, *CamRay* switches from one camera to the next in the array to capture their faces. These *Performer*s counted the "jumps" of the video window to *"roughly figure out how much I should move to the left/right"* (P23).

Surprising, only half the *Instructor*s ranked *Follow-Remote* first. *Follow-Local* was ranked first 10 times, and *Side-by-Side* 2 times. This was confirmed by participants when asked to describe how they used the video in *Follow-Local* and *Side-by-Side*: *"to see if [the Performer] was moving the object or not"* (P5), *"to know if my partner was focusing on the same task"* (P6), *"to get gaze direction and gestures, not position"* (P7) and *"to confirm verbal instructions"* (P20).

These findings partially support *H3*: almost all *Performer*s preferred *Follow-Remote*, but half of the *Instructor*s preferred having the video in front of them or on the side. *Instructor*s liked seeing their remote collaborator as they performed instructions to check for understanding.

To summarize our results, in *Follow-Remote* participants:

- used more deictic instructions than in other VIDEO conditions, supporting *H1*;
- classified discs more efficiently, used fewer words and produced fewer misunderstanding errors, supporting *H2*; and,
- preferred this condition as *Performer*, but half did not prefer it as *Instructor*s, partially supporting *H3*.

DISCUSSION

The above results provide evidence that the increased performance of *Follow-Remote* is related to (1) *Performer*s more closely following the *Instructor*s' position and gaze (*cursor-person difference*, *gaze-position difference*); and, (2) *Instructor*s using more deictic instructions (*IS*), leading to fewer errors (*PE*); and, (3) *Instructor*s using fewer words (*WC*).

First, *Performer*s were able to predict the target container as *Instructor*s moved and looked at the display: once an *Instructor* found a target container, the *Performer* would already be hovering a disc nearby and gazing in the vicinity, requiring less time to move and drop the disc. Second, as *Performer*s made fewer errors they saved time. Third, awareness of the remote person's actions allowed for short and simple instructions, such as *"just there!"* or *"one above!"*.

These findings can be explained by the natural tendency to minimize communication costs when generating common ground. Let us consider Clark's costs of grounding [8] in mediated communication for our experiment. Certain costs do not exist: there is no *start-up* time, and no *delay* nor *asynchrony* since communication was synchronous and real-time. Other costs are the same across VIDEO conditions: *production*, *reception* and *speaker change*, since all conditions used video-mediated communication; *fault* and *repair* since the severity of a fault and the time and effort to repair it dependended mainly on the task. We are thus left with three costs: *formulation*, *understanding* and *display*.

Formulation cost states that *"it costs more to plan complicated than simple utterances"* and *"to formulate perfect than imperfect utterances"* [8]. Different strategies have different costs: an instruction that relies on a coordinate system for absolute mapping, e.g., *"on container 3, 2"*, or a relative mapping to a container, e.g., *"two up, one down"* are costlier than pointing and using a pure deictic pronoun, e.g., *"there!"*. This suggests that *Follow-Remote* had a lower formulation cost.

Understanding cost states that *"the costs can be compounded when contextual clues are missing"* [8]. This explains why, when using deictic pronouns in *Follow-Local* or *Side-by-Side*, participants produced more errors: the cost of understanding is higher since the context to interpret instructions is missing.

Finally, *display* cost states that *"In media without copresence, gestures cost a lot, are severely limited, or are out of the question. In video teleconferencing, we can use only a limited range of gestures."* [8]. This explains why in *Follow-Remote*, *Instructor*s were able to use more deicitc gestures and these were understood more accurately by *Performer*s, reducing the display cost. This also explains why *Performer*s preferred *Follow-Remote* when interpreting instructions, while half the *Instructor*s preferred *Follow-Local* or *Side-by-Side*, since they could more easily check the *Performer*s for understanding.

In summary, by presenting video according to the remote collaborator's location in *Follow-Remote*, we enabled participants to use and understand deictic instructions, reducing the overall cost of communication.

Implications for Design

The above analysis leads to a set of recommendations for the design of telepresence systems for wall-sized displays.

Camera Arrays support remote collaboration in large interactive spaces that allow physical navigation. An array of cameras placed at eye's level can capture people's faces as they move across a wall-sized display. Remote displays can present this video feed in various ways to enable collaboration.

Follow-Remote supports deictic instructions when collaborating remotely across wall-sized displays. By displaying the remote participant's video in the context of the shared space, it creates an instance of Buxton's Reference Space, *"the space within which the remote party can use body language to reference the work—things like pointing, gesturing [and] the channel through which one can sense proximity, approach, departure, and anticipate intent"* [6]. Collaborative data manipulation tasks can particularly benefit from this setup, as they often require deictic instructions.

Users should control video position in order to better support different tasks: when interpreting deictic instructions, *Follow-Remote* provides an image of the remote person in the context of the shared space; when checking for understanding, creating a virtual face-to-face with *Follow-Local* makes the remote person's gaze and facial expressions directly available.

CONCLUSION AND FUTURE WORK

This paper introduces *CamRay*, a platform for remote collaboration that captures and presents video feeds of remote participants while working in front of wall-sized displays. *CamRay* is based on consumer hardware and open software; it can be incorporated into existing wall-sized displays to add telepresence capabilities.

We ran an experiment where we used *CamRay* to support collaboration on an asymmetric data manipulation task. The video feed either followed the local person's position (*Follow-Local*), the remote person's position (*Follow-Remote*) or was on the side (*Side-by-Side*). We investigated how the position of the video feed affects collaboration. Participants were able to manipulate data more efficiently, taking less time, making fewer errors and using fewer words when video followed the remote collaborator. This can be explained by the fact that video enabled them to use and better understand deictic instructions, reducing the cost of communication. However, many participants liked having video always visible, either in front of them or on the side, when checking their partner's understanding of instructions.

We found that both *Follow-Local* and *Follow-Remote* have their own advantages. With *Follow-Remote*, people are positioned in the context of shared content, allowing them to communicate using deictic gestures. With *Follow-Local*, nonverbal cues, such as facial expressions and eye contact, are made visible, supporting face-to-face communication. We believe that both approaches can be used in a telepresence system to support different moments in the collaboration. We recommend that collaboration systems for wall-sized displays present video feeds according to the local and remote users' position, and provide a way to transition between them.

This is only a first step for telepresence in large interactive spaces. We believe that *CamRay* can be used to further explore the role of video in remote collaboration across wall-sized displays. We plan to explore how *Follow-Local* can support tasks that require discussion or benefit from seeing each others' faces, such as data visualization or sense making. We are also interested in exploring the benefits of collaboration using asymmetric video positions. We observed that people preferred different video behaviors depending on their role in the task, and we believe there are further benefits in positioning the video feeds independently from each other.

Finally, we are interested in exploring how camera arrays can support collaboration with more than two users and two sites. From a technical perspective, we need to solve the challenge of selecting and displaying multiple video and audio feeds as multiple collaborators are present in multiple sites. From the perspective of collaboration, we need to support the variety of collaboration styles that occur spontaneously in larger groups.

ACKNOWLEDGMENTS

This work was partially supported by European Research Council (ERC) grants n° 321135 CREATIV: Creating Co-Adaptive Human-Computer Partnerships and n° 695464 ONE: Unified Principles of Interaction; and by EquipEx DIGISCOPE (ANR-10-EQPX-26-01), operated by the French Agence Nationale de la Recherche (ANR) as part of the program "Investissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02).

REFERENCES

1. Ignacio Avellino, Cédric Fleury, and Michel Beaudouin-Lafon. 2015. Accuracy of Deictic Gestures to Support Telepresence on Wall-sized Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2393–2396. DOI: http://dx.doi.org/10.1145/2702123.2702448
2. Robert Ball, Chris North, and Doug A. Bowman. 2007. Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 191–200. DOI: http://dx.doi.org/10.1145/1240624.1240656
3. S. Beck, A. Kunert, A. Kulik, and B. Froehlich. 2013. Immersive Group-to-Group Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 616–625. DOI: http://dx.doi.org/10.1109/TVCG.2013.33
4. Sara A. Bly, Steve R. Harrison, and Susan Irwin. 1993. Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. *Commun. ACM* 36, 1 (Jan. 1993), 28–46. DOI: http://dx.doi.org/10.1145/151233.151235
5. Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. 2002. Effects of Four Computer-mediated Communications Channels on Trust Development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 135–140. DOI: http://dx.doi.org/10.1145/503376.503401
6. Bill Buxton. 2009. Mediaspace – Meaningspace – Meetingspace. In *Media Space 20 + Years of Mediated Life*, Steve Harrison (Ed.). Springer, London, 217–231. DOI: http://dx.doi.org/10.1007/978-1-84882-483-6_13
7. Milton Chen. 2002. Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 49–56. DOI: http://dx.doi.org/10.1145/503376.503386
8. Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on socially shared cognition*, L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.). American Psychological Association, Washington, DC, US, 127–149.
9. Herbert H Clark and Catherine R Marshall. 1981. *Definite reference and mutual knowledge*. Cambridge University Press.
10. Herbert H. Clark, Robert Schreuder, and Samuel Buttrick. 1983. Common ground at the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior* 22, 2 (April 1983), 245–258. DOI: http://dx.doi.org/10.1016/S0022-5371(83)90189-5
11. M. Dou, Y. Shi, J. M. Frahm, H. Fuchs, B. Mauchly, and M. Marathe. 2012. Room-sized informal telepresence system. In *2012 IEEE Virtual Reality Workshops (VRW)*. 15–18. DOI: http://dx.doi.org/10.1109/VR.2012.6180869
12. Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2004. Action As Language in a Shared Visual Space. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 487–496. DOI: http://dx.doi.org/10.1145/1031607.1031687
13. Jörg Hauber, Holger Regenbrecht, Mark Billinghurst, and Andy Cockburn. 2006. Spatiality in Videoconferencing: Trade-offs Between Efficiency and Social Presence. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. ACM, New York, NY, USA, 413–422. DOI: http://dx.doi.org/10.1145/1180875.1180937
14. Keita Higuchi, Yinpeng Chen, Philip A. Chou, Zhengyou Zhang, and Zicheng Liu. 2015. ImmerseBoard: Immersive Telepresence Experience Using a Digital Whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2383–2392. DOI: http://dx.doi.org/10.1145/2702123.2702160
15. Ellen A. Isaacs and John C. Tang. 1993. What Video Can and Can'T Do for Collaboration: A Case Study. In *Proceedings of the First ACM International Conference on Multimedia (MULTIMEDIA '93)*. ACM, New York, NY, USA, 199–206. DOI: http://dx.doi.org/10.1145/166266.166289
16. Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 525–532. DOI: http://dx.doi.org/10.1145/142750.142977
17. Clemens N. Klokmose, James R. Eagan, Siemen Baader, Wendy Mackay, and Michel Beaudouin-Lafon. 2015. Webstrates: Shareable Dynamic Media. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 280–290. DOI: http://dx.doi.org/10.1145/2807442.2807446
18. Lars Lischke, Sven Mayer, Katrin Wolf, Niels Henze, Albrecht Schmidt, Svenja Leifert, and Harald Reiterer. 2015. Using Space: Effect of Display Size on Users' Search Performance. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1845–1850. DOI: http://dx.doi.org/10.1145/2702613.2732845
19. Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. 2016. Shared Interaction on a Wall-Sized Display in a Data Manipulation Task. In *Proceedings of the 2016 CHI Conference on Human Factors in*

Computing Systems (CHI '16). ACM, New York, NY, USA, 2075–2086. DOI: http://dx.doi.org/10.1145/2858036.2858039

20. Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E. Mackay. 2014. Effects of Display Size and Navigation Type on a Classification Task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4147–4156. DOI: http://dx.doi.org/10.1145/2556288.2557020
21. Paul K. Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. 2015. Flexible Ecologies And Incongruent Locations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 877–886. DOI: http://dx.doi.org/10.1145/2702123.2702286
22. Wendy E. Mackay. 1999. Media Spaces: Environments for multimedia interaction. In *Computer-Supported Cooperative Work*, Michel Beaudouin-Lafon (Ed.). Wiley and Sons, Chichester, 55–82.
23. Andrew F. Monk and Caroline Gale. 2002. A Look Is Worth a Thousand Words: Full Gaze Awareness in Video-Mediated Conversation. *Discourse Processes* 33, 3 (2002), 257–278. DOI: http://dx.doi.org/10.1207/S15326950DP3303_4
24. David T. Nguyen and John Canny. 2007. Multiview: Improving Trust in Group Video Conferencing Through Spatial Faithfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1465–1474. DOI: http://dx.doi.org/10.1145/1240624.1240846
25. David T. Nguyen and John Canny. 2009. More Than Face-to-face: Empathy Effects of Video Framing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 423–432. DOI: http://dx.doi.org/10.1145/1518701.1518770
26. Judy Robertson and Maurits Kaptein. 2016. *Modern Statistical Methods for HCI*. Springer. DOI: http://dx.doi.org/10.1007/978-3-319-26633-6
27. Abigail Sellen, Bill Buxton, and John Arnott. 1992. Using Spatial Cues to Improve Videoconferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 651–652. DOI: http://dx.doi.org/10.1145/142750.143070
28. Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C. Tang, and Bill Buxton. 2010. Three's Company: Understanding Communication Channels in Three-way Distributed Collaboration. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 271–280. DOI: http://dx.doi.org/10.1145/1718918.1718969
29. John C. Tang and Scott Minneman. 1991. VideoWhiteboard: Video Shadows to Support Remote Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. ACM, New York, NY, USA, 315–322. DOI: http://dx.doi.org/10.1145/108844.108932
30. John C. Tang and Scott L. Minneman. 1990. VideoDraw: A Video Interface for Collaborative Drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 313–320. DOI: http://dx.doi.org/10.1145/97243.97302
31. Elizabeth S. Veinott, Judith Olson, Gary M. Olson, and Xiaolan Fu. 1999. Video Helps Remote Work: Speakers Who Need to Negotiate Common Ground Benefit from Seeing Each Other. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 302–309. DOI: http://dx.doi.org/10.1145/302979.303067
32. Malte Willert, Stephan Ohl, Anke Lehmann, and Oliver Staadt. 2010. The Extended Window Metaphor for Large High-resolution Displays. In *Proceedings of the 16th Eurographics Conference on Virtual Environments & Second Joint Virtual Reality (EGVE - JVRC'10)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 69–76. DOI: http://dx.doi.org/10.2312/EGVE/JVRC10/069-076

GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration

Khanh-Duy Le[1], Ignacio Avellino[2], Cédric Fleury[3], Morten Fjeld[1], and Andreas Kunz[4]

[1] Chalmers University of Technology, Sweden
{khanh-duy.le,fjeld}@chalmers.se
[2] ISIR, CNRS, Sorbonne Université, France
ignacio.avellino@sorbonne-universite.fr
[3] LRI, Univ. Paris-Sud, CNRS, Inria, Université Paris-Saclay, France
cedric.fleury@lri.fr
[4] ETH Zurich, Switzerland
kunz@iwf.mavt.ethz.ch

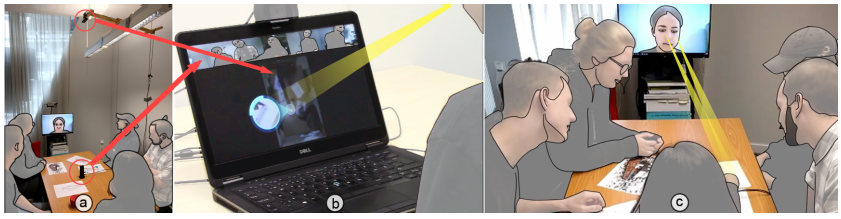

Fig. 1: *GazeLens* system. (a) On the hub side, a 360° camera on the table captures coworkers and a webcam mounted on the ceiling captures artifacts on the table. (b) Video feeds from the two cameras are displayed on the screen of the remote satellite worker; a virtual lens strategically guides her/his attention towards a specific screen area according to the observed artifact. (c) The satellite's gaze, guided by the virtual lens, is aligned towards the observed artifact on the hub space.

Abstract. In hub-satellite collaboration using video, interpreting gaze direction is critical for communication between hub coworkers sitting around a table and their remote satellite colleague. However, 2D video distorts images and makes this interpretation inaccurate. We present *GazeLens*, a video conferencing system that improves hub coworkers' ability to interpret the satellite worker's gaze. A 360° camera captures the hub coworkers and a ceiling camera captures artifacts on the hub table. The system combines these two video feeds in an interface. Lens widgets strategically guide the satellite worker's attention toward specific areas of her/his screen allow hub coworkers to clearly interpret her/his gaze direction. Our evaluation shows that *GazeLens* (1) increases hub coworkers' overall gaze interpretation accuracy by 25.8% in comparison to a conventional video conferencing system, (2) especially for physical artifacts on the hub table, and (3) improves hub coworkers' ability to distinguish between gazes toward people and artifacts. We discuss how screen space can be leveraged to improve gaze interpretation.

Keywords: remote collaboration · telepresence · gaze · lens widgets.

1 Introduction

In hub-satellite communication, a remote team member (satellite) collaborates at a distance with colleagues at the main office (hub). Typically, hub coworkers sit around a table with artifacts such as paper printouts, with a screen placed at one edge of the table showing a video feed of the satellite worker. The satellite worker sees the hub office in a deep perspective as it is captured by a camera placed at the edge of the table. Hub coworkers see a closer view of their colleagues, with a much more shallow perspective. Simply put, due to these differences in perspective, it is difficult to interpret the satellite's gaze (where s/he's looking at). While video conferencing systems can support non-verbal cues as people can see each others' faces and gestures, it is not always coherent: non-verbal cues such as gaze and deictic gestures are disparate between hub coworkers and the satellite, making communication asymmetric as co-located hub coworkers easily understand each others' non-verbal cues but not those of the satellite worker.

Gaze is important in collaboration - it is a reliable predictor of conversational attention [1, 4], offering effortless reference to spatial objects [29], supporting remote instruction [5, 30], and improving users' confidence in distributed problem solving on shared artifacts [2]. Kendon [27] argues that gaze is a signal through which a person relates their basic orientation and even intention toward another. Falling short on conveying gaze in remote collaboration can lead to confused communication [3], reduce social intimacy [3], decrease effectiveness [32] and increase effort for collaborative tasks [2, 5].

Previous work has tried to improve gaze perception in remote collaboration, but has mainly focused on conveying either gaze awareness between distant coworkers [7, 8, 25] or gaze on shared digital content [11, 14], leaving the problem of conveying gaze on physical artifacts rather under attended. Achieving this often requires specialized and complex hardware setups on the satellite side [9, 13, 16], which might be unrealistic for traveling workers. We focus on designing a mobile solution to improve hub coworkers' interpretation of the satellite worker's gaze both toward themselves and hub physical artifacts using minimal equipment.

We present *GazeLens*, a hub-satellite video conferencing system that improves hub coworker's accuracy when interpreting the direction of a satellite worker's gaze. At the hub side, *GazeLens* captures two videos: a view of the coworkers, using an off-the-shelf 360° camera, and a view of the artifacts on the table, using a ceiling-mounted camera. The system presents these videos simultaneously on the satellite worker's laptop screen, eliminating the need for stationary or specialized hardware on the remote end. *GazeLens* displays lenses on the satellite's screen, which the satellite worker can move to focus on different parts of the two videos, such as a hub coworker on the 360° view and an artifact on the table view. These lenses are strategically positioned to explicitly guide the direction of the satellite worker's gaze. As with conventional video conferencing, hub coworkers simply see a video stream of their remote colleague's face shown on the screen placed on the edge of their table. Our aim is to provide a more

coherent picture for hub coworkers of where exactly their satellite colleague is directing his/her attention, thus improving clarity of communication.

