MAGIC: Manipulating Avatars and Gestures to Improve Remote Collaboration

Catarina Gonçalves Fidalgo

Thesis to obtain the Master of Science Degree in

Aerospace Engineering

Supervisors: Prof. Joaquim Armando Pires Jorge
Dr. António Maurício Lança Tavares de Sousa

Examination Committee

Chairperson: Prof. Paulo Jorge Coelho Ramalho Oliveira
Supervisor: Prof. Joaquim Armando Pires Jorge
Member of the Committee: Prof. Carlos Alberto Pacheco dos Anjos Duarte

October 2020

Acknowledgments

First and foremost, I would like to thank Professor Joaquim Jorge for giving me the opportunity to integrate this project and not giving up on me when I doubted myself. For his insight, support, and sharing of knowledge that made this Thesis possible.

I would also like to express my sincere gratitude to Maurício Sousa for his continuous help, patience, motivation, and enthusiasm. For always being there for me and believing in me.

Additionally, I would like to thank Daniel Medeiros, Daniel Mendes, Rafael Dos Anjos, and Karan Singh for their contributions to this work.

Besides, a special thank you to all the people that participated in the experimental sessions and were willing to help me made this work possible even in times of pandemic.

I would also like to thank my friends Emanuel, Henrique, Inês, Jinho, João, Meni, Rochette, Saraiva, Sofi, and Tavares for being by my side from our first year of university, for helping me grow as a person and being always there for me during the good and bad times in my life. I will forever cherish all the moments we spent together and all the new things you made me discover. Furthermore, a special thank you to Henrique, Jacob, Jinho, Rita, and Moita for their support during the development of this work. For giving me the motivation to work when I needed to and knowing when was the right time to have fun. For always stimulating exciting discussions on the most diverse topics and making me become a more instructed person. I will miss us living all together.

A special thank you to my dearest friend Henrique for being with me from the begging of all, for supporting me through all these years of studies and showing me never to give up. For helping me build a home in Lisbon and making all my days happier.

I also want to express my deepest gratitude to my friend Jinho. Thank you for your love and support and for always making me do the right thing. Thank you for making me believe in myself. I could not have accomplished this without you. May you rejoice my life for many more years.

Last but not least, I would like to thank my mom for her friendship, encouragement, and caring over all these years, for always being there for me and believing in me even I didn't. I would also like to thank my brother, my godmother, and my grandparents, for their understanding and support.

To each and every one of you, my warmest thank you.

Abstract

Remote work plays an essential role in projects in which geographically separated people need to perform joint tasks. Apart from saving time and resources, collaborating remotely is helpful in situations when confinement is required. Life-sized face-to-face telepresence promotes the sense of ”being there” and improves collaboration by allowing an immediate understanding of nonverbal cues. However, when discussing shared 3D content in a face-to-face setting, having different points-of-view of the model paired with occlusions raises ambiguities in analysis, decreasing workspace awareness. In this dissertation, we introduce MAGIC, a novel telepresence approach that improves remote collaboration in shared 3D workspaces by allowing participants to communicate through nonverbal cues while sharing the same workspace perspective, integrating task-, person-, and reference-space seamlessly. To enable a face-to-face setting with shared perspective, we manipulate the remote participant's representations and gestures so that they correctly apply to the local person's reference space. To evaluate our approach, we developed a Virtual Reality prototype that combines the virtual spaces of two remote collaborators so that they can work together in the same worktable. We capture the collaborators' hands and head positions in real-time and use them to animate fully rigged avatars through inverse kinematics. Results from a user evaluation suggest that MAGIC is effective in improving task performance and increasing workspace awareness. Furthermore, manipulations improved the sense of presence between remote collaborators despite being unnoticed.

Keywords

Remote Collaboration; Workspace Awareness; Nonverbal Communication; Perception Manipulation; Virtual Reality.