We evaluate the performance of *GazeLens* in two studies, where we compare it to conventional video conferencing (*ConvVC*) using a wide-angle camera on the hub side. The first study shows that *GazeLens* helps hub coworkers distinguish whether a satellite worker is looking at a person or at an artifact on the hub side. The second study shows that *GazeLens* helps hub coworkers interpret which artifact on the hub table the satellite worker is looking at. Early feedback on usability show the benefits for satellite workers, by improving visibility of hub artifacts and hub coworkers' activities while maintaining their spatial relations. We show that screen space can be better leveraged through strategic placement of interface elements to support non-verbal communication in video conferencing and thus convey a satellite worker's gaze direction.

2 Related Work

2.1 Gaze Awareness Among Remote coworkers

One-to-one Remote Collaboration: Gemmel et al. [24] and Giger et al. [25] proposed using computer vision to manipulate eye gaze in the remote worker's video. They focused on achieving direct eye contact by correcting the disparity between the location of the video conferencing window and the camera.
Multi-party Remote Collaboration: In the Hydra system [7], each remote party was represented by a hardware device containing a display, a camera, and a microphone. These were spatially arranged in front of the local worker, helping convey the worker's gaze.
Group-to-group Remote Collaboration: For each participant, MultiView [8] used one camera and one projector to capture and display each person on one side from the perspective of each person on the other. Similarly, MMSpace [13] placed multiple displays around the table of a local group, each representing a worker on the remote side, replicating the sitting positions of the remote workers. Both systems maintained correct gaze awareness between remote coworkers.
Hub-satellite Remote Collaboration: Jones et al. [15] installed a large screen on the satellite's side to display the hub's video stream and employed multiple cameras to construct a 3D model of the satellite worker's face to help hub coworkers perceive their gaze. Pan and Steed [9] and Gotsch et al. [16] also used an array of cameras to capture the satellite worker's face from different angles, selectively displaying the images to hub coworkers on a cylindrical display.

While most previous work focused on direct eye contact and gaze awareness between remote coworkers, only a few have attempted to provide correct interpretation of gaze toward shared artifacts - either virtual artifacts shared through a synchronized system or physical artifacts at either location - mandatory in hub-satellite collaboration that involves shared objects on the hub table. In addition, the above systems often require specialized hardware, which is not suitable for a traveling satellite worker who needs a lightweight and mobile device.

2.2 Gaze Support for Shared Virtual Artifacts

ClearBoard [11] creates a write-on-glass metaphor by overlaying a shared digital canvas on the remote coworker's video feed, inherently conveying gaze between two remote people working on the canvas. Similarly, Holoport [14] captures images of the hub's workers using a camera behind a scree, which helps convey coworkers' gaze among each other as well as towards on-screen artifacts. GAZE and GAZE-2 [1, 18] introduce a 3D virtual environment where the video stream of each worker is displayed on a 3D cube that could change direction to convey the worker's gaze toward others. Hauber et al. [12] evaluated a setup where workers were equipped with a tabletop showing a shared display, coupled with a screen showing the video feeds of the remote workers. A camera was mounted on top of the screen to capture remote workers' faces. They also compared this technique with the a 3D virtual environment of GAZE [1]. Finally, Avellino et al. [31] showed that video can be used to convey gaze and deictic gestures toward shared digital content in large wall-sized displays.

All these systems convey gaze direction to shared virtual artifacts by keeping the spatial relation of the video feed to the digital content. While they demonstrate that people can interpret gaze direction from a video feed, these techniques are not applicable in the context of hub-satellite collaborations involving physical artifacts on the hub table. These systems are designed for symmetric and specific settings such as large interactive whiteboards or wall-sized displays, which are not appropriate for mobile workers or small organisations.

2.3 Gaze Support for Physical Artifacts

Visualizations that indicate remote gaze direction have been explored for supporting physical collaborative tasks [2, 5, 29]. Otsuki et al. [17] developed ThirdEye, an add-on display that conveys the remote worker's gaze into the 3D physical space. It projects a 2D graphic element, controlled by eye tracking data of the remote worker, onto a hemispherical surface that looks like an eye. However, such mediated representations might introduce spatial disparities when compared to unmediated gaze, potentially leading to confusion and reducing the value of the satellite's video feed. These solutions add complexity to the satellite worker's setup, by adding specialized hardware such as an eye tracker.

Xu et al. [6] introduce an approach for conveying the satellite worker's attention in hub-satellite collaboration. The satellite worker can view a panoramic video stream of the hub on their screen, captured by a 360° camera, and manually select the area of interest in the video. A tablet on the hub's side, horizontally placed under the 360° camera, showed an arrow pointing at the area selected by the satellite worker. This solution cannot convey the satellite's attention toward physical artifacts as it lacks the vertical dimension of their gaze, and using an arrow to represent gaze might also be distracting and unnatural for the hub coworkers as compared to an unmediated gaze.

Finally, CamBlend [19] used video effects to blur the 180° video of the remote side, encouraging the user to focus on an area of interest in order to view it in

high resolution. This mimics a human's visual system, where foveal (central) vision has much higher acuity than peripheral vision [20]. CamBlend did not however aim to convey the satellite's gaze. *GazeLens* leverages this technique to provide the satellite worker with an overview of the hub's space, while guiding the satellite worker's attention to strategic locations in order to explicitly convey their gaze to the hub workers.

3 GazeLens Design

GazeLens is designed to improve the hub coworkers' perception of the satellite worker's gaze. It is motivated by the limitations of current video conferencing systems in conveying gaze.

3.1 Gaze Perception in Video Conferencing

Stokes [22] and Chen [10] showed that when the angle between the gaze direction and the camera is less than 5° in video conferencing between two people, the remote person perceives direct eye contact. Moreover, when one person looks towards the right of the camera, the remote person feels they are looking at their right shoulder, and so on. While this effect can be leveraged to establish eye contact between pairs of video conferencing endpoints [10], it may also be used for gaze interpretation in groups, such as hub-satellite settings.

Fig. 2: Hub table captured by a camera placed (a) below and (b) above the hub screen (image courtesy requested).

3.2 Limitations of Hub-Satellite Communication Systems

The screen on the hub side showing the satellite's video often uses a wide-angle camera that captures an overview of the hub environment, so the satellite worker can view the hub (Figure 2). Two typical placements for this camera are just above the screen, such as in the Cisco MX Series [34] and Polycom RealPresence Group Series [35] or below the screen, as in the Cisco SX80 [36] or AVS solutions [37]. Neither of these setups effectively conveys the satellite worker's gaze back to hub coworkers nor at the artifacts on the table. When the camera is placed below the screen, artifacts on the table are largely occluded or difficult to see, but the satellite sees hub worker's faces straight on (Figure 2a).

With a placement above the screen, the hub's artifacts are less occluded, but the hub's environment as a whole seems distant, with a distortion of deep perspective where coworkers appear small (Figure 2b, note the distant table edge). This "mapping" of the hub's environment onto the satellite's computer screen leads to hub coworkers being unable to distinguish the satellite's gaze toward different people and artifacts. Additionally, hub coworkers near the camera appear lower on the satellite's computer screen, making it harder for them to discern whether the satellite worker is looking at a coworker or at an object on the table.

3.3 Design Requirements

With these limitations in mind, we derived the following design requirements for a video conferencing system that can convey the satellite worker's gaze toward their hub coworkers and physical artifacts:

DR1: the system needs to display a view of the hub to the satellite worker in which they can see both the hub coworkers' faces and the artifacts on the table without occlusions.

DR2: the system should allow hub coworkers to clearly distinguish if the satellite worker is gazing toward individual coworkers or hub table artifacts.

DR3: the system should allow hub coworkers to accurately interpret the satellite worker's gaze toward physical artifacts.

DR4: the system should only rely on video to convey the satellite worker's gaze, and avoid mediated gaze representations such as arrows, pointers or virtual arms, which introduce spatial and representational disparities.

DR5: the system for the satellite worker should consist of a lightweight and mobile device which does not require any calibration, suitable for traveling.

3.4 *GazeLens* Implementation

Hub side: to ensure *DR1*, *GazeLens* captures the panoramic video of the coworkers sitting around the hub's table using a 360° camera placed at the center of it, and it captures the scene using a camera mounted on the ceiling to avoid occluding artifacts on the hub table (see Figure 1a).

Satellite side: GazeLens presents the two video feeds to the satellite worker on a standard laptop with a camera, satisfying *DR5* (see Figure 1b). Their presentation is designed so that it improves the interpretation of the satellite's gaze. To fulfill *DR2*, the video feed displaying hub coworkers should be placed near the satellite's laptop camera, located above the screen. This panoramic video is then segmented on the satellite's display to maintain spatial fidelity: the hub coworkers sitting in front of their screen are shown in the center of the satellite's video, while those on the sides of the hub table are displayed on their corresponding sides. The overview video of the hub table is displayed below the panoramic video of the hub coworkers (Figure 3).

To address *DR3*, the video of the hub table view is scaled to fit the satellite's screen and to maximize the size of any objects on it, although this leads to different gaze patterns depending on table shape. Stretching this video to maintain

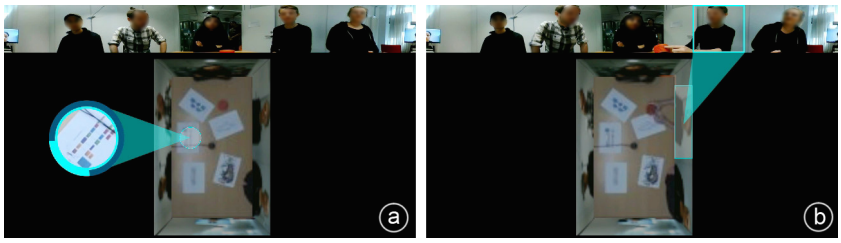

Fig. 3: *GazeLens* interface with (a) a lens showing a close-up of an artifact on the table and (b) a lens highlighting a hub worker's position around the table. Lenses are triggered when users click on the video feeds, the lens on artifacts is rotated either by dragging the handle or simply clicking on the border at the desired direction.

a specific size would solve this problem, but would also distort the objects. Instead, we chose a focus-based approach mimicking foveal and peripheral vision to maximize variation of the satellite worker's gaze, while preventing the hub table representation from becoming distorted.

The focus-based interaction is implemented as a widget in the form of a virtual lens that focuses on content. The hub's table video is displayed in the table's actual aspect ratio and "out of focus," using a video effect to mimic the indistinct quality of peripheral vision. The satellite worker thus sees an arrangement of artifacts but not their details. To see an object's details, the satellite selects it and a round virtual lens appears on the side showing a high-resolution detail of the selected area. This lens is strategically placed on the satellite's screen so that when the satellite worker gazes at it, the hub coworkers are able to correctly interpret which object is being looked at based solely from the direction of the satellite's gaze (see Figure 1c). This supports *DR4*. The lens position is interpolated by mapping the hub table's video onto as much screen space as possible below the hub coworker's panoramic video. As artifacts can be placed on the table from different directions, we implemented a rotation control on the lens, which the satellite worker can use to rotate its content if needed (Figure 3a).

To keep the satellite worker aware of the hub coworkers' spatial arrangement around the table, we segmented the hub's panoramic video based on the table's aspect ratio, and placed the segments around the table at their corresponding sides. These segments are then also displayed out of focus. When the satellite worker wants to look at one of their hub coworkers, they select it within the panoramic view at the top of the screen a square lens widget appears to guide their gaze toward a specific person (see Figure 3b).

GazeLens is implemented using C# and .NET 4.5 framework[5]. In its current implementation, the panoramic video height is equal to 20% of the entire screen height. When displayed on a 14-inch 16:9 conventional laptop screen, this creates a distance of around 4cm from the built-in camera to the top edge of the screen showing the panoramic video of the hub. Assuming that the satellite worker is 45

[5] https://docs.microsoft.com/en-us/dotnet/framework/

cm from their screen, and the hub screen is placed at the center of the table edge, this 4 cm distance creates the desired visual angle of 5° between the satellite's camera and the panoramic video showing their hub coworkers, establishing direct gaze [10]. This size is also sufficient to avoid distortions in the panoramic video. The lenses are activated by a left-button mouse click event on non-touch screen computers, and by a touch down event on touch devices.

4 Study 1: Accuracy in Interpreting Satellite's Gaze

We evaluate whether *GazeLens* can improve hub coworkers' ability to interpret a satellite's gaze by comparing it to a conventional video conference system (*ConvVC*). *ConvVC* displays the hub side in full screen on the satellite's screen, still guiding their gaze towards right direction. To our knowledge, no off-the-shelf video conferencing interfaces for laptop/tablet offer better unmediated gaze towards people and artifacts than a *ConvVC*. We test the following hypotheses:

- H1: *GazeLens* improves accuracy of gaze interpretation compared to *ConvVC*;
- H2: *GazeLens* outperforms *ConvVC* for gaze interpretation accuracy both when the hub coworker sits in front of and to the side of screen; and,
- H3: *GazeLens* incurs a lower perceived workload than *ConvVC*.

4.1 Method

The study has a within-subjects design with the following factors:

- INTERFACE used by the satellite worker with levels: *GazeLens* and *ConvVC*;
- POSITION of the hub participants around the hub table with levels: *Front* and *Side* of the screen.

We controlled two secondary factors ACTOR and TARGET. We recorded 3 video sets of different ACTORs to mitigate possible effects tied to one of them in particular. Each ACTOR gazed at 14 TARGETs located on and around the table (Figure 4) as if he/she was the satellite worker.

Conditions were grouped by POSITION, then by INTERFACE and then by ACTOR. The presentation order of these three conditions was counterbalanced using Latin squares. Each Latin square row was repeated when necessary. For each POSITION × INTERFACE × ACTOR condition, the order of the 14 TARGETs was randomized so that successive videos never showed the same target as the previous one (and with a different ACTORs). Participants performed in total 168 trials (2 POSITIONs × 2 INTERFACEs × 3 ACTORs × 14 TARGETs).

4.2 Participants

Twelve participants (7 male), aged 22 to 33 (median = 25), with backgrounds from computer science, interaction design, and social science participated in the study. This sample size is the average one reported in CHI studies [38] and also used in related work [17, 31]. Pilot studies determined that effects are strong

enough to be observed with this sample size. All participants had normal or corrected-to-normal vision. Three never used video conferencing applications, two used them on a monthly basis, five on a weekly basis, one on a daily basis and one multiple times a day. Each received a movie ticket for their participation.

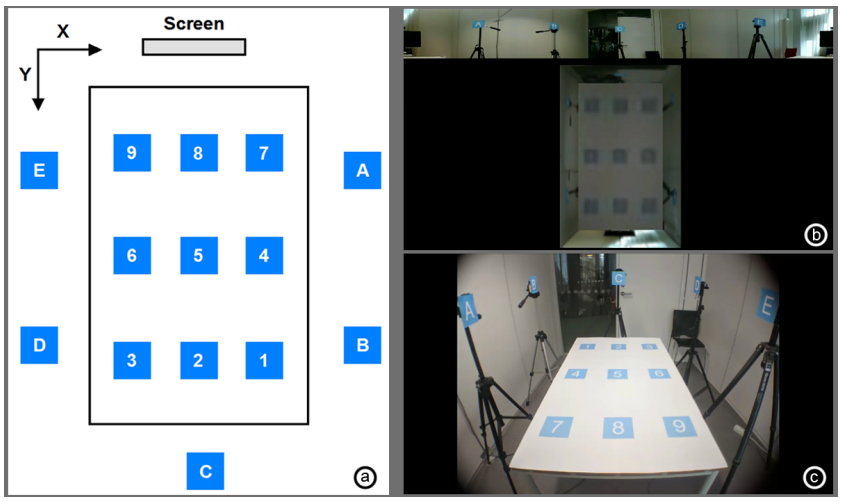

Fig. 4: (a) Hub space with target arrangement as used in Study 1. (b) *GazeLens* interface and (c) conventional interface *ConvVC* with the experimental setup.

4.3 Hardware and Software

For video recording the stimuli, we used a conventional laptop, a typical commercial foldable laptop with a screen size of $11-17$ inches (here, 14 inches) with a built-in front facing camera, for the satellite worker, as it is still one of the most common device used by travelers. Due to the low resolution of our laptop's built-in camera, we used a Plexgear 720p webcam[6] mounted at the same position as the laptop's built-in camera to record ACTORs. Using a high-res camera does not reduce the validity of the study as we are not investigating the effect of image quality. Also, many current conventional laptop models have high resolutions.

On the hub side, we used a 80cm × 140cm rectangular table with a height of 60cm to accommodate 6 people. The 14 TARGETs ($12cm \times 12cm$) were divided into two groups: 9 (labeled from 1 to 9) arranged in a 3×3 grid on the hub table represented artifacts, and 5 (labeled from A to E) around the table representing coworker targets. This left one edge occupied by the screen, two targets on each 140cm vertical edge and one on the 80cm horizontal edge (Figure 4). We used 5 hub coworkers in the study as it is a typical small team size and offers sufficient

[6] https://www.kjell.com/se/sortiment/dator-natverk/datortillbehor/webbkameror/plexgear-720p-webbkamera-p61271

challenge for interpreting gaze. Hub coworkers were 65cm higher than the table, approximated to the average eye-level height of a person sitting on a 45cm-high chair, the average for office chairs [23]. The distance from a coworker to its nearest neighbor (proxy or hub screen) was 80cm, and to the nearest edge of the table was 35cm. We used a 25-inch monitor to display the satellite's video stream, placed on a stand with the same height as the table.

To capture all the hub's targets in the *ConvVC* condition, due to our laboratory's hardware constraints, we simulated a wide-angle camera by coupling the 12-megapixel back camera of a LG Nexus 5X phone with a 0.67X wide-angled lens. The phone was mounted above the screen and adjusted so that the proxies were captured near the top edge of the video to convey the satellite worker's direct eye contact when looking at these targets (Figure 4-right/bottom). For the *GazeLens* condition, we used a Ricoh R 360° camera to capture panoramic video and a Logitech HD Pro Webcam C910 to capture the overview video of the table (Figure 4-right/top).

Participants sat at two positions around the hub's table: in *Front*, opposite the screen and at the of the screen (positions A and C respectively on 4a). The distance from the participant's body to the nearest edge was around 35cm. As the table was symmetric, the evaluation result from one side could be applied to the other. We chose position A as it was closer to the screen than B or D, causing the so-called Mona Lisa effect (where the image of a subject looking into the camera is seen by remote participants as looking at them, irrespective of their position) that could affect the gaze interpretation. The recorded videos were displayed at full screen. The videos' aspect ratio (4:3) mismatched the hub's screen (16:9). However, we did not modify the videos' size to avoid a partial or distorted view of the hub.

4.4 Procedure

After greeting participants, they signed a consent form and read printed instructions. Participants answered a pre-study questionnaire providing their background and self-assessing their technological expertise. They completed a training session before starting the experiment. We encouraged them to take a 5-minute break between the two POSITION conditions and a 2-minute break between each 21 videos (middle and at the end of each INTERFACE condition). It took 1 hour 15 minutes for a participant to complete the study. Finally, they answered the post-study questionnaires and received a movie ticket.

Video Recordings. We recorded 6-second videos of 3 different ACTORs gazing at 14 targets in the satellite worker interface displayed on a 14-inch laptop screen, for both *GazeLens* and *ConvVC*. We observed in pilots that 6 seconds are long enough to make ACTOR's gaze movements perceivable while avoiding fatigue. *ActorA* was a 29-year-old man with brown medium-length hair and hazel eyes, *ActorB* a 34-year-old man with short blonde hair and brown eyes, and *ActorC* a 44-year-old woman with brown pulled back hair and hazel eyes.

ACTORs sat on a 45cm-high office chair 45cm away from the laptop screen, which was placed on a 70cm-high office desk. In order to recreate a more realistic

gaze, actors first looked at a starting point and then at the target. This causes relative movements in the satellite worker's gaze, which provides a context with easier interpretation for the viewer. A target's starting point was decided by choosing its nearest neighboring label at an arbitrary point of 50cm, with the exception of labels on the same grid row and column as the target. As humans are less sensitive to vertical changes of gaze [10] and the distance between two targets on the same row in conventional videos is smaller, satellite workers eye movements become be noticeable. For each target, we recorded actors gazing at them from three different starting points. We did not use a chin-rest for the actors to make the recording realistic, however they were instructed to keep their head straight. They were also instructed to look at the targets in natural ways (i.e. they could turn their head if needed).

Task. Participants were advised to sit upright at POSITIONs, and could lean back if they got tired. However, if seated at the *Side*, they were not allowed to lean toward the screen. Participants watched each video playing in an infinite loop to avoid missing gaze movements due to distractions. There was thus no time pressure for the participants as we focused on accuracy. When they were ready to answer which target the ACTOR was looking at, they tapped a large "Stop" button on an Asus Nexus 9 tablet. The tablet then showed a replica of the table with the targets and hub coworkers laid out in the same fashion as on the participant's screen to make selection easier.

4.5 Data Collection and Analysis

We collected participants' responses for each trial, i.e. which target they thought was being gazed at, and their confidence in their answer (on 5-point Likert scale: 1 = not confident, 5 = very confident). We also recorded response time. When two INTERFACE conditions for each POSITION were completed, participants answered a post-questionnaire indicating their perceived workload (based on NASA TLX [33]), perceived ease to differentiate gazes at targets on and around the table, perceived ease to interpret the satellite's gaze and their interpretation strategies in both conditions.

We define *Gaze Interpretation Accuracy* as the proportion of correct trials. We define *Differentiation Accuracy*, i.e. the participant's ability to differentiate gaze at targets *around* or *on* the table, as the proportion of trials with gaze at the correct set of targets on or around table.

4.6 Results

To analyze *Gaze Interpretation Accuracy* we perform a two-way factorial ANOVA (INTERFACE × POSITION). The result (Figure 5) shows an effect of INTERFACE ($F_{1,44} = 7.33$, $p < 0.001$), POSITION ($F_{1,44} = 6.88$, $p < 0.01$) and no interaction effect INTERFACE × POSITION ($p > 0.1$). *GazeLens* significantly improves interpretation of the satellite gaze in comparison to *ConvVC* ($31.45\% \pm 4.67\%$ vs $25\% \pm 3.51\%$, an increase of 25.8%, 6.45% effect size), supporting H1. As expected, participants interpreted the satellite's gaze significantly more precisely

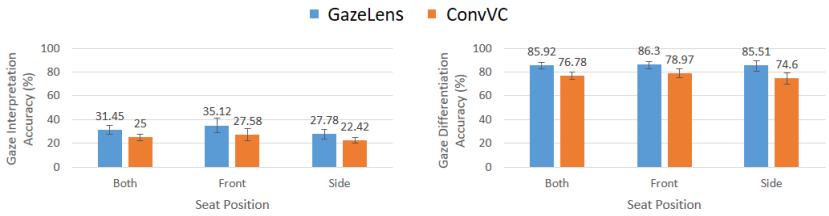

Fig. 5: *Gaze Interpretation Accuracy* (in %) (left) and *Gaze Differentiation Accuracy* (in %) (right) for INTERFACE ×POSITION. Error bars show 95% confidence interval (CI).

at the *Front* than *Side* POSITION (31.35% ± 4.96% vs 25% ± 3.13%). As there is no interaction effect, we cannot reject H2. Data in Figure 5 suggests that participants performed better with *GazeLens* at both sitting positions.

To analyze *Differentiation Accuracy* we perform a two-way factorial ANOVA (INTERFACE × POSITION). The result (Figure 5) shows an effect of INTERFACE ($F_{1,44} = 20.77$, $p < 0.001$), no effect of POSITION nor an interaction effect of INTERFACE ×POSITION ($p > 0.1$). Participants using *GazeLens* could better differentiate gaze at targets around or on the table compared to *ConvVC* (85.92%±2.38% vs. 76.78%±3.14%, $p < 0.001$).

A two-way factorial ANOVA analysis (INTERFACE × POSITION) did not yield any effect of INTERFACE × POSITION on perceived workload (not supporting H3), neither for answer time nor confidence (all p's > 0.1). Finally, we did not find any effect of ACTOR on *Gaze Interpretation Accuracy* nor *Differentiation Accuracy* at targets on or around table, neither learning effects.

5 Study 2: Accuracy in Interpreting Gaze at Hub Artifacts

Study 1 showed that *GazeLens* improves gaze interpretation accuracy in general. We wanted to further investigate how accurately hub coworkers can interpret the satellite worker's gaze towards physical artifacts on the hub table. In reality, arrangements of physical artifacts on the table can vary from sparse (e.g. meetings with some paper documents) to dense (e.g. brainstorming with sticky notes, phones, physical prototypes). This prompts the need to explore how the granularity of artifact arrangement impacts a hub coworker's gaze interpretation. We also investigate if *GazeLens* can increase hub coworkers' accuracy at interpreting the satellite's gaze compared to *ConvVC*, especially regarding error distance along the two table dimensions: horizontal (X) and vertical (Y). We used a similar experiment design to Study 1, where participants have to determine the satellite's gaze in prerecorded videos displayed on the screen at the hub table.

We operationalize artifacts arrangements through the granularity of layouts:

- 3×3 (9 objects in a 3×3 grid): low-granularity arrangements to investigate gaze interpretation accuracy in meeting scenarios involving paper documents,
- 5×5 (25 objects in a 5×5 grid): high-granularity arrangements to investigate gaze interpretation accuracy in scenarios such as brainstorming.

We formulate the following hypotheses:

- H1: *GazeLens* improves the hub coworkers' interpretation accuracy for gaze toward objects on the hub table compared to *ConvVC*;
- H2: *GazeLens* outperforms *ConvVC* for gaze interpretation accuracy at both levels of granularity;
- H3: *GazeLens* reduces X and Y error distance compared to *ConvVC*.

5.1 Method

The within-subject study design has the following factors:

- INTERFACE used by the satellite to view the targets, with two conditions: *GazeLens* and *ConvVC*; and,
- LAYOUT of the artifacts on the table with two conditions: 3×3 and 5×5 grid.

For each participant, the conditions were grouped by LAYOUT, then by INTERFACE and then by ACTOR. ACTORs were the same as in Study 1.

The order of presentation was counterbalanced across conditions using Latin squares for the first three conditions and randomized order for TARGET. Each Latin square was repeated when necessary. For each LAYOUT × INTERFACE × ACTOR condition, the order of the targets (9 for 3×3 and 25 for 5×5) was randomized so a different succession of videos was shown for each target. Participants took a 5-minute break between the two layout conditions, and performed a training session before starting the experiment, where we ensured they covered all TARGETs, INTERFACEs and LAYOUTs.

5.2 Participants

Twelve participants—different from those in Study 1—8 males, aged 22 to 38 (median = 29), with backgrounds from computer science, interaction design, and social science participated in the study. All had normal or corrected-to-normal vision. Three used their computer on daily basis, eight multiple times a day and one on a weekly basis. One had never used video conferencing applications, eight used them on a monthly basis, one on a weekly basis, one on a daily basis, and one multiple times a day. Each received a movie ticket for their participation.

5.3 Hardware and Software

We used the same cameras, hub table, hub screen, and screen placement as in Study 1. To investigate gaze interpretation accuracy for different artifact sizes, we used two different layouts on the table (Figure 4). We removed targets representing hub coworkers in Study 1 to avoid distracting actors and participants.

5.4 Procedure

We employed a similar procedure as in Study 1. However, participants took a 2-minute break after every 18 videos in the 3×3 layout, and after every 15 videos in the 5×5 layout (the dense layout was more tiring).

Video Recordings. We recorded 306 6-second videos for the hub's targets of the same three ACTORs as in study 1: 81 videos for the 3×3 layout and 225 for the 5×5 layout. We used the same laptop, camera, placements of the devices and ACTORs as in Study 1. Each video was also recorded in a similar procedure as in Study 1: each ACTOR first looked at a starting point and then at the target. We used the same criteria for choosing starting points for targets.

Task. We used a similar task as in Study 1, although participants only sat at position A (*Front*) to watch the videos. The positions of the screen and those of participants in relation to it remained the same.

5.5 Data Collection and Analysis

We collected data as in Study 1. We measure *Gaze Interpretation Accuracy* as in Study 1 and two error measures: *X-Axis Error* and *Y-Axis Error*, denoting the error between the correct and selected target along the table's horizontal and vertical orientation (X and Y axis in Figure 4a) respectively.

5.6 Results

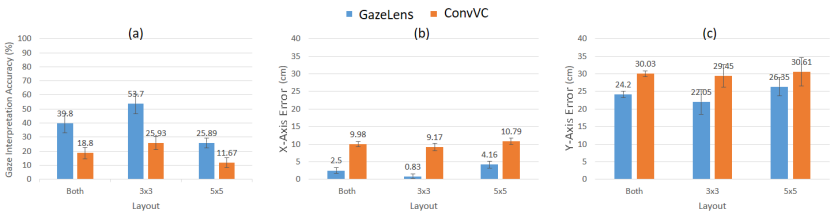

Fig. 6: (a) Gaze Interpretation Accuracy (in %) for each INTERFACE × LAYOUT condition. (b) X-Axis Error (in cm) and (c) Y-Axis Error (in cm) for each INTERFACE × LAYOUT condition. Bars indicate 95% CI.

We analyze *Gaze Interpretation Accuracy* as in Study 1 by performing a two-way factorial ANOVA (INTERFACE and LAYOUT). The result shows an effect of INTERFACE ($F_{1,44} = 69.26$, $p < 0.001$), supporting H1, LAYOUT ($F_{1,44} = 69.50$, $p < 0.001$) and INTERFACE × LAYOUT ($F_{1,44} = 7.214$, $p < 0.05$). A post-hoc Tukey HSD test showed that *GazeLens* significantly improves *Gaze Interpretation Accuracy* in both 3×3 layout (53.7%±7.06% vs 25.93%±4.81%,$p <$ 0.001) and 5×5 layout (25.89%± 3.55% vs 11.67%± 3.5%, $p < 0.001$) supporting H2. Post-hoc Tukey HSD tests showed significant differences between *GazeLens* with 3×3 and *GazeLens* with 5×5 ($p < 0.001$), between *GazeLens* with 3×3 and *ConvVC* with 5×5 ($p < 0.001$), between *ConvVC* with 3×3 and *ConvVC* with

5×5 ($p < 0.01$). Figure 6a shows gaze interpretation accuracy in each INTERFACE × LAYOUT condition.

To examine *X-Axis Error*, we perform a two-way factorial ANOVA analysis with INTERFACE and LAYOUT as factors. The analysis shows an effect of INTERFACE ($F_{1,44} = 251.59$, $p < 0.001$), partially supporting H3, LAYOUT ($F_{1,44} = 27.54$, $p < 0.001$) and no effect of INTERFACE × LAYOUT ($F_{1,44} = 3.255$, $p > 0.05$). For *Y-Axis Error* we perform a two-way factorial ANOVA analysis with INTERFACE and LAYOUT as factors. The analysis shows an effect of INTERFACE ($F_{1,44} = 11.29$, $p < 0.005$), partially supporting H3, but no effect of LAYOUT and INTERFACE × LAYOUT (all $p > 0.1$).

We did not find any effect of INTERFACE on answer time, self-confidence, perceived workload and perceived ease of gaze interpretation (all $p > 0.1$). No learning effect was found in term of gaze interpretation accuracy, X and Y-axis error. Figure 7 visualizes the X and Y error distances at each target by INTERFACE and LAYOUT. Figure 6 (b,c) shows X and Y error distance in each INTERFACE × LAYOUT condition.

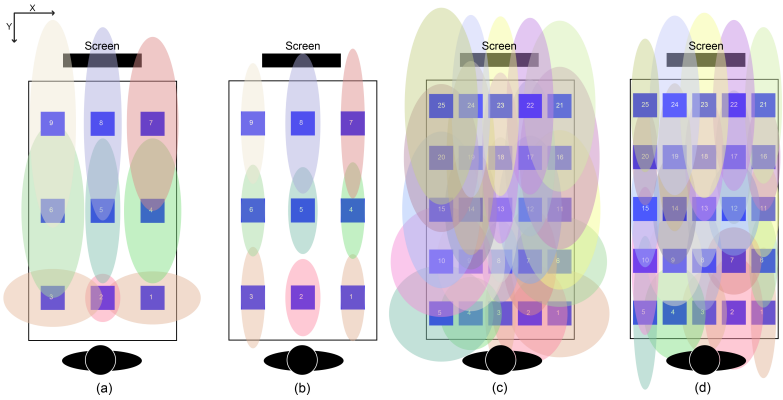

Fig. 7: X and Y-Axis Error visualization at each target in Study 2, in (a) 3×3 layout using *ConvVC*, (b) 3×3 layout using *GazeLens*, (c) 5×5 layout using *ConvVC*, (d) 5×5 layout using *GazeLens*. Zero error is shown by an ellipse-axis equal to the target size.

6 Early User Feedback of GazeLens

The two previous experiments evaluated *GazeLens* on the hub side. We gather in a last study early user feedback on its usability from the satellite worker's perspective. We recruited five pairs of participants (8 male, 2 female, aged from 23 to 50, median 31) to solve a remote collaborative task. Participants had various backgrounds from computer science, software engineering, and social sciences. Participants in each pair knew each other well. Designing an experimental collaborative task for hub-satellite collaboration involving physical artifacts is complicated by the complex communication required between coworkers and artifacts.

To our knowledge, there is still no standardized experimental task for this. As we focus on gathering feedback on the satellite's side, for simplicity, we chose a standard task commonly used when investigating remote collaboration on physical tasks: solving a puzzle by arranging a set of pieces into a predefined picture.

Each pair of participants consisted of a *worker* on the hub side and an *instructor* on the satellite side. The *worker* had all the puzzle pieces on the hub table, but did not know the solution. The *instructor* knew the solution and communicated with the worker via audio and video to guide them selecting and arranging the pieces. This task can trigger movements of the hub worker, their hands, and the artifacts on and around the table, which could be perceived differently by the satellite worker on different video conferencing interfaces. Each puzzle consists of 16 rectangular pieces, chosen so that they were hard to be verbally described by color and visual patterns. We used the same laboratory setup as in Study 1 and 2.

Participants performed the tasks in both interfaces on the satellite side, *GazeLens* and *ConvVC*, in order to have comparative views on their usability. They familiarized themselves for about 15 minutes with each interface, and had 10 minutes to solve the task in each condition. Two different $50cm \times 80cm$ puzzles with comparable levels of visual difficulty were used for two conditions. The conditions and puzzle tasks were counter-balanced. We gathered qualitative feedback in an interview after participants went through both conditions.

Only one *instructors* our of four reported perceiving inconvenient using *GazeLens* lenses. He reported that activating the lens on the table by mouse click was quite tiring and suggested using mouse wheel scroll events to make the activation similar to zooming. All *instructors* reported that it was easier for them to see the puzzle pieces' content in *GazeLens*, as those at the far edge of the table were hard to see in *ConvVC*, where they sometimes had to ask the performer to hold and show it to the camera. One *instructor*, who often uses Skype for hub-satellite meetings, really liked the concept of people around the table in a panoramic video connected via a virtual lens. He could imagine that it could help him clearly see everyone while knowing where they are sitting around the table and what they are doing on the table during a meeting. Besides that, two *instructors* reported that when the virtual lens was on the top of *GazeLens*' table area it might obscure *workers'* hand gestures.

7 Discussion

7.1 GazeLens Improves Differentiating Gaze Towards People vs. Artifacts

Study 1 showed that *GazeLens* improves the satellite worker's gaze interpretation accuracy toward hub coworkers and, in particular, they are able to distinguish with more than 85% accuracy if the satellite worker is gazing towards them or towards the physical artifacts on the table. This is due to the position of the panoramic video of hub coworkers' at the top of the satellite interface, close to the

camera, and the position of the artifact overview at the bottom of the interface. To further improve this, we could explore increasing the gap between these two views to obtain a larger, more distinguishable, difference in gaze direction.

7.2 GazeLens Improves Gaze Interpretation Accuracy

Study 2 showed that *GazeLens* improved gaze interpretation accuracy for table artifacts not only in sparse (3×3) but also dense (5×5) arrangements. This can be explained by how the entire laptop screen is used to make the satellite's worker's gaze better aligned with the hub artifacts as compared to the *ConvVC* condition, as stated by P10: *"To determine where the satellite worker was gazing, I could imagine a line of sight from his eyes to the table/person"*. We argue that in *ConvVC*, due to the perspective projection of the hub table, distances between objects at the far edge of the table appear too small, making gaze toward them indistinguishable. This was confirmed by participant comments about the satellite's gaze in *ConvVC*: *"It was hard compared to the other condition [*GazeLens *] because they just stared at the table and I had no clue which number was the exact one" (P6)*; *"the differences between different gazes felt very small" (P5)* and *"They were looking more at the center" (P9)* . In contrast to *ConvVC*, participants perceived the satellite worker's gaze in the *GazeLens* condition as *"more obvious"*, *"clearer"* and *"easier to determine where they are looking"*.

Small between-object distances in *ConvVC* also caused negligible eye movements in the videos where the satellite worker gazes from the starting point to the target: *"They moved less and thus gave me fewer references to be able to get a picture of what they were looking at"*; *"Eye movements were very small and the angles were hard to calculate in my head"* and *"The eyes did not move and I got confused"*. In contrast, participants perceived eye movement in *GazeLens* condition as *"clearer"*, *"easy to distinguish from side to side"*, *"enough to follow"*, *"sometimes added with head movements, easier to determine"*.

When investigating X-Axis and Y-Axis Error, it was not surprising that horizontal gaze changes were perceived more accurately than vertical ones, as people are more sensitive to horizontal gaze changes, especially when the gaze is below the satellite's camera [10]. Furthermore, laptops have landscape screens, leaving less vertical space to position the lens than in the horizontal direction-making gaze differences more distinguishable in the horizontal orientation. In future work, we want to explore how to improve gaze perception in the vertical dimension.

7.3 Limitations and Future Work

Although most of the participants reported that the satellite worker's gaze was clear and easy to interpret with *GazeLens*, two participants in Study 1 reported that they did not feel the satellite worker was looking at any markers in particular, and their answers were just an approximation based on gaze. This can be explained by the fact that at that moment *GazeLens* did not precisely calculate the screen mapping based on the actual size of the table and the distance from

the coworkers to the hub table. Achieving geometrically corrected gaze in video communication is almost impossible, as it depends on several parameters that cannot all be easily acquired in real-life hub-satellite scenarios, such as camera focal length, camera position, video size, camera-scene, and screen-viewer distance. *GazeLens*' mapping strategy is effective at improving gaze interpretation and yet simple enough to be deployed in realistic scenarios. In future, we will consider replacing the ceiling-mounted camera with a depth-sensing camera, which can acquire the table size and coworkers's distance from the table in order to improve mapping. We are also interested to further study *GazeLens* with different hub table shapes, sizes and layouts, using the current screen mapping strategy and others. Likewise, due to the emerging use of tablets for work purposes, it would be valuable to study *GazeLens* on tablet devices both in portrait and landscape display mode.

In our last study, participants perceived *GazeLens* positively, without usability issues. Still, we plan to improve the system by making the virtual lens over the table less occlusive in the future setup using a depth-sensing camera, by detecting the presence of hub workers' hand gestures and dynamically adjusting the opacity of the lens. Besides that, we plan to study how expertise might influence time needed to learn *GazeLens*, as we think that probably this is not enough to make an impact on the hub side, which is unaware of what is shown on the satellite's interface. Lastly, we plan to extended *GazeLens* to support multiple satellites, for instance by representing each one by a screen placed around the hub table and the corresponding video feeds adjusted accordingly (e.g. re-segment panoramic video, change orientation of the table's video).

8 Conclusion

While conventional hub-satellite collaboration typically employs video conferencing, it is difficult for hub coworkers to interpret the satellite worker's gaze. Previous work supporting gaze between remote workers has not addressed shared physical artifacts used in collaboration, and support for conveying gaze in remote collaboration with asymmetric setups is still limited. We designed *GazeLens*, a novel interaction technique supporting gaze interpretation that guides the attention of the satellite worker by means of virtual lenses focusing on either hub coworkers or artifacts. In our first study, we showed that *GazeLens* significantly improves gaze interpretation over a conventional video conferencing system; and also that it improves hub coworkers' ability to differentiate the satellite's gaze toward themselves or artifacts on the table. In our second study, we found that *GazeLens* improves hub coworkers' interpretation accuracy for gaze toward objects on the table, for both sparse and dense arrangements of artifacts. Early user feedback informed us about the advantages and potential drawbacks of *GazeLens*' usability. *GazeLens* shows that the satellite worker's laptop screen can be fully leveraged to guide their attention and help hub coworkers more accurately interpret their gaze.

References

1. Vertegaal, R.: The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In: CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 294–301, ACM New York,1999.
2. Akkil, D., James, J.M.,Isokoski, P., Kangas, J.: GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In: CHI EA '16 Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1151–1158, ACM New York,2016.
3. Vertegaal, R.,van der Veer, G., Vons, H.: Effects of Gaze on Multiparty Mediated Communication. In: Proceedings of Graphics Interface 2000, pp. 95–102, ACM New York,2010.
4. Vertegaal, R., Slagter, R., Van der Veer, G., Nijholt, A.: Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: CHI '01 Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 301-308, ACM New York, 2001.
5. Higuch, K., Yonetani, R., Sato, Y.: Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In: CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5180–5190, ACM New York, 2016.
6. Xu, B., Ellis, J., Erickson, T.: Attention from Afar: Simulating the Gazes of Remote Participants in Hybrid Meetings. In: DIS '17 Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 101–113, ACM New York, 2017.
7. Sellen, A., Buxton, B., Arnott, J.: Using spatial cues to improve videoconferencing. In: CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 651–652, ACM New York, 1992.
8. Nguyen, D., Canny, J.: MultiView: spatially faithful group video conferencing. In: CHI '05 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 799–808, ACM New York, 2005.
9. Pan, Y., Steed, A.: A Gaze-preserving Situated Multiview Telepresence System. In: CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2173–2176, ACM New York, 2014.
10. Chen, M.: Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconferencing. In: CHI '02 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 49–56, ACM New York, 2002.
11. Ishii, H., Kobayashi, M.: ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In: CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 525–532, ACM New York, 1992.
12. Hauber, J., Regenbrecht, H., Billinghurst, M., Cockburn, A.: Spatiality in videoconferencing: trade-offs between efficiency and social presence. In: CSCW '06 Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, pp. 413–422, ACM New York, 2006.
13. Otsuka, K.: MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In: Proceedings of 2016 IEEE Virtual Reality (VR), IEEE, 2016.
14. Küchler, M., Kunz, A.: Holoport-a device for simultaneous video and data conferencing featuring gaze awareness. In: Proceedings of Virtual Reality Conference 2006, pp. 81–88, IEEE, 2006.
15. Jones ,A. , Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., Debevec, P.: Achieving eye contact in a one-to-many 3D video teleconferencing system. In: ACM Transactions on Graphics (TOG), **28**(3),2009.

16. Gotsch, D., Zhang, X., Meeritt, T., Vertegaal, R.: TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-size 3D Human Telepresence. In: CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 552, ACM New York, 2018.
17. Otsuki, M., Kawano, T., Maruyama, K., Kuzuoka, H., Suzuki, Y.: ThirdEye: Simple Add-on Display to Represent Remote Participant's Gaze Direction in Video Communication. In: CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5307–5312, ACM New York, 2017.
18. Vertegaal, R., Weevers, I., Sohn, C., Cheung, C.: GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In: CHI '03 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 521–528, ACM New York, 2003.
19. Norris, J., Schnädelbach, H., Qiu, G.: CamBlend: An Object Focused Collaboration Tool. In: CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 627-636, ACM New York, 2012.
20. Bailey, R., McNamara, A., Sudarsanam, N., Grimm, C.: Subtle gaze direction. In: ACM Transactions on Graphics (TOG) **28**(4), 2009.
21. Hata, H., Koike, H., Sato, Y.: Visual Guidance with Unnoticed Blur Effect. In: AVI '16 Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 28–35, ACM New York, 2016.
22. Stokes, R.: Human Factors and Appearance Design Considerations of the Mod II PICTUREPHONE Station Set. In: ACM Transactions on Graphics (TOG) **17**(2), 1969.
23. Average Human Sitting Posture Dimensions Required in Interior Design, https://gharpedia.com/average-human-sitting-posture-dimensions-required-in-interior-design/.
24. Gemmell, J., Toyama, K., Zitnick, C.L., Kang, T., Seitz, S.: Gaze awareness for video-conferencing: a software approach. In: IEEE MultiMedia **7**(4),2000.
25. Giger, D., Bazin, J-C, Kuster, C., Popa, T., Gross, M.: Gaze correction with a single webcam. In: 2014 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2014.
26. Venolia, G., Tang, J., Cervantes, R., Bly, S., Robertson, G., Lee, B., Inkpen, K.: Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In: CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1049–1058, ACM New York, 2010.
27. Kendon, A.: Some functions of gaze-direction in social interaction. In: Acta Psychologica **26**, 1967.
28. Brennan, S. E., Chen, X., Dickinson, C.A., Neider, M.B., Zelinsky, G.J.: Coordinating cognition: the costs and benefits of shared gaze during collaborative search. In: Cognition **106**(3),2008.
29. Akkil, D., Isokoski, P.: I See What You See: Gaze Awareness in Mobile Video Collaboration. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA '18, p. 32, ACM New York, 2018.
30. Yao, N., Brewer, J., D'Angelo, S., Horn, M., Gergle, D.: Visualizing Gaze Information from Multiple Students to Support Remote Instruction. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, p. LBW051, ACM New York, 2018.
31. Avellino, I., Fleury, C. and Beaudouin-Lafon, M.: Accuracy of Deictic Gestures to Support Telepresence on Wall-sized Displays. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 2393–2396, ACM New York, 2015.

32. Monk, A.F., Gale, C.: A look is worth a thousand words: Full gaze awareness in video-mediated conversation. In:Discourse Processes **33**(4), pp. 257-278, 2002.
33. Hart, S.G. and Staveland, L.E: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Advances in psychology **52**, pp. 139-183, 1998.
34. Cisco TelePresence MX Series, https://www.cisco.com/c/en/us/products/collaboration-endpoints/telepresence-mx-series/index.html. Last accessed 26 Jan 2019
35. RealPresence Group Series, http://www.polycom.com/products-services/hd-telepresence-video-conferencing/realpresence-room/realpresence-group-series.html. Last accessed 26 Jan 2019
36. Cisco CTS-SX80-IPST60-K9 TelePresence (CTS-SX80-IPST60-K9), https://www.bechtle.com/ch-en/shop/cisco-cts-sx80-ipst60-k9-telepresence–896450-40–p. Last accessed 26 Jan 2019
37. Enterprise Video Conference, http://www.avsolutions.com/enterprise-video-conference. Last accessed 26 Jan 2019
38. Caine, K.: Local standards for sample size at CHI. In: Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 981–992, ACM New York, 2016.

Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users

ARTHUR FAGES, Université Paris-Saclay, CNRS, Inria, LISN, France
CÉDRIC FLEURY, IMT Atlantique, Lab-STICC, UMR CNRS 6285, France
THEOPHANIS TSANDILAS, Université Paris-Saclay, CNRS, Inria, LISN, France

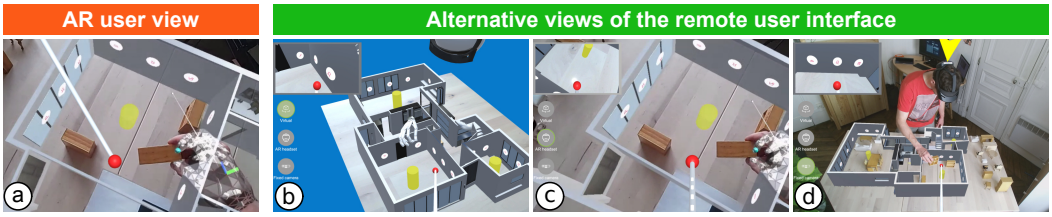

Fig. 1. An AR user and a remote desktop collaborator perform a physical furniture arrangement task around a virtual 3D house model. (a) AR user's view displayed in the headset and three alternative views of the remote collaborators available in the ARgus interface: (b) a fully virtual view, (c) a first-person view streamed from the headset, and (d) an external view streamed from a depth camera. Project page: https://argus-collab.github.io

Establishing an effective collaboration between augmented-reality (AR) and remote desktop users is a challenge because collaborators do not share a common physical space and equipment. Yet, such asymmetrical collaboration configurations are common today for many design tasks, due to the geographical distance of people or unusual circumstances such as a lockdown. We conducted a first study to investigate trade-offs of three remote representations of an AR workspace: a fully virtual representation, a first-person view, and an external view. Building on our findings, we designed ARgus, a multi-view video-mediated communication system that combines these representations through interactive tools for navigation, previewing, pointing, and annotation. We report on a second user study that observed how 12 participants used ARgus to provide remote instructions for an AR furniture arrangement task. Participants extensively used its view transition tools, while the system reduced their reliance on verbal instructions.

CCS Concepts: • **Human-centered computing** → **Computer supported cooperative work**; **Mixed / augmented reality**; **Collaborative interaction**.

Additional Key Words and Phrases: Augmented reality, remote collaboration, video-mediated communication.

ACM Reference Format:
Arthur Fages, Cédric Fleury, and Theophanis Tsandilas. 2022. Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users. *Proc. ACM Human-Computer Interaction* 6, CSCW2, Article 549 (November 2022), 27 pages. https://doi.org/10.1145/3555607

549

1 INTRODUCTION

Augmented reality (AR) technologies radically change the way 3D design teams work together. AR users can move away from the screen of their computer to interact directly with the objects of a virtual scene and naturally navigate in their physical space. AR also strengthens collaboration by

Authors' addresses: Arthur Fages, Arthur.Fages@lisn.upsaclay.fr, Université Paris-Saclay, CNRS, Inria, LISN, Orsay, France; Cédric Fleury, Cedric.Fleury@lisn.upsaclay.fr, IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest, France; Theophanis Tsandilas, Theophanis.Tsandilas@inria.fr, Université Paris-Saclay, CNRS, Inria, LISN, Orsay, France.

adding virtual aids [39] while preserving traditional communication channels, such as voice, gaze and gestures. Previous work has investigated the use of AR for a diverse range of collaborative tasks, from interior design for couples [48] and science teaching [51] to industrial manufacturing [60]. Unfortunately, real-time collaboration is a challenge when users work remotely and, consequently, they do not share the same physical environment and do not all have access to AR equipment. Such situations have become commonplace during the still ongoing COVID19 pandemic [5]. Many design and research teams have found themselves to work remotely, relying on video-communication software to collaborate together [63]. Some experts predict that such situations are not temporary – they will largely persist after the pandemic [10]. HCI research thus needs to better understand how different remote workspace configurations support collaboration in these new contexts.

While screen sharing has been a valuable tool of collaboration for remote desktop users, sharing the workspace of a collaborator wearing an AR headset requires a new set of tools that considers both the physical and the virtual space of the AR user. In this direction, several AR technologies such as the Microsoft HoloLens enable AR users to video-stream their view. Yet, such views are not interactive and do not offer independent camera control to remote viewers. According to Tait and Billinghurst [52], increased view independence results in stronger collaboration performance. However, view independence requires that the physical environment of the AR user is reconstructed in real time, such that it can be smoothly integrated into the 3D virtual scene. Unfortunately, existing solutions for reconstructing independent AR views have serious limitations. For example, techniques based on multiple depth sensors [7, 9] require heavyweight instrumentation, consume large volumes of bandwidth, while the quality of their reconstructed models is still limited and largely unrealistic [25]. Other 3D reconstruction techniques [19, 36] pose significant constraints on the view possibilities of remote users.

The alternative approach that we investigate here is to offer remote users multiple view representations, where each provides a different aspect of the workspace of the local AR worker. We focus, in particular, on tasks that require access both to a virtual model and to its physical context, or to physical objects that interact with the virtual model. In this case, remote collaborators must make decisions about which representation to use to effectively complete the task. We study three complementary representations: (i) a first-person view as provided by the AR headset, (ii) an augmented third-person view as captured by a fixed camera with a depth sensor, and (iii) a fully virtual representation. The first two representations show the real-word scene but do not support view independence. The last representation, in contrast, supports full view independence but does not capture the real-world scene. However, by providing tools for switching between these representations, we expect that remote users will develop strategies that leverage their complementary roles. We frame our research questions as follows:

- **RQ_1: How do remote users perceive the trade-offs of the three representations when providing instructions to an AR worker?** Several past studies [18, 20, 51] have studied the trade-offs of first-person and third-person views, but as we discuss in this paper, their results are somehow contradictory and non-conclusive. Others [52] have studied fully independent views, but only ones that rely on the 3D reconstruction of the real scene.
- **RQ_2: If we offer remote users the possibility to switch between representations, how will they make use of them?** To explore answers to this question, we integrate the three representations into ARgus (see Fig. 1), a remote collaboration system. A key contribution of ARgus is on how its user interface merges representations through a collection of interactive tools for previewing, between- and within-view navigation, camera control, 3D pointing, and annotation.

We report on the results of two user studies, one for each question. The first study examines strengths and weaknesses of the three representations, focusing on the collaboration experience of remote users when communicating spatial instructions. The second user study investigates how 12 remote participants use ARgus to guide a local AR user to complete an AR furniture arrangement task. Our results provide a fresh perspective on the trade-offs of each representation. They also help us characterize participants' view-switching strategies, evaluate the perceived effectiveness and utility of ARgus, and understand whether and how it assists remote communication.

2 RELATED WORK

Our research builds upon a rich volume of previous HCI work on remote collaboration.

The role of viewpoint in video-mediated collaboration. When people do not share the same space, video is the most common communication medium. Its role is to bring a common ground of understanding (or *conversational grounding* [18]) and support *workspace awareness* [14] or a *"shared person space"* that includes *"facial expressions, voice, gaze and body language"* [12].

The HCI literature has long examined the role of different views in video-mediated communication, especially in the context of physical tasks that involve spatial object manipulation and construction. Back in the 90s, Kuzuoka [31] investigates spatial workspace collaboration through SharedView, a video communication system. Kuzuoka's study requires a remote expert to explain a 3D task to a local worker in a machining center and shows that the viewpoint of the video can affect the efficiency of communication. Gaver et al. [20] study the use of five camera views for a remote-collaboration design task. Their task requires a participant in a local office to arrange the furniture in a dollhouse in collaboration with a remote partner. Their results show that participants largely preferred task-centered views than face-to-face communication. The authors also observe that view switching can be problematic. In particular, multiple views can interfere with establishing a common frame of reference, introduce discontinuities, and impede coordination.

Ten years later, Fussel et al. [18] compare two remote-view configurations: (i) a head-mounted camera with eye tracking, and (ii) a scene camera placed at the back of the worker, providing a wider but fixed view of the working environment. The scene camera is shown to be preferable and improve communication efficiency, while the head-camera view does not add any benefit compared to an audio-only condition. Similarly, adding a second, head-camera view to a scene-camera view seems to deteriorate rather than to improve collaboration performance. A more recent study [51] in the context AR video collaboration for 3D guidance tasks also shows that a third-person view results in better task performance and higher user satisfaction than a first-person view.

However, other studies show advantages in combining multiple alternative views. For example, Schafer and Bowman [45] study a virtual furniture arrangement task and observe that the availability of two alternative representations (virtual 3D and floor plan) *"enabled the users to investigate different aspects of the space."* Ranjan et al. [41] find that remote users complete complex lego-construction tasks faster with automatic pan-tilt-zoom camera than with a static camera. Giusti et al. [21] investigate how a local user and a remote expert configure a mobile phone and a tablet to repair a Lego model or replace a punctured bike tube. When both a phone and a tablet was available, local users tended to fix the tablet's camera view to show an overview of their workspace and sometimes their face, while they used the camera of their mobile phone when they needed to zoom in on specific parts to show details. Lanir et al. [32] investigate user performance and behavior with respect to who (the local vs. the remote user) has the camera control. Their conclusion is that the outcome depends on the situation and task at hand. Overall, results are far from conclusive but seem to suggest that the most suitable strategy is to give users control over alternative views, each

adapted to a different type of task. Our goal is to verify this hypothesis and investigate mechanisms that help users effectively control these views.

Finally, in the context of remote AR collaboration, Tait and Billinghurst [52] evaluate how varying degrees of view independence affect collaboration. They find that more independent views result in faster task-completion time, higher user confidence, and fewer verbal instructions. Unfortunately, the approach of Tait and Billinghurst [52] requires the physical environment of the local user to be reconstructed as a virtual 3D model. This model does not capture dynamic changes in the real environment, is limited in space and resolution, and does not include a natural representation of the local AR user, who is represented instead as a virtual view frustum. Furthermore, to detect the manipulation of physical objects and communicate it to remote users, the authors use a sophisticated optical tracking system and attach infrared markers to a small set of preregistered physical objects. Clearly, such configurations are extremely hard to set up and do not scale to real-world collaboration tasks. Next, we discuss the limitations of 3D reconstruction methods in more depth and present the state-of-the-art of view transition techniques.

Remote representation of an AR workspace. A key challenge for remote AR collaboration is how to communicate information about the physical space while enabling users to navigate in the scene and manipulate objects. A common solution is using virtual replicas of the physical objects. For example, Oda et al. [38] focus on remote collaboration between an expert wearing a VR headset and a local worker wearing an AR headset. Their system enables the expert to provide guidance by moving or annotating the 3D model of an existing physical part (virtual replica) in the worker's virtual space.

Unfortunately, virtual replicas provide partial only information about the physical environment of the local AR user. Feick et al. [17] combine, instead, two parallel views for a remote expert user: (i) a video feed showing the other user manipulating a physical object, and (ii) a 3D scene that allows the expert to gesture over a virtual proxy of the object. Kumaravel et al. [57] take this approach even further. They study two representations that communicate the virtual and physical workspace of a local user: (i) a 2D video stream and (ii) an *hologlyph*, a 3D representation of spatial data captured by depth cameras and rendered as a point cloud. In other mixed-reality systems, remote collaborators can switch between a 360° panorama video and a 3D reconstructed scene [56] or even navigate in a point-cloud representation of the remote workspace through multiple depth cameras that produce a real-time 3D reconstruction of the scene [9]. Other research explores techniques for communicating cues about the gaze of collaborating users [24].

Despite their technical sophistication, the above systems have serious limitations. First, they either support static 3D models or require remote users to have access to specialized and hard to set up equipment. Second, even the most compelling systems suffer form artifacts that limit the realism of the reconstructed workspace. For example, the system of Bai et al. [9] (one of the very few to support real-time scene reconstruction) can only display low-resolution 3D panoramas and simplistic avatar representations of the local user. But as Jones et al. [25] observe, the reduced quality of a full 3D reconstruction can distort collaborators' expressiveness and make them experience an *"uncanny valley of XR [extended reality] telepresence."* The authors also report that *"the more immersive an XR Telepresence system is, the more amplified technical issues such as latency, video quality, and control become"* [25].

Other very active research in AR mobile collaboration [19, 36, 49] has introduced techniques that enable remote users to interact with a reconstructed 3D representation of the remote workspace. These techniques have similar limitations. Based on KinectFusion [23], BeThere [49] requires the local user to pre-capture the 3D geometry of the workspace with a mobile depth camera and the remote user to use a device with a depth sensor to interact with it. SLAM systems [19] provide a

limited range of 3D navigation that is constrained by the image viewpoints seen by the camera of the local user. Finally, systems based on light fields [36] lack depth information, making occlusion management problematic.

Since we do not expect the above problems to be solved any time soon, we limit our scope to techniques of augmented video-mediated communication, as those require more lightweight setups, consume less bandwidth, and do not suffer from 3D reconstruction problems. Furthermore, as we study tasks that involve both virtual and physical objects, we are also interested in how streamed video can be coupled with fully virtual representations that afford free navigation.

View transition techniques. Purely virtual environments offer considerable freedom for remote collaboration through arbitrary virtual cameras and views. For example, Photoportals [30] and Spacetime [62] provide a range of imaginative techniques for viewpoint control in VR. In contrast, AR collaboration is largely constrained by the position and coordination of physical cameras in the environment of the local user. Previous work has tried to deal with this problem in different ways. Rasmussen and Huang [42] show previews from multiple cameras to remote users who can then switch between them. Sukan et al [50] enable mobile AR users to quickly switch between snapshots of their past views. Komiyama et al. [28] provide techniques for a smooth transition among the views of multiple physical cameras. Finally, Tatzgern et al [55] study how to seamlessly transition between AR and VR views. Our system design draws inspiration from all this line of work.

3 DESIGN PROBLEM

We are interested in asymmetric collaboration setups that involve a *local user* with an AR headset (e.g., a Microsoft HoloLens) and *remote collaborators* who participate from distance through a desktop application. In contrast to approaches that require users at both ends to wear an AR or a VR headset [9, 58], such setups are relatively lightweight and easy to employ, as they only require the local user to have access to AR equipment. These setups thus offer high flexibility to the *remote collaborators*, allowing them to work in many different situations, such as while traveling or in a crowded open office where physical space is limited.

Video has become the most common medium of remote collaboration and has taken a dominant role during the ongoing COVID19 pandemic [63]. Our goal is not to replace video communication but to enhance it with new visual and interaction modalities that leverage the benefits of AR systems. A major challenge is how to deal with the asymmetry in the views of remote collaborators, in particular how to enable them to easily navigate in the 3D environment of the AR user, inspect the virtual content, and provide directions that require spatial orientation and awareness.

As we already discussed, we also dismiss solutions that require the reconstruction of the physical workspace [9, 26, 52, 56], either because they cannot keep track of dynamic changes in the environment of the local user, or because they provide a largely unrealistic representation of the scene and the local user, break the collaborators' experience due to the *"uncanny valley of XR telepresence"* [25], and amplify network outage problems [8].

We restrict our design space to lightweight configurations that use a single external depth camera in addition to the camera of the AR user's headset. This external camera could be replaced by a webcam or a smartphone since more and more devices are now equipped with a depth sensor. We may even rely on standard webcams or smartphones in the near future, as a single monocular camera can be sufficient to provide depth data [34].

Focusing on the views of the *remote collaborators*, we investigate three design dimensions:

Workspace representation. It refers to the representation used by the system to help collaborators perceive each other and their shared workspace. This representation may consist of a virtual 3D scene, video, or alternatively a combination of these two. Ideally, it should

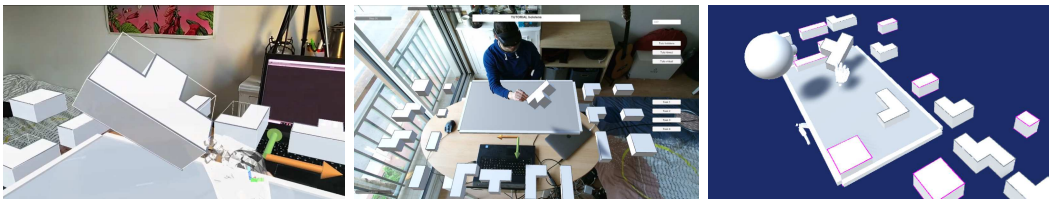

Fig. 2. Remote-view configurations tested by our first study: Headset View (left), External View (middle) and Virtual View (right). A remote participant gives oral instructions to the AR user on how to position 3D shapes on a virtual support.

provide spatial information about both virtual and physical objects in the workspace but also information about the actual AR user, such as her or his body position and gestures.

Scene viewpoint. It determines which virtual and physical objects are visible at a given moment or during the whole collaborative task and from which perspective. Previous literature often makes a distinction between a *first-person* and a *third-person* perspective (e.g., see Komiyama et al. [28]). The former refers to the perspective of the AR user. It can be captured by a head-mounted and communicated to remote collaborators. The latter refers to an out-of-body perspective as captured by external cameras.

View independence. A key problem is how to enable remote users to independently navigate in the 3D space of the AR user to obtain a convenient view, e.g., a view that helps them inspect details of the virtual model or avoids occlusions, and point to a position in space, e.g., to indicate a physical or virtual object to the local user.

Additional dimensions, such as display configuration and means of communication, can emerge from this design space. We chose dimensions that focus on the collaboration process itself rather than ones that deal with how collaboration is made possible, since many tools have already been presented for this purpose [20, 51, 56, 61]. To simplify our user studies, we also decided to focus on one-to-one collaboration. We defer the study of the more general case where multiple remote collaborators participate to our future work.

4 USER STUDY 1

We conducted a user study to investigate our first research question (RQ_1). The study examines trade-offs of different workspace representations and scene viewpoints. In particular, it observes how users provide remote instructions under three configurations:

Headset View is an augmented video from a first-person viewpoint. We capture the video directly from the AR headset to simulate the situation where the remote user sees the scene "through the eyes" of the local user (see Fig. 2-left). The video feed integrates the virtual 3D content into the physical scene of the AR user. The key strength of this configuration is that collaborators share a common frame of reference. So they do not need to mentally rotate the 3D space [47] to communicate.

External View is an augmented video from an external (third-person) viewpoint. We use a depth camera (Microsoft Kinect V2) to provide an overview of the full workspace of the local user. A key question is how to optimally position the camera. In previous studies [18, 51], in which the local worker remains seated, the external camera is positioned at the back left (or right) side of the worker. This way, the two collaborators view the scene from a similar perspective. Unfortunately, in such configurations, the face, hands, and other key parts of the worker's body may not be visible. Furthermore, if the worker freely moves around the model of interest, his or her body may occlude parts of the workspace. For these reasons,

we excluded this alternative. For optimal visibility, the camera is positioned in front of the AR user (at 2m height and approximately 2.5m away) and oriented 30° downwards. We also ensure that the board on which the local user places objects is centered in the recorded image. The video feed of this camera is augmented with the virtual 3D content visible in the AR user's workspace (see Fig. 2-middle). Compared to first-person views, external views have been shown to increase communication efficiency [18] and improve performance and satisfaction [51]. Other authors observe that users strongly prefer them for *"placing objects recommended by themselves"* [11].

Virtual View is a fully virtual representation with a free viewpoint. Remote collaborators see a virtual representation of the 3D scene. A simplified avatar shows the head and hands of the AR user (see Fig. 2-right). Remote users can freely navigate in the 3D scene and choose their preferred viewpoint. This approach follows the naive metaphor of birds that can choose the most convenient position to observe the AR user. Previous results [52] suggest that this additional freedom in the choice of views can improve both the performance and the confidence of remote collaborators.

The study took place during the COVID19 pandemic. To eliminate risks of contamination, the experimenter (first author) acted as the local user wearing the AR headset for all sessions of the study. Participants acted as remote collaborators and completed the study tasks from their home or office environment. The experimental protocols of our studies were approved by a local ethical committee.

4.1 Participants

24 volunteers (11 women and 13 men) participated. They were 21 to 41 years old (Median = 26.5 years). All were frequent or occasional users of at least one video-communication tool, such as Skype or Zoom. Seven participants frequently or occasionally used an AR or a VR headset. 11 participants were frequent or occasional users of 3D games, game engines, or 3D modeling environments. Participants were recruited by word of mouth and responses to a recruitment email sent to our lab's mailing lists. No compensation was given.

4.2 Apparatus

The experimenter set up the workspace in his home environment and interacted with the scene through a Microsoft HoloLens 2. For the calibration, the experimenter defined the HoloLens origin by manually positioning a 3D object on an AprilTag [59] marker. The Kinect camera was automatically calibrated by detecting this marker using the ViSP library [35]. Communication between the participants and the experimenter was established through commercial video-communication software (Skype or Discord). The Headset View and External View were presented to participants through screen sharing. For the Headset View, we used the Microsoft HoloLens 2 video-sharing application [2] to stream live video from the headset. For the External View, our implementation considered potential occlusions between virtual and real objects as "seen" by the Kinect camera. For each pixel, a shader chose to display either the streamed video or the virtual object by respecting their depth information from the camera. For the Virtual View, participants downloaded and executed a client application, which rendered an interactive 3D scene synchronized with the HoloLens application via a remote server. This architecture is implemented with Unity 2019.4 and used the Unet library [6] for network communication. Participants could pan, zoom, and rotate the 3D scene using their mouse and keyboard. Finally, we used a website to guide participants in the course of the experiment (see Fig. 3-a). This website provided information and instructions regarding the configurations and the task and linked to our online questionnaires.

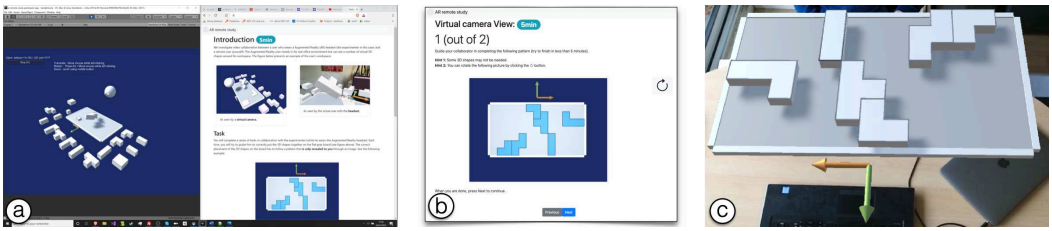

Fig. 3. (a) Remote participant interface used for our first study: tested view configuration on the left (Virtual View in this example) and website used to give instructions on the right. (b) Close-up of the website showing the target pattern: the UI widget on the right allows participants to rotate the pattern image. (c) Zoom-in on the AR user workspace showing the virtual board with the finalized task: colored axes help participants make the correspondence between the pattern on the image and the virtual board shown on the view.

4.3 Task

Participants were asked to place 3D pieces of nine different shapes on a virtual board by giving oral instructions to the experimenter who acted as a surrogate (see Fig. 2). The experimenter used close and distant manipulation tools provided by the Microsoft HoloLens 2: its direct manipulation gestures, its hand-ray tool and air-tapping for selection.

The solution to the task was a 2D top-view pattern that described how to position pieces in any order. The pattern was randomly generated to contain eight pieces out of 18 pieces available in the workspace. It was presented to participants as an image on the website and was unknown to the experimenter (see Fig. 3-b). Its default orientation shown to participants reflected the experimenter's perspective. The pattern was thus inverted with respect to the External View. To help participants adapt the orientation of the 2D pattern as they would do with a piece of paper, we included UI widgets for rotating the pattern. We also added colored axes both in the views and pattern images to make correspondence clear. The virtual board was composed of a 9×5 grid of squares with side length 10 cm (see Fig. 3-c). When a 3D piece was placed on the board, it was snapped to the grid. Pieces had a maximum length of 30 cm (i.e., three grid squares).

As Kuhlen and Brennan [29] discuss, using a confederate in studies that involve conversations between humans is a common research method, but its practice *"might be hazardous"* to collected data. In particular, if confederates have an active, uncontrolled participation in the dialog and are aware of the hypotheses of the study, they can bias the results. To reduce the risk of bias, we established a minimalistic communication protocol for the experimenter. The experimenter followed the participant's instructions and only verbally intervened: (i) to ask the participant to repeat an instruction if the instruction was not understood; (ii) to request confirmation for a planned action; and (iii) to request confirmation for a completed action. The experimenter could also answer questions concerning the user interface or the task, but we tried to respond to such questions as much as possible during training. In contrast, the task required participants to take the initiative as speakers, as Kuhlen and Brennan [29] also recommend.

4.4 Design

To keep sessions short, we simplified the experimental design by dividing the user study into two independent parts. Focusing on the viewpoint (first-person vs. third-person) of AR video, Part I compared the Headset View with the External View. Focusing on the workspace representation (virtual vs. AR) and the type of navigation control (remote user vs. local user control), Part II compared the Headset View with the Virtual View. We divided our participants into two groups of 12 participants, one for each part, trying to balance gender. We followed a within-participant

design, where all 12 participants tested both configurations. Half of them were first exposed to the Headset View, and the other half starts with the second condition. For each configuration, participants completed two main tasks, preceded by a training task with a simplified pattern with three only pieces.

4.5 Procedure

After signing a consent form, participants completed an online demographic questionnaire. Participants went through a short tutorial that explained the two communication configurations. They were then introduced to the training and two main tasks of each configuration. Participants evaluated the configurations and the task through a set of questions divided into multiple short questionnaires. Each participant answered seven questionnaires in total: one after each task (2 tasks × 2 configurations), one after each configuration (2 configurations), and one after the full session. The full procedure lasted approximately 50-70 minutes.

4.6 Data Collection and Measures

We collected: (i) participants' answers to the online questionnaires, (ii) recordings of the participants' voice during the tasks, and (iii) logs of low-level software events (view positions, trajectories, and time stamps). As we discussed above, the presence and collaboration role of the experimenter adds bias in the way tasks are completed. As a result, task performance measures such as task-completion time and errors are not reliable, and we do not consider them here. We focus instead on how participants perceived difficulty for different components of the task. We also report on the participants' preferences and their feedback about trade-offs of the compared conditions. Finally, we examine the strategies that participants followed to complete the tasks. Consider that our analyses are exploratory and should be interpreted as such.

4.7 Results

We present our main results. Anonymized data from this study and the R code of our analyses are available as supplementary material at https://osf.io/g7xas.

Perceived task difficulty. Participants rated the difficulty for each sub-task through 5-point Likert items (1 = very difficult, 5 = very easy). We miss the answers of one participant for these questions in Part I. The analysis of ordinal data with metric models is generally problematic [33]. We therefore use state-of-the-art *cumulative probit* regression models [13, 33] that enable us to map ordinal scales to a latent (i.e., not observable) continuous variable and then express estimates of differences between conditions as standardized effect sizes. For an extensive justification of this method and a comprehensive tutorial, we refer the interested reader to Bürkner and Vuorre [13]. The method is based on a Bayesian statistics [27] framework, but we emphasize that we do not use informative priors here. Figure 4 presents the results of our analysis, where we compare the perceived difficulty of our configurations through estimates of mean standardized differences expressed as 95% *credible intervals*[1]. Those are differences over a continuous (rather than ordinal) physiological variable of difficulty and are expressed in standard deviation (SD) units. In contrast to common non-parametric significance tests that rely on rank transformations, the approach enables us to estimate the magnitude of the observed effects by means of probabilistic interval estimates and effect sizes and thus better evaluate the statistical evidence about these effects.

The results indicate that participants perceived that the External View was easier than the Headset View for searching pieces in their collaborator's environment. In contrast, the External

[1]A credible interval is the Bayesian analog of a confidence interval. Unlike a 95% confidence interval, which is often misinterpreted, a 95% credible interval expresses a range in which the parameter of interest lies with 95% probability [27].

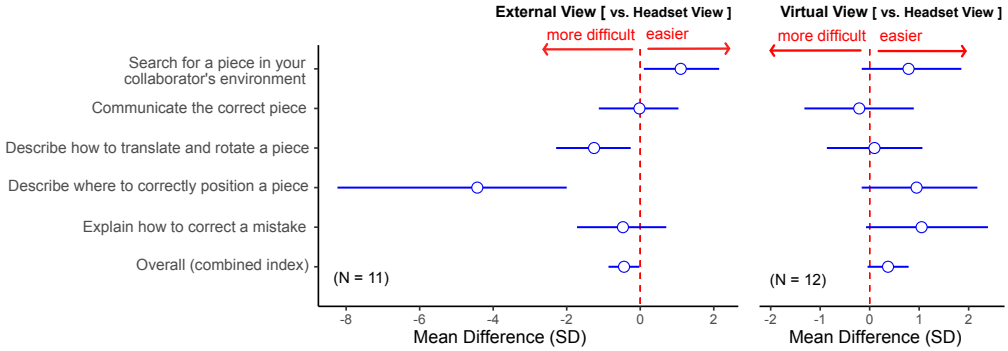

Fig. 4. Comparing the perceived difficulty of different subtasks between configurations. For our analysis, we use Bayesian ordinal (cumulative probit) models [13], which map the original ordinal scale of Likert items to a *latent* continuous variable. The bars in the graph represent 95% *credible intervals* of mean differences over this continuous variable and can be treated as estimates of standardized effect sizes. Note that the unit of these differences is the standard deviation (SD) of the distribution of the latent variable.

View was more difficult for describing how to translate or rotate a piece and how to correctly position a piece. This latter effect is especially pronounced. When exposed to the External View, several participants struggled to correctly map their image of the pattern to the workspace of the AR user. Because of the position of the external camera, the participants had to mentally perform a rotation transformation to give the correct instructions. We further discuss this problem below. For the other subtasks (communicate the correct piece and explain how to correct a mistake), we do not observe any clear difference between the two configurations.

Differences between the Virtual View and the Headset View are more uncertain. There is a trend that the Virtual View was perceived as easier for searching pieces in their collaborator's environment, for describing how to correctly position a piece, and for explaining how to correct a mistake. However, the low size of the sample does not let us draw clear conclusions.

Preferences. We also asked participants to compare the configurations that they tested on six different aspects of the collaboration task. Figure 5 summarizes our results. We observe that participants see different benefits in each configuration. They appreciated the ability of the External View to provide awareness about the remote environment and help them search and locate pieces effectively. However, most participants expressed an overall preference for the Headset View, as it helped them perceive their collaborator's actions, facilitated communication, and helped them complete the task more effectively. The Virtual View, in turn, was especially appreciated for helping participants search and locate pieces effectively but also complete the task more effectively than the Headset View. Overall preferences between the Virtual View and the Headset View were equally split.

Trade-offs. Open-ended questions in the questionnaires asked participants to elaborate on the strengths and weakness of each configuration. All 12 participants of Part I reported that providing a global view of the workspace was the main strength of the External View. *"The strongest aspect was being able to see the overview of the scene and the entire puzzle we are building as a whole"* (P1).

"The fixed camera implies that all the items always stay in view of the distance person, easier if the collaborator cooperates less" (P11). As a comparison, in the Headset View *"the environment is reduced, and it takes more time to find your way around and locate all the items"* (P3). *"I do not have an autonomy of my vision angle, I only see what he sees" (P5).*

However, most participants evaluated this very same property of the Headset View as its strongest aspect: *"giving directions is much easier because I can just tell the partner to what I am*

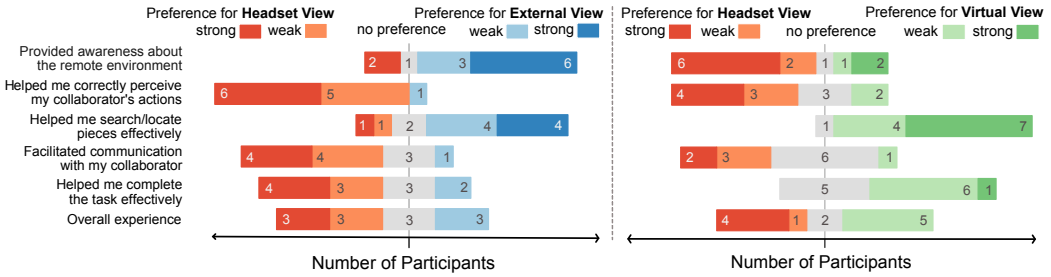

Fig. 5. Distribution of participants' preferences: Headset View vs. External View (left) and Headset View vs. Virtual View (right).

doing!" (P5). According to P12, *"you see through the eyes of [your partner], so you could exactly guide his gestures like a puppet."* In contrast, eight participants explicitly mentioned the inversion of left and right as a major problem of the External View: *"You are located on the opposite side so everything is going to be the reverse to explain."* (P12). Even though we allowed users to rotate the reference image with the solution pattern (see Fig. 3-b), only half of them used this function, and even this strategy did not seem to solve the problem for them.

As additional limitations of the External View, participants complained about distance distortions (P10), difficulties in correctly perceiving depth (P1), a sense of *"distantiation"* (P12), and a weaker sense of participation (P3).

The responses of the participants of Part II focused on the same qualities and drawbacks of the Headset View but raised additional concerns that the camera can be *"shaky"* (P19) and can *"induce motion sickness"* (P13). Concerning the Virtual View, participants especially appreciated its navigation capabilities: *"The user may navigate independently of the operator, make it possible to change point of view, or see things out of the operator's sight"* (P13); *"you are totally autonomous on the vision of the environment"* (P21). However, participants also identified several weaknesses: *"I do not really know where my collaborator is looking at"* (P15); *"lack of information about the real environment of the other user"* (P18); *"less points of reference than the previous configuration"* (P22); *"users need to be used to 3D applications in order to place [their] view correctly"* (P20).

Communication strategies. All participants frequently referred to their partner's "left" and "right" to communicate orientation. A common approach for indicating specific objects was to verbally describe their shape, e.g., by means of a letter of a similar shape ("Z", short "L", long "L", etc). A small number of participants (four in total) responded that they sometimes or frequently made use of physical objects in the experimenter's space as reference for the two AR views. To provide directions about how to rotate objects, strategies were more diverse. Several participants described the angle (90 or 180 degrees) of the rotation and its direction (clockwise/anticlockwise or left/right), while two participants acknowledged difficulties in finding an efficient strategy. For translations, most participants used the edges and corners of the virtual table for reference, but for higher precision, they also referred to the borders of other pieces on the table. In the Virtual View, participants' dominant approach was to place the virtual camera above the head of the avatar of their partner to obtain a similar viewpoint. According to our logs, four participants moved around the board to discover a better viewpoint but also ended up placing the camera at this position.

User feedback. Two participants proposed to place the camera of the External View slightly behind (P5) or above the head (P12) of the AR user, while P2 proposed to approach the camera closer to the table. P14 and P21, instead, wondered about the possibility to increase the field of view of the Headset View, e.g., by adding extra cameras, while three participants (P12, P13, P15) proposed

to combine multiple views together. Finally, several participants made suggestions about pointing techniques: a cursor for *"indicating locations"* (P17), a laser to *"target specific parts"* (P18), clicking with the mouse to *"illuminate a piece"* (P22) or to *"ping"* at a certain position as the Headset View moves (P16), and *"add a vocabulary to easier describe pieces"* (P9).

4.8 Discussion

The task required the AR user to manipulate virtual only objects. This choice was made to ensure that participants could complete the task under all three configurations. Clearly, it overrates the utility of the Virtual View, which lacks support for physical objects. Furthermore, we notice that some participants expressed strong preference for the Headset View over the External View. The External View was also rated as more difficult for certain subtasks. This finding is somehow at odds with results of past studies [18, 51], suggesting that the specificities of the task and the camera viewpoint may have an important influence to the success of a representation. In particular, in the external views that those two studies compared, the camera was conveniently located to the back left of the worker. As Shepard and Metzler [47] have shown, the time needed to perform a *mental rotation* in 3D space linearly increases with the angular offset of a viewer's viewpoint. This mental-rotation model predicts longer reaction times for our 180° camera configuration and implies a greater mental effort. An 180° offset also requires collaborators to reverse their wording, e.g., to replace every egocentric "right" with a "left" [46].[2]

Despite the above shortcomings, the External View presents several benefits over the Headset View. First, the view provided global awareness about the remote environment. Second, most participants felt that it helped them search for and locate pieces with less effort (see Fig. 4-5). The External View is also the only configuration that allows remote users to see the face and real full body of their collaborators. Although the role of such information was not directly evaluated with our task, it can be essential for supporting empathy [53] between participants and establishing communication awareness [14].

5 *AR*GUS: A MULTI-VIEW COLLABORATION SYSTEM

The results of our first study show that each view configuration has unique qualities that are difficult to substitute by the other two. The External View supports global awareness about the physical environment of the local worker and helps the remote user search for objects that are spread around the workspace. The Virtual View supports independent navigation, helping the remote user to provide instructions (e.g., about how to correct mistakes) from a convenient but also stable point of view. Finally, the Headset View is especially effective for perceiving the actions of the AR user and communicating egocentric instructions. Our research efforts thus focus on how to combine them and how to give remote desktop collaborators direct control over their use. To this end, we developed ARgus, a multiview collaboration system for 3D modeling (see Fig. 1). ARgus's implementation reflects three design goals:

DG$_1$. Communicate both real and virtual representations but without requiring the 3D reconstruction of the local workspace. We rely instead on video for capturing the physical environment of the local AR user and his or her real body. As we discussed in previous sections, this approach avoids problems associated with the 3D reconstruction of a physical workspace.

DG$_2$. Support both first-person and third-person views of varying levels of view independence. This goal is consistent with the results of our formative study and recommendations of

[2] A mirror configuration would transfer the problem to rotational directions, e.g., a "clockwise" direction should become "anticlockwise." Given their complexity, we suspect that the mental effort of such transformations would be even greater.

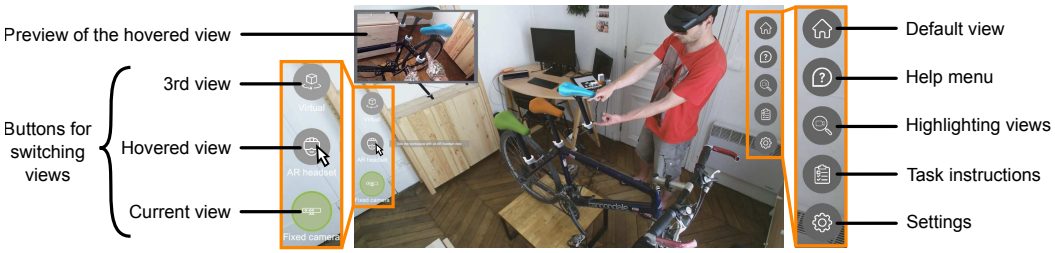

Fig. 6. Desktop interface of ARgus used by a remote collaborator for the redesign of a bicycle saddle.

several older studies [20, 26, 40, 43, 52]. A challenge for ARgus was how to design effective and consistent mechanisms for switching and navigating between and within views.

DG$_3$. Provide tools that minimize communication effort and facilitate coordination. According to Schober [46], speakers try to minimize the mental effort of their addresses and their own by replacing *speaker-centered descriptions* (e.g., at "my left" or "your right") by *neutral descriptions*. ARgus provides aids for neutral descriptions via direct-pointing and spatial-annotation tools.

Below, we present the main features of ARgus. Although the system supports bidirectional communication, we focus in this paper on its design for remote desktop collaborators.

5.1 Combining Multiple Views

ARgus receives the augmented video streams from both an AR headset and an external depth camera located in the AR user's physical space. Furthermore, it maintains a synchronized version of the virtual 3D scene and can generate virtual views from any workspace location. Remote users can seamlessly switch between virtual and augmented video representations, as well as freely navigate to any viewpoint on the 3D scene.

ARgus also offers the possibility to display live previews of all three views (Headset View, External View, and Virtual View). These previews are video thumbnails of alternative views displayed in a small embedded window on top of the user's current view. They allow users to take a quick look at a different view, e.g., to inspect details of the physical environment that are not visible in the current view or to decide whether it is worth switching views. This mechanism aims to prevent the short bursts of switching between views observed by Gaver et al. [20] and facilitates coordination when users ask their collaborator to temporarily switch to their viewpoint to approve the veracity of their discovery [40].

5.2 Supporting Navigation

We provide several solutions for displaying previews, switching between views, and navigating in the 3D scene.

Main user interface. The main window of ARgus' user interface displays three circular buttons for selecting views and getting feedback about the active view (see Fig. 6). When users hover over a button, a live video preview is displayed on the top-left corner of the window. Clicking on the button activates the view. We use a trajectory and field-of-view interpolation based on Cinemachine [1] to animate the virtual camera in the 3D scene.

This solution ensures visual consistency among views, helps users understand the location of distant viewpoints, and avoids disorientation. We also use a blur effect to smooth out transitions between augmented video and virtual representations. We let users customize the duration of view transitions.

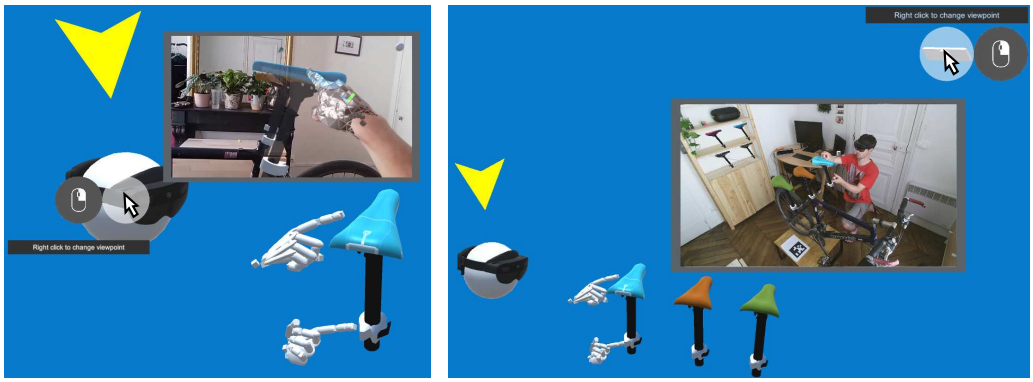

Fig. 7. The remote user hovers the mouse over the headset of the 3D avatar (left) and Kinect 3D model (right) to display the preview of the Headset View and the External View respectively.

Interacting with the 3D scene. The 3D scene of ARgus' Virtual View serves as the basis for 3D navigation. It also offers an alternative solution for switching between views through interactive virtual camera representations. In the Virtual View, users can use the mouse to rotate their viewpoint around the center of the 3D scene and translate it (pressing ALT). The same navigation capabilities are available in the two augmented-video representations, the Headset View and the External View. However, since remote users do not have direct control of the position of the two physical cameras (i.e., the external and the head-mounted camera), navigation actions within these views immediately cause the view representation to turn to virtual. This design approach ensures that interaction is consistent across all views.

The 3D scene includes virtual representations of the physical cameras themselves. Users can interact with them to preview or activate their corresponding views. For example, Fig. 7 shows the active Virtual View of a desktop user who remotely collaborates for the redesign of a bicycle saddle. The virtual view does provides no information about the real scene. Therefore, the remote user hovers the mouse over the headset of the 3D avatar to better understand what her partner sees (Fig. 7-left). She then hovers over the model of the Kinect camera (Fig. 7-right) to compare how the three saddle designs look together with the bicycle's real frame. Users may also decide to click the mouse to switch to this view. Finally, the 3D scene includes guides (arrows and highlighting effects) that help users locate the cameras and orient themselves in the 3D space.

Navigating with spherical views. Using base 3D rotation and translation interactions to closely inspect specific parts of a 3D model can be tedious and time consuming. To facilitate such tasks, we adapt *Navidget* interaction technique [22] and integrate it into ARgus' user interface as a Spherical View tool. Activated with a mouse right-click within either the External View or the Virtual View, the tool visualizes a sphere centered on the selected point. Users can move a virtual camera on the surface of the sphere, and a camera preview is shown (see Fig. 8-left). The sphere radius can be adjusted with the mouse wheel, causing the virtual camera to zoom in or out. Users can release the mouse to switch to a desired view or press ESC to keep the current viewpoint.

Viewpoint recording. Following the approach of Sukan et al. [50], we allow users to record viewpoint locations (pressing a key) when they spot interesting views that they want to later reuse. Viewpoint recordings are represented as virtual cameras. As all other cameras (see above), they have a visual representation in the 3D scene, and users can interact with them to preview or switch to their views.

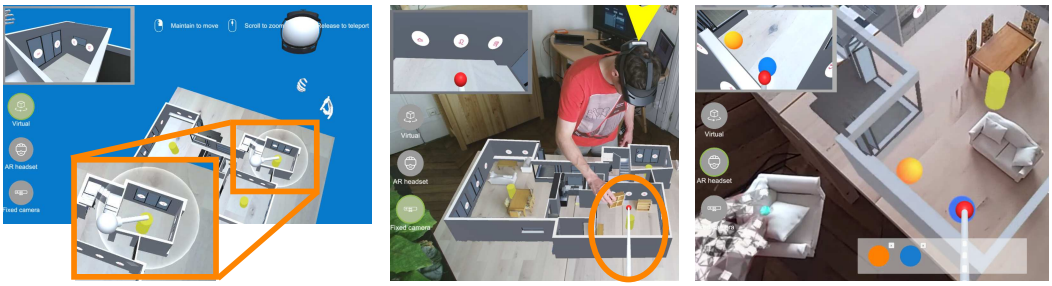

Fig. 8. Tools available in ARgus: Spherical View (left), Virtual Stick (middle) and annotations (right).

5.3 Facilitating Communication

Other tools in ARgus focus on how to facilitate the communication of users (DG_3).

AR user representation. The Virtual View includes a synchronized representation of the AR user with a simplified avatar composed of a sphere wearing the 3D model of a Microsoft HoloLens 2 and virtual hands (see Fig. 7). Each hand is represented by 24 joints, connected by canonical shapes, such as cylinders and squares. Both hands and head positions are retrieved from the MRTK libraries [4]. In the External View, a vertical arrow on top of the real head of the AR user communicates an interaction point for previewing and selecting the Headset View.

Pointing stick. As several participants of our formative study proposed, it is often useful to directly point in the remote scene, e.g., to indicate an object or provide instructions about where to place it. ARgus provides such functionality through a Virtual Stick (see Fig. 8-middle). The stick starts from the viewpoint's origin. Its direction is controlled with the mouse, while its length can be adjusted with the mouse wheel. A small sphere represents its tip, which is red if colliding with a 3D element and grey otherwise. A dotted line indicates its pointing direction, starting from its tip and projected until its collision with a 3D model in the scene. We considered results by Brown et al. [11], who report that users express a strong preference for surface-constrained pointing under all circumstances. A virtual camera is attached to the tip of the stick, and a preview of this camera is displayed on the top-left corner of the main window, helping users perceive depth and understand where the Virtual Stick is pointing at. In the Headset View, the view is frozen from the time users activate the Virtual Stick until they stop using it. Like in TransceiVR [58], freezing the moving view allows users to focus on an interesting viewpoint and achieve more accurate pointing.

Annotations. Overlaying information in an AR workspace in a spatially meaningful way can improve human performance and decrease mental workload [54]. Likewise, using shared virtual landmark increase user experience and facilitate spatial referencing in collaboration [37, 46]. In all views of ARgus, remote users can use the Virtual Stick to add annotations represented as colored spheres. In Fig. 8-right, for example, the remote user has added a yellow and a blue annotation to suggest target locations for placing furniture. The user interface shows a list of all activate annotations (up to five in our evaluation study), allowing users to quickly review and remove them.

5.4 Architecture and Implementation

ARgus was developed in Unity 2019.4. Its architecture relies on a client-server model connecting a remote desktop user and a local AR headset to a local server (see Fig. 9). The server keeps a synchronized version of the 3D scene and records the AR user's physical workspace with the external depth camera. It generates the External View by augmenting the camera video feed with the objects of the 3D scene. Occlusions between the virtual objects and the real objects are

[3]The figure includes icons made by Freepik and Good Ware from www.flaticon.com.

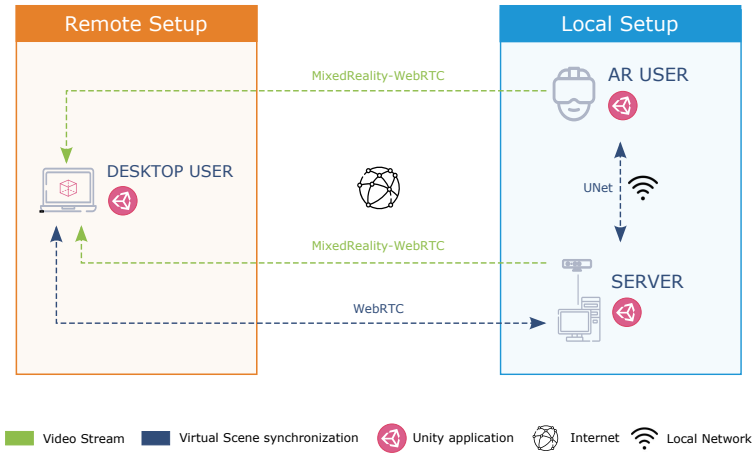

Fig. 9. System's architecture and implementation.[3]

managed through the depth map of the external camera: for each pixel, a shader displays either the streamed video or the virtual object according to their depth information.

The AR headset is connected to the server as a client using the Unet library. It maintains a synchronized version of the 3D scene, which is used both to render AR user's augmented view in the headset and to generate the video feed of the Headset View. It also transmits the AR user's head and hands positions and orientations. To calibrate the AR headset reference frame and the depth camera reference frame, the virtual space origin is defined (i) manually by the AR user who needs to position a 3D object on an AprilTag [59] marker and (ii) automatically by the depth camera, which detects this marker using the ViSP library [35].

The application of the remote user is also connected to the server as a client using WebRTC. We built a custom protocol based on this technology to synchronize 3D object states (position and rotation) and software events (tools, logs, etc.). The application can thus render a synchronized version of the 3D scene to create the Virtual View. In addition, it receives the video feeds from the AR headset and the server based on the Mixed-Reality WebRTC libraries [3] to display the External View and the Headset View.

6 USER STUDY 2

We conducted a second user study that investigates our second research question (RQ_2). The study examines how remote collaborators use ARgus to provide instructions to a local AR designer.

As for our first study, we opted for an experimental design that avoids contamination risks due to the COVID19 pandemic. The experimenter (first author) acted as the local user wearing the AR headset, while participants acted as remote collaborators and completed the tasks from their home or office. A preregistration [16] of the study is available at https://osf.io/6dhzn.

6.1 Participants

12 volunteers (4 women and 8 men) participated in the study with an age ranging from 24 to 29 years old (Median = 27.5 years). All were frequent or occasional users of at least one video-communication application. Two participants frequently or occasionally used an AR or a VR headset, while five participants had no previous experience with AR/VR technologies. Eight participants were frequent or occasional users of 3D games, game engines, or 3D modeling environments. Before starting

the tasks, we verified that all participants had a stable internet connection (we replaced four initial participants who could not continue due to connection problems). We followed the same recruitment process as for our first study.

6.2 Apparatus and Conditions

As for the first study, the experimenter interacted with a Microsoft HoloLens 2 in a workspace created in his home environment. We evaluated a simplified version of the ARgus (written here as ARgus) to help participants quickly master the key features of the interface. More specifically, we deactivated its support for viewpoint recording since it was not useful in our experimental task. We also used pre-selected positions for the spherical view, suitable for the 3D model used in this study. To activate the tool, participants had to right-click on a yellow cylinder located at three relevant positions of the model (one for each room of a house model). The cylinder then became the rotation center of the Spherical View. As we observed in our first study, finding a good placement for the external depth camera is not trivial and largely depends on the task. We decided to use the same configuration as for the first study: we positioned the camera at 2m height, 30° downwards to face the experimenter and to capture his moving body and his augmented workspace, minimizing occlusions. We used the Headset View as control condition. As in our first study, this condition did not provide any interaction capabilities.

Participants downloaded and executed a single Unity application for both conditions on their personal computer. The user interface had a fixed-size window with a 1920 × 1080 resolution. A step-by-step tutorial about the system functionality and the tasks was directly embedded in the system. For verbal communication between the participants and the experimenter, we used a commercial application (Skype or Discord).

6.3 Task

ARgus' functionalities can support remote mixed-reality participatory design in a range of domains, such as furniture arrangement [11] and urban planning [15, 44]. We decided to focus on a furniture arrangement task because it was used in the past by other related studies [20, 26, 45]. As in our formative study, this task requires participants to search for 3D pieces in the workspace of the experimenter, find a target location for them, and instruct the experimenter to place them correctly. In contrast, we now looked for tasks that would involve both physical and virtual objects in a scene. We considered two alternatives: (i) the AR user manipulates virtual pieces within a larger physical frame of reference (e.g., as in Fig. 6); or (ii) the AR user manipulates physical pieces (miniature furniture) within the virtual model of a house. We opted for the second alternative (see Fig. 1), as it provides richer opportunities for virtual navigation and better captures the trade-offs of different representations. The task simulates the situation where a remote buyer communicates with a furniture designer (or seller). The furniture designer follows instructions to try miniature models of his or her collection in a virtual model of the buyer's house.

We introduced several constraints to create various arrangement tasks unknown to the experimenter. Zodiac symbols were randomly displayed on pre-defined positions on the virtual house model's walls. We chose these symbols as they are easy to identify but hard to verbally describe. This way, we forced participants to rely on intrinsic landmarks of the model for communicating positions, rather than artifacts that are absent in real-world tasks. Two symbols were randomly assigned to each participant. In each room, these two symbols were located on perpendicular walls and defined a cross-shaped forbidden area: the line in front of each symbol was not available to place furniture.

Participants were asked to arrange furniture for three *thematic spots* randomly chosen among nine. The functional aspect of these spots was described textually. For example, a "living spot" was

described as "a place where people can meet and spend some time together". To perform this task, participants could choose miniature furniture among six storage cabinets, four tables and ten chairs (see Fig. 1-d). To complicate the task, we required each miniature chair to be appropriately oriented so that sitting people can see a virtual window without moving their head too much. To be valid, a spot had to include at least two pieces of furniture, meet the placement constraints and represent an harmonious layout (according to the participant's preferences). The symbols, constraints and spot description were communicated to participants at the beginning of each task and made available at any time in a specific panel of the interface (see Fig. 6). This information was unknown to the AR local user.

As for Study 1, we tried to reduce the experimenter's influence [29] by constraining his verbal interventions to only ones required for the completion of the task, i.e., asking the participant to repeat instructions, and asking for confirmation of planned or completed actions.

6.4 Design and Procedure

We followed a within-participant design, where all 12 participants tested both user interface configurations. Half of them were first exposed to the Headset View. The other half were first exposed to ARgus. After signing a consent form, participants completed an online demographic questionnaire. They were then introduced to the two configurations. For ARgus, participants went through a tutorial presenting each tool step-by-step. For each configuration, participants completed a practice and main task. The practice task required the arrangement of one thematic spot.

At the end, participants completed a questionnaire that evaluated their experience with the two configurations that they tested. The full procedure lasted approximately 70 to 90 minutes.

6.5 Data Collection and Measures

We collected participants' answers to a pre- and a post-questionnaire. The post-questionnaire evaluated the efficiency of each user interface configuration with a Likert scale of four items with seven levels (*1 = Inefficient, 7 = Efficient*). It also assessed the importance of verbal communication for each configuration with a Likert scale of four items with five levels (*1 = Not important, 5 = Very important*). The questionnaire further evaluated the utility of the views and interactive tools of ARgus configuration and collected participants feedback about their use. We also collected logs of low-level events that describe the use of interactive tools and view transitions during the task. Due to technical problems, logs were not collected for one participant (P5).

Finally, we recorded and manually transcribed participants' voice during the tasks. We then distinguished between phrases that provide remote instructions and other non-instructional content, such as transitional ("ok", "now") and thinking-aloud sentences. Instructions were further classified into three subtask categories: identifying & reaching an object, manipulating an object, and moving in the scene. These categories cover the full set of instructions that we identified and do not overlap. We started with a finer-grained coding scheme. In particular, we initially tried to differentiate between instructions on identifying and reaching objects or locations, and between instructions that concerned different types of manipulation actions. However, these categories were often fused, which made their coding uncertain and unreliable. We thus finally opted for larger categories.

The first and second author decided together on how to segment the transcripts and code the segments by inspecting the data of the first participant. They independently coded the transcripts of three additional participants. They then discussed and finalized the segmentation and coding scheme. As a last step, the first author re-coded all the transcripts, while the second author independently coded the transcripts of the last two participants. We calculated inter-coder reliability at the word level both for distinguishing between instructions and non-instructions (Krippendorff's $\alpha = .98$, 95% CI [.97, .99]) and for the overall classification that also considers the type of instruction

(Krippendorff's α = .97, 95% CI [.96, .98]). Inter-coder reliability scores are high, so we count and analyze the words in participants' transcripts for all the above categories.

6.6 Questions and Hypotheses

We expected that participants might develop diverse strategies to complete the tasks. Our goal was to observe and understand these strategies. We were particularly interested in two questions:

Q_1: Will participants find the three views of ARgus useful, and how will they make use of them?

Q_2: Will participants find the user interface tools useful, and how will they make use of them?

Furthermore, we wanted the participants to reflect about how they completed tasks with the two configurations and report on their trade-offs. Tait and Billinghurst [52] found that increased view independence reduces the number of verbal instructions between collaborators. Likewise, we expected that ARgus would reduce reliance on verbal communication, because it gives more viewing freedom to remote users and provides opportunities for completing the task more efficiently. More formally, we tested the following three hypotheses:

H_1: The mean perceived efficiency will be higher for ARgus.

H_2: The mean perceived importance of verbal communication will be lower for ARgus.

H_3: The mean number of words for communicating instructions will be lower for ARgus.

Like Tait and Billinghurst [52], we are interested in the link between view independence and communication performance. However, our studies are distinct from each other. First, since we do not reconstruct the model of the real scene, we investigate view independence through complementary views with different levels of navigation control. Therefore, we also try to identify the view-control strategies that participants develop to carry out the task. Second, our system includes an external view, which also shows the real body of the local user. Note that Tait and Billinghurst [52] recognize the potential benefits of an external view and identify it as a promising configuration for future studies. Third, Tait and Billinghurst [52] test the positioning of physical objects on a physical table. We study instead a more complex task that requires collaborators to position physical pieces within a larger virtual model. In our case, collaborators need to deal with occlusions in the AR scene, thus both physical and virtual navigation are essential for completing the task. Finally, annotations in their system are virtual replicas of a small collection of physical objects, which are conveniently placed on the surface of a table. Our annotation mechanism is simpler but more generic, as it lets remote participants mark any virtual or physical object and location in the 3D workspace with little manipulation effort.

6.7 Results

Anonymized data from this study and the R code of our analyses are available as supplementary material at https://osf.io/3nqrg. Here, we summarize our results.

Use of tools and view representations. We first summarize the strategies that participants used to complete the task under the ARgus condition. For each participant (except for P5), Figure 10 visualizes the active views during the task and the use of previews, the pointing stick, and the spherical view. We emphasize that we did not encourage participants to be fast, and the time range that we show does not always reflect active collaboration time. Some participants (e.g., P1) spent initial time to think about the constraints of the task and further explore the available tools. It is not a surprise that the slowest participants in Figure 10 were exposed to ARgus first (in circle).

Overall, all participants frequently transitioned between views during the task, which demonstrates the utility of our approach. However, we observe that the Virtual View and the Headset View dominated the participants' choices. The External View was heavily used by P1 and P3 and

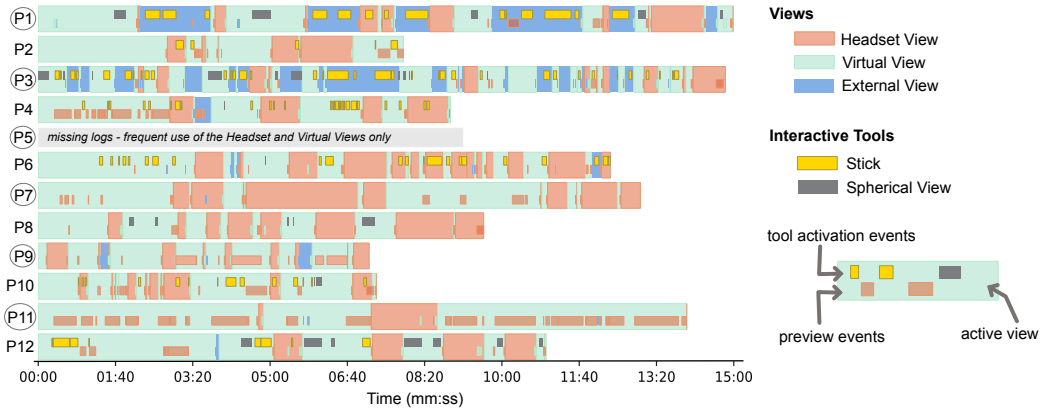

Fig. 10. Use of the three view representations, the pointing stick, and the spherical view by the participants of the evaluation study for the main task under ARgus. Circled participants were exposed to ARgus first.

sparingly by three other participants. Participants' questionnaire responses are consistent with these patterns.

Three only participants found the External View to be useful (P3) or very useful (P1, P9).

P2 explained that he did not *"feel the need"* to use it but *"in a bigger environment it could have been useful to guide the partner quickly from one point to another."*

The three view representation were used in two different ways: (i) as main active views or (ii) through the preview window. Figure 10 shows that several participants (P2, P4, P7, P9, P10, and P11) extensively used the Headset View in preview mode from the Virtual View. According to P4, *"the headset view caused dizziness [...] I stayed in the virtual view and watched the headset view from the window."* P2 agrees that *"having the headset view showing in the corner while navigating and pointing in virtual view was the ideal setup."*

The stick was activated in all three representations either as a pointing or as an annotation tool. For example, P1 and P3 regularly used it from the External View to indicate furniture pieces. P4, P6, and P12 used in combination with the Virtual View to indicate target positions. Other participants did not feel the need to use it: *"I did not use the stick as the rooms had enough identifiable elements to allow my partner to understand my instructions"* (P5). Finally, a smaller group of participants made use of the spherical view. According to P1, it is *"the best to manage the constraints"* but other participants did not agree: *"I was comfortable enough with virtual navigation not to feel the need to resort to the spherical view"* (P2); *"I tried to use the spherical view but I am not enough comfortable with in comparison with rotate and translate so I abandoned."* (P6); *"I would have liked a 2D mapping"* (P5). The spherical mapping that we used is generic but may not be the most appropriate for the specific task. Alternative mappings that better adapt to the geometry of the virtual model might indeed improve the usability of the tool.

Perceived efficiency. We compare the efficiency of the two user interface configurations as perceived by our participants. We use again Bayesian cumulative probit models [13] for our analysis (see Section 4). Figure 11-left summarizes our results. Overall, participants rated ARgus as more efficient (see Hypothesis H_1). This was especially the case for verifying the constraints in the scene. For this task, free navigation through the virtual view seemed to be crucial. According to P6, the Headset View causes *"seasickness"*, while P9 commented that its resolution *"was not so effective to perceive accurately the symbols on the walls when having a wide point of view."* In contrast, seven participants rated the Headset View as more efficient for helping them to perceive

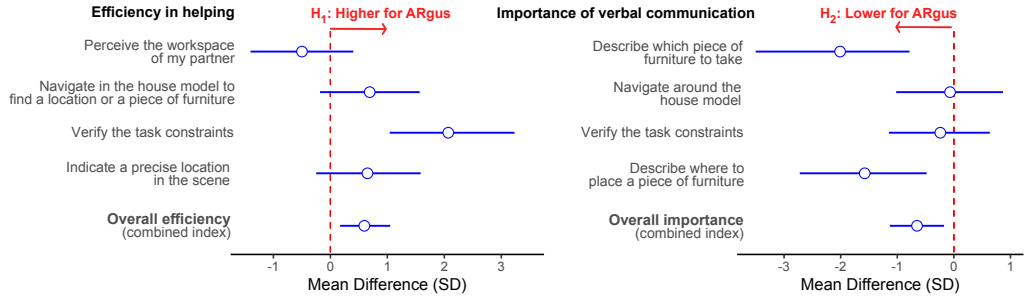

Fig. 11. Comparing the perceived efficiency of the two user interface configurations and the importance of verbal communication for each of them ($N = 12$). We use again Bayesian ordinal (cumulative probit) models [13]. The bars in the graph represent 95% credible intervals of mean differences over a latent continuous variable and can be treated as estimates of standardized effect sizes.

the workspace of their partner despite the fact that the ARgus configuration provided a richer set of views and options for observing the remote space. The added complexity of this interface can explain this result: *"Having only one solution forces to rely on it and in the case of the headset, forces to establish an efficient communication with the partner, that can be lacking when overwhelmed by all the possibilities of the different views and the difficulty to master them all"* (P3).

Reliance on verbal instructions. Figure 11-right compares the mean difference between configurations in participants' perception about the importance of verbal communication. Overall, verbal communication was perceived as less important for ARgus (see Hypothesis H_2), particularly for describing which pieces of furniture to take and where to place them. P10 explained that verbal communication is more important for the Headset View *"because you cannot point with as much precision as with the stick and you cannot see equally well symbols and distances."*

Our transcript analysis provides additional information about how participants verbally communicated instructions. Figure 12 summarizes our results. Overall, the ARgus user interface reduced the number of words that belonged to instructions by 151.8, 95% CI [25.7, 278.0], $t(11) = 2.65$, $p = .023$ (see Hypothesis H_3). To put this number in perspective, participants pronounced on average 834.5 words with the Headset View, where 435.6 of these words were instructions. We observe that clear differences between conditions only concern instructions that ask the experimenter to move around the model. Surprisingly, there is no clear difference in the number of words used by participants to guide the experimenter on how to identify, reach, and manipulate (e.g., translate or rotate) objects. A possible explanation of this result is the fact that five participants did not at all use the stick (see Fig. 10) and relied on verbal instructions for these subtasks. Indeed, a post hoc analysis shows a strong correlation between the use of the stick (binary variable) and the difference of words used for these subtasks (*Point-biserial correlation* = .79, 95% CI [.40, .94]). Seven participants who used the stick pronounced 156.9 fewer words (95% CI [41.8, 271.9]) with ARgus when they provided instructions for these subtasks. This result, however, must be treated with caution because uncontrolled ordering effects may exaggerate the difference.

7 DISCUSSION

Overall, our results confirm that remote desktop collaborators can benefit from the multiple views of ARgus, since each view is best adapted to a different aspect of the task. The Virtual View makes navigation in the virtual model easier and independent of the position and visual focus of the local AR user. The External View provides a static overview of the workspace, showing both virtual and physical objects. Finally, the Headset View allows remote users to directly observe the view and

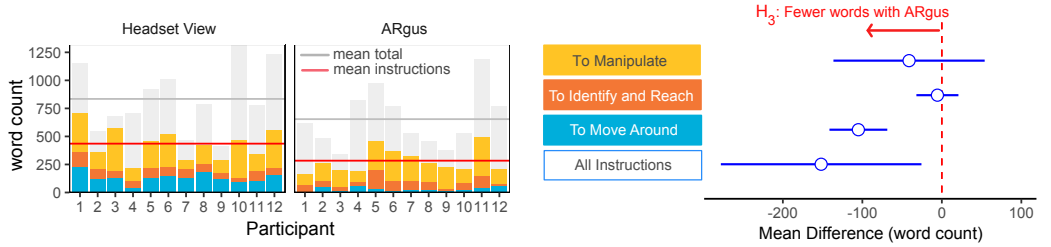

Fig. 12. Results of transcript analysis. We compare the number of words pronounced by the 12 participants to provide instructions. The grey boxes at the left show the total number of words with non-instructions. The error bars at the right represent 95% confidence intervals derived from the t-distribution.

actions of their local partner and provide direct instructions. Our participants demonstrated various strategies on how to combine these views with the tools of ARgus. Given previous results [18, 51], we expected a more extensive use of the External View. However, using all three views can be complex, increasing cognitive costs. So many participants judged that the Virtual View and the Headset View were enough for completing the task. Nevertheless, mastering all combinations of views and previews, as well as developing strategies to use them effectively in various steps of the collaboration, may require a long learning process that we did not assess in our studies. Finding a good viewpoint for an external camera also remains a problem. A solution may be to reposition the external camera on the fly depending on the collaborative situation, as explored in Giusti et al. [21]. The nature of the task may also explain why most participants largely relied on the Virtual View to complete the task. It is reasonable to expect that if key objects and landmarks in the scene were mostly physical rather than virtual, the Virtual View might be less appropriate, while the two other views might be more frequently used. Clearly, there are trade-offs in the choice of each view that largely depend on where the task falls in the continuum between virtual and physical.

The results support our three hypotheses. Participants perceived on average that ARgus was more efficient than the control Headset View condition (H_1) and lessened the importance of verbal communication (H_2). We also found that ARgus reduced the average number of words of remote instructions (H_3), which corroborates previous evidence [52] that increased view independence reduces the prevalence of verbal instructions.

We acknowledge that our experimental method and setup present several limitations. The experimenter took the role of the local collaborator in all experimental sessions, which inevitably limits the external validity of our results. The variable quality of the internet connection and the limited resolution of the HoloLens frontal camera may have had an effect as well. Furthermore, we studied one only part of the bilateral collaboration, neglecting how the local AR user perceives and interprets instructions given by the remote collaborator through multiple complementary views. Future user studies should thus examine the collaboration strategies (verbal communication, physical navigation, and gestural interaction) of local users, and their need for awareness of remote user actions.

Another interesting problem is how to extend ARgus to support multiple remote users and enable them to collaboratively interact with but also edit a shared AR scene. This problem poses significant challenges for the user interface of both local and remote users, since users will now have to coordinate and follow an increased number of viewpoints. Finally, we are interested in enriching ARgus' pointing, annotation, and hybrid navigation tools and evaluate their collaboration effectiveness with more specialized experimental tasks.

8 CONCLUSION

We studied how different views can help a remote desktop user to collaborate with a local user wearing an AR headset on design tasks that may require manipulation of virtual and physical objects. We presented a user study that compared three view representations: (i) a Headset View, augmented video from a first-person viewpoint, (ii) an External View, augmented video from an external third-person viewpoint, and (iii) a Virtual View, a virtual representation with a free viewpoint. Structured as two independent sub-studies with 12 participants each, the study confirmed that each view presents different benefits, targeting a different aspect of a collaboration task. Based on these insights, we developed ARgus, a multi-view collaboration system that provides tools for effectively switching between views, virtually navigating in the remote AR workspace, pointing, and annotating the model. We then ran a second user study to evaluate how 12 remote participants used ARgus to provide instructions to a local user wearing an AR headset for a furniture arrangement task. We observed that participants frequently switched between views or concurrently used them through ARgus' preview functionality. Our results also suggest that the added flexibility of ARgus' multi-view interface allows remote users to verify spatial constraints more efficiently and reduces their reliance on verbal instructions. Future work needs to understand the role of such as a multi-view system from the perspective of the local AR user and extend its scope to multiple remote users.

ACKNOWLEDGMENTS

This work was supported by French government funding managed by the National Research Agency under the Investments for the Future program (PIA) grant ANR-21- ESRE-0030 (CONTINUUM). We thank Olivier Gladin for his precious help during the development of ARgus.

REFERENCES

[1] 2021. *Cinemachine.* Suite of modules for operating the Unity camera https://unity.com/fr/unity/features/editor/art-and-design/cinemachine. Retrieved July 28th 2021.

[2] 2021. *Microsoft HoloLens Application.* Microsoft Hololens 2 Desktop Application https://www.microsoft.com/en-us/p/microsoft-hololens/9nblggh4qwnx. Retrieved July 26th 2021.

[3] 2021. *MixedReality-WebRTC.* Microsoft Mixed Reality WebRTC libraries. https://github.com/microsoft/MixedReality-WebRTC. Retrieved July 26th 2021.

[4] 2021. *MRTK.* Mixed Reality Toolkit for Unity https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/?view=mrtkunity-2021-05. Retrieved July 28th 2021.

[5] 2021. *UN/DESA Policy Brief 92: Leveraging digital technologies for social inclusion.* United Nation, Department of Economic and Social Affairs, Economic Analysis. https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-92-leveraging-digital-technologies-for-social-inclusion/. Retrieved November 26th 2021.

[6] 2021. *UNet.* Unity Multiplayer and Networking. https://docs.unity3d.com/Manual/UNet.html. Retrieved January 22th 2021.

[7] Matt Adcock, Stuart Anderson, and Bruce Thomas. 2013. RemoteFusion: Real Time Depth Camera Fusion for Remote Collaboration on Physical Tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Hong Kong, Hong Kong) *(VRCAI '13).* Association for Computing Machinery, New York, NY, USA, 235–242. https://doi.org/10.1145/2534329.2534331

[8] Tooba Ahsen, Zi Yi Lim, Aaron L. Gardony, Holly A. Taylor, Jan P de Ruiter, and Fahad Dogar. 2021. The Effects of Network Outages on User Experience in Augmented Reality Based Remote Collaboration - An Empirical Study. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 313 (oct 2021), 27 pages. https://doi.org/10.1145/3476054

[9] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376550

[10] Nickolas Bloom. 2020. Working from Home and the Future of U.S. Economic Growth under COVID. https://www.youtube.com/watch?v=jtdFIZx3hyk.

[11] Gordon Brown and Michael Prilla. 2019. Evaluating Pointing Modes and Frames of Reference for Remotely Supporting an Augmented Reality User in a Collaborative (Virtual) Environment: Evaluation within the Scope of a Remote Consultation Session. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 713–717. https://doi.org/10.1145/3340764.3344896

[12] William A. S. Buxton. 1992. Telepresence: Integrating Shared Task and Person Spaces. In *Proceedings of the Conference on Graphics Interface '92* (Vancouver, British Columbia, Canada). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 123–129.

[13] Paul-Christian Bürkner and Matti Vuorre. 2019. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science* 2, 1 (2019), 77–101. https://doi.org/10.1177/2515245918823199 arXiv:https://doi.org/10.1177/2515245918823199

[14] Gutwin Carl and Greenberg Saul. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW)* 11, 3 (Sept. 2002), 411–446. https://doi.org/10.1023/A:1021271517844

[15] T. Chassin, J. Ingensand, M. Lotfian, O. Ertz, and F. Joerin. 2019. Challenges in creating a 3D participatory platform for urban development. *Advances in Cartography and GIScience of the ICA* 1 (2019), 3. https://doi.org/10.5194/ica-adv-1-3-2019

[16] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173715

[17] Martin Feick, Anthony Tang, and Scott Bateman. 2018. Mixed-Reality for Object-Focused Remote Collaboration. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings* (Berlin, Germany) *(UIST '18 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 63–65. https://doi.org/10.1145/3266037.3266102

[18] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 513–520. https://doi.org/10.1145/642611.642701

[19] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 449–459. https://doi.org/10.1145/2642918.2647372

[20] William W. Gaver, Abigail Sellen, Christian Heath, and Paul Luff. 1993. One is Not Enough: Multiple Views in a Media Space. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 335–341. https://doi.org/10.1145/169059.169268

[21] Leonardo Giusti, Kotval Xerxes, Amelia Schladow, Nicholas Wallen, Francis Zane, and Federico Casalegno. 2012. Workspace Configurations: Setting the Stage for Remote Collaboration on Physical Tasks. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (Copenhagen, Denmark) *(NordiCHI '12)*. Association for Computing Machinery, New York, NY, USA, 351–360. https://doi.org/10.1145/2399016.2399071

[22] Martin Hachet, Fabrice Decle, Sebastian Knoedel, and Pascal Guitton. 2008. Navidget for Easy 3D Camera Positioning from 2D Inputs. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. United States, 83–88. https://hal.archives-ouvertes.fr/hal-00308251

[23] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 559–568. https://doi.org/10.1145/2047196.2047270

[24] Allison Jing, Kieran William May, Mahnoor Naeem, Gun Lee, and Mark Billinghurst. 2021. EyemR-Vis: Using Bi-Directional Gaze Behavioural Cues to Improve Mixed Reality Remote Collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 283, 7 pages. https://doi.org/10.1145/3411763.3451844

[25] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-Ing into the Uncanny Valley of XR Telepresence. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 59 (apr 2021), 31 pages. https://doi.org/10.1145/3449133

[26] Nicolas Kahrl, Michael Prilla, and Oliver Blunk. 2020. Show Me Your Living Room: Investigating the Role of Representing User Environments in AR Remote Consultations. In *Proceedings of the Conference on Mensch Und Computer* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 267–277. https://doi.org/10.1145/3404983.3405520

[27] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4521–4532. https://doi.org/10.1145/2858036.2858465

[28] Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. JackIn Space: Designing a Seamless Transition between First and Third Person View for Effective Telepresence Collaborations. In *Proceedings of the 8th Augmented Human International Conference* (Silicon Valley, California, USA) *(AH '17)*. Association for Computing Machinery, New York, NY, USA, Article 14, 9 pages. https://doi.org/10.1145/3041164.3041183

[29] Anna K. Kuhlen and Susan E. Brennan. 2013. Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review* 20, 1 (2013), 54–72. https://doi.org/10.3758/s13423-012-0341-8

[30] André Kunert, Alexander Kulik, Stephan Beck, and Bernd Froehlich. 2014. Photoportals: Shared References in Space and Time. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1388–1399. https://doi.org/10.1145/2531602.2531727

[31] Hideaki Kuzuoka. 1992. Spatial Workspace Collaboration: A SharedView Video Support System for Remote Collaboration Capability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 533–540. https://doi.org/10.1145/142750.142980

[32] Joel Lanir, Ran Stone, Benjamin Cohen, and Pavel Gurevich. 2013. Ownership and Control of Point of View in Remote Assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2243–2252. https://doi.org/10.1145/2470654.2481309

[33] Torrin M. Liddell and John K. Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79 (2018), 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

[34] Xuan Luo, Jia-Bin Huang, Richard Szeliski, Kevin Matzen, and Johannes Kopf. 2020. Consistent Video Depth Estimation. *ACM Trans. Graph.* 39, 4, Article 71 (jul 2020), 13 pages. https://doi.org/10.1145/3386569.3392377

[35] E. Marchand, F. Spindler, and F. Chaumette. 2005. ViSP for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robotics and Automation Magazine* 12, 4 (December 2005), 40–52.

[36] Peter Mohr, Shohei Mori, Tobias Langlotz, Bruce H. Thomas, Dieter Schmalstieg, and Denis Kalkofen. 2020. Mixed Reality Light Fields for Interactive Remote Assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376289

[37] Jens Müller, Roman Rädle, and Harald Reiterer. 2017. Remote Collaboration With Mixed Reality Displays: How Shared Virtual Landmarks Facilitate Spatial Referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6481–6486. https://doi.org/10.1145/3025453.3025717

[38] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 405–415. https://doi.org/10.1145/2807442.2807497

[39] Niklas Osmers and Michael Prilla. 2020. Getting out of Out of Sight: Evaluation of AR Mechanisms for Awareness and Orientation Support in Occluded Multi-Room Settings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376742

[40] Kyoung S. Park, Abhinav Kapoor, and Jason Leigh. 2000. Lessons Learned from Employing Multiple Perspectives in a Collaborative Virtual Environment for Visualizing Scientific Data. In *Proceedings of the Third International Conference on Collaborative Virtual Environments* (San Francisco, California, USA) *(CVE '00)*. Association for Computing Machinery, New York, NY, USA, 73–82. https://doi.org/10.1145/351006.351015

[41] Abhishek Ranjan, Jeremy P. Birnholtz, and Ravin Balakrishnan. 2007. Dynamic Shared Visual Spaces: Experimenting with Automatic Camera Control in a Remote Repair Task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1177–1186. https://doi.org/10.1145/1240624.1240802

[42] Troels A. Rasmussen and Weidong Huang. 2019. SceneCam: Using AR to Improve Multi-Camera Remote Collaboration. In *SIGGRAPH Asia 2019 XR* (Brisbane, QLD, Australia) *(SA '19)*. Association for Computing Machinery, New York, NY, USA, 36–37. https://doi.org/10.1145/3355355.3361892

[43] Patrick Salamin, Daniel Thalmann, and Frédéric Vexo. 2006. The Benefits of Third-Person Perspective in Virtual and Augmented Reality?. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (Limassol, Cyprus) *(VRST '06)*. Association for Computing Machinery, New York, NY, USA, 27–30. https://doi.org/10.1145/1180495.1180502

[44] Sheree May Saßmannshausen, Jörg Radtke, Nino Bohn, Hassan Hussein, Dave Randall, and Volkmar Pipek. 2021. *Citizen-Centered Design in Urban Planning: How Augmented Reality Can Be Used in Citizen Participation Processes.* Association for Computing Machinery, New York, NY, USA, 250–265. https://doi.org/10.1145/3461778.3462130

[45] Wendy A. Schafer and Doug A. Bowman. 2005. Integrating 2D and 3D Views for Spatial Collaboration. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '05).* Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/1099203.1099210

[46] Michael F. Schober. 1995. Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes* 20, 2 (1995), 219–247. https://doi.org/10.1080/01638539509544939 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01638539509544939

[47] R. N. Shepard and J. Metzler. 1971. Mental Rotation of Three-Dimensional Objects. *Science* 171, 3972 (Feb. 1971), 701–703. https://doi.org/10.1126/science.171.3972.701

[48] Joon Gi Shin, Gary Ng, and Daniel Saakes. 2018. Couples Designing Their Living Room Together: A Study with Collaborative Handheld Augmented Reality. In *Proceedings of the 9th Augmented Human International Conference* (Seoul, Republic of Korea) *(AH '18).* Association for Computing Machinery, New York, NY, USA, Article 3, 9 pages. https://doi.org/10.1145/3174910.3174930

[49] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D Mobile Collaboration with Spatial Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 179–188. https://doi.org/10.1145/2470654.2470679

[50] Mengu Sukan, Steven Feiner, Barbara Tversky, and Semih Energin. 2012. Quick Viewpoint Switching for Manipulating Virtual Objects in Hand-Held Augmented Reality Using Stored Snapshots. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (ISMAR '12).* IEEE Computer Society, USA, 217–226. https://doi.org/10.1109/ISMAR.2012.6402560

[51] Hongling Sun, Yue Liu, Zhenliang Zhang, Xiaoxu Liu, and Yongtian Wang. 2018. Employing Different Viewpoints for Remote Guidance in a Collaborative Augmented Environment. In *Proceedings of the Sixth International Symposium of Chinese CHI* (Montreal, QC, Canada) *(ChineseCHI '18).* Association for Computing Machinery, New York, NY, USA, 64–70. https://doi.org/10.1145/3202667.3202676

[52] Matthew Tait and Mark Billinghurst. 2015. The Effect of View Independence in a Collaborative AR System. *Comput. Supported Coop. Work* 24, 6 (dec 2015), 563–589. https://doi.org/10.1007/s10606-015-9231-8

[53] Chiew Seng Sean Tan, Kris Luyten, Jan Van Den Bergh, Johannes Schöning, and Karin Coninx. 2014. The Role of Physiological Cues during Remote Collaboration. *Presence: Teleoperators and Virtual Environments* 23, 1 (02 2014), 90–107. https://doi.org/10.1162/PRES_a_00168 arXiv:https://direct.mit.edu/pvar/article-pdf/23/1/90/1625375/pres_a_00168.pdf

[54] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2002. Experimental Evaluation of Augmented Reality in Object Assembly Task. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality (ISMAR '02).* IEEE Computer Society, USA, 265.

[55] Markus Tatzgern, Raphael Grasset, Denis Kalkofen, and Dieter Schmalstieg. 2014. Transitional Augmented Reality navigation for live captured scenes. In *2014 IEEE Virtual Reality (VR).* 21–26. https://doi.org/10.1109/VR.2014.6802045

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

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

[58] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. 2020. TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20).* Association for Computing Machinery, New York, NY, USA, 182–195. https://doi.org/10.1145/3379337.3415827

[59] John Wang and Edwin Olson. 2016. AprilTag 2: Efficient and robust fiducial detection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).*

[60] Peng Wang, Shusheng Zhang, Mark Billinghurst, Xiaoliang Bai, Weiping He, Shuxia Wang, Mengmeng Sun, and Xu Zhang. 2020. A comprehensive survey of AR/MR-based co-design in manufacturing. *Engineering with Computers* 36 (2020), 1715–1738. Issue 4. https://doi.org/10.1007/s00366-019-00792-3

[61] Michael Wittkämper, Irma Lindt, Wolfgang Broll, Jan Ohlenburg, Jan Herling, and Sabiha Ghellal. 2007. Exploring Augmented Live Video Streams for Remote Participation. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA) *(CHI EA '07).* Association for Computing Machinery, New York, NY, USA, 1881–1886. https://doi.org/10.1145/1240866.1240915

[62] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. 2018. Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 853–866. https://doi.org/10.1145/3242587.3242597

[63] Longqi Yang, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, and Jaime Teevan. 2021. The effects of remote work on collaboration among information workers. *Nature Human Behaviour* (2021). https://doi.org/10.1038/s41562-021-01196-4

Received January 2022; revised April 2022; accepted August 2022

université PARIS-SACLAY

Titre : Favoriser la collaboration dans les grands espaces interactifs

Mots clés : interaction humain-machine, travail coopératif assisté par ordinateur, réalité virtuelle, réalité augmentée, téléprésence, interaction 3D

Résumé : Avec la croissance exponentielle de la quantité et de la complexité des données numériques produites par notre société, le besoin d'outils informatiques pour collaborer n'a jamais été aussi important. Permettre à des groupes d'utilisateurs de manipuler, d'analyser et de comprendre ces données, tout en conservant le contrôle sur la façon dont l'intelligence artificielle les traite, est devenu un défi majeur. Dans ce contexte, mes recherches étudient comment les grands espaces interactifs, tels que les murs d'images, les systèmes immersifs de réalité virtuelle ou les espaces de réalité augmentée, peuvent favoriser la collaboration entre les utilisateurs.

La première partie de mon travail explore de nouveaux paradigmes d'interaction permettant aux utilisateurs de maîtriser les caractéristiques inhabituelles des grands espaces interactifs. Au-delà de l'interaction à un niveau individuel, il s'agit d'étudier comment ces systèmes peuvent favoriser la collaboration entre utilisateurs co-localisés. La seconde partie de mon travail porte sur la collaboration à distance entre espaces interactifs. Elle propose à la fois des solutions techniques pour connecter des plateformes hétérogènes, et des solutions pour favoriser la perception mutuelle et la communication entre les collaborateurs distants. Plutôt que de chercher à reproduire la collaboration dans le monde physique, mon travail propose d'aller au-delà en exploitant les capacités numériques et le grand espace physique qui entoure les utilisateurs.

Mes travaux futurs se concentreront sur comment exploiter au mieux le continuum de la réalité mixte pour permettre à des utilisateurs d'interagir et de collaborer à différents niveaux de ce continuum. L'objectif principal est de pouvoir s'adapter aux différentes phases de la collaboration dans des situations hybrides impliquant à la fois des participants co-localisés et distants.

Title: Supporting Collaboration in Large Interactive Spaces

Keywords: human-computer interaction, computer-supported cooperative work, virtual reality, augmented reality, telepresence, 3D interaction

Abstract: As the quantity and complexity of digital data produced by our society grow exponentially, the need for computer-supported collaboration has never been higher. Empowering groups of users to manipulate, analyze and understand this data, while preserving control over how artificial intelligence processes it, has become a major challenge. In this context, my research investigates how large interactive spaces, such as wall-sized displays, immersive virtual reality systems or augmented reality spaces, can foster collaboration among users.

A first part of my work investigates new interaction paradigms that provide users with the ability to master the unusual characteristics of such large interactive spaces. Beyond individual interaction, it investigates how these systems can foster co-located collaboration by providing appropriate collaborative interaction among users. A second part of my work focuses on remote collaboration across large interactive spaces. It explores technical solutions to connect heterogeneous platforms, as well as telepresence systems providing appropriate awareness and communication cues among the remote collaborators. Rather than mimicking collaboration in the physical world, it aims to push collaboration beyond "being there" by leveraging digital cues and taking advantage of the large physical space surrounding users.

My future research will concentrate on exploiting the mixed reality continuum to enable collaborators to interact across time and space by seamlessly transitioning between heterogeneous interaction modalities. The overall objective is to support the different phases of a collaboration in hybrid situations, involving both co-located and remote participants.