Resumo

O trabalho remoto desempenha um papel fundamental em projetos nos quais pessoas geograficamente separadas necessitam de realizar tarefas conjuntas. Além de economizar tempo e recursos, poder colaborar remotamente é útil em situações em que o confinamento é necessário. Sistemas de telepresença frente a frente promovem a sensação de presença e melhoram a colaboração, permitindo a compreensão imediata de sinais não verbais. No entanto, aquando da discussão de conteúdo 3D frente a frente, ter diferentes pontos de vista e oclusões de partes do modelo levantam ambiguidades na análise, diminuindo a consciência do espaço de trabalho. Nesta dissertação, apresentamos o MAGIC, uma nova abordagem de telepresença que melhora a colaboração remota em espaços de trabalho 3D. O MAGIC permite aos colaboradores usar sinais não verbais enquanto partilham a mesma perspetiva do espaço de trabalho, integrando-o com os espaços da pessoa e de referência de forma contínua. Para permitir uma disposição frente a frente com partilha de perspetiva, manipulamos a representação do colaborador remoto para que os seus gestos façam sentido no ambiente local. Para avaliar a abordagem, desenvolvemos um protótipo em Realidade Virtual que combina os espaços virtuais de dois colaboradores remotos para que possam trabalhar juntos na mesma mesa. Capturamos a posição das mãos e da cabeça dos colaboradores e usamo-las para animar avatares com cinemática inversa. Resultados de uma sessão de avaliação com utilizadores sugerem que o MAGIC é eficaz em melhorar o desempenho de tarefas e aumentar a consciência do espaço de trabalho. Adicionalmente, as manipulações melhoraram a sensação de presença entre colaboradores apesar de não terem sido notadas.

Palavras Chave

Colaboração Remota; Consciência do Espaço de Trabalho; Comunicação Não Verbal; Manipulação de Percepção; Realidade Virtual.

Contents

List of Figures

List of Tables

Acronyms

AIXR	Academy of International Extended Reality
AR	Augmented Reality
CAD	Computer Aided Design
CH	Convex Hull
CRV	Centro de Realidade Virtual
CSCW	Computer Supported Cooperative Work
CSG	Constructive Solid Geometry
FABRIK	Forward And Backward Reaching Inverse Kinematic
HMD	Head Mounted Display
IST	Instituto Superior Técnico
MR	Mixed Reality
SAS	Secondary Air System
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UI	User Interface
VR	Virtual Reality
WHO	World Health Organization
WYSIWIS	What You See Is What I See

1

Introduction

Contents

Many leading organizations have been operating globally for decades, and, nowadays, global working has become a universal concept. People work in increasingly connected, interdependent, and integrated teams, collaborating across time zones and geographies. Statistics[1] from *FlexJobs & Global Workplace Analytics* show that from 2005 to 2017, the number of people working remotely in the U.S. increased 159% as it allows for considerable savings in time and resources. Furthermore, the current global circumstances caused by the Covid-19 outbreak demonstrated that collaborating remotely is of significant importance in situations where social confinement is required. Being able to work remotely minimizes the risk of contracting and spreading the virus, saving lives, and supporting the worldwide economy.

For remote collaboration to work, traditional approaches using voice and video have been used for decades. However, these technologies neglect essential aspects of interpersonal nonverbal communication such as gaze, body posture, and gestures to indicate objects referred to in speech, essential to a face-to-face meeting. These limitations negatively affect how remote people interact with a local user, especially for operations that require 3D spatial referencing and action demonstration.

In fact, 3D models play a crucial role in many fields. In engineering and architectural scenarios, 3D models are used instead of traditional physical models in different stages of a project, from its conception to its presentation to possible investors, assisting in the design of new technological devices, vehicles, and structures (see Figure 1.1 - A and B). In the medical field, 3D models are used for tissue and organ fabrication and creation of customized prosthetics and implants [1](see Figure 1.1 - C). Furthermore, detailed 3D models of organs, constructed from multiple MRI's and CT scans [2], can improve the understanding of a patient anatomy for pre-surgical planning [3] (see Figure 1.1 - F). This type of models are also important in the geology field (Figure 1.1 - D), and the video game and movie industries (see Figure 1.1 - E).

That said, the development of 3D models constitutes a very interdisciplinary field and often involves the collaboration of people from different areas of specialization, which may need to collaborate from geographically separated areas. In this scenario, remote collaboration through video conference compromises the operation's efficiency, as the analysis of complicated 3D structures using only words can be ambiguous and vague.

[1] "159% Increase in Remote Work Since 2005: FlexJobs Global Workplace Analytics Report", Brie Weiler Reynolds (Accessed: 2020-08-10): https://www.flexjobs.com/blog/post/flexjobs-gwa-report-remote-growth/

[2] Prosthetic Leg 3D Model, Sakib Ahmed (Accessed: 2020-08-24) : https://grabcad.com/library/66595

[3] Familiar House, Marie-Rose Gonçalves

[4] Gas Turbine Engine 3D Model, Sara Diogo (Accessed: 2020-08-24) : https://grabcad.com/library/gas-turbine-engine-1

[5] 3D geological and mineral potential mapping, Kenex (Accessed: 2020-08-24) : http://kenex.com.au/Services/3dmodel.asp

[6] Linhi Character 3D Model, Quang Phan (Accessed: 2020-08-24) : https://sketchfab.com/3d-models/linhi-character-preview-3e3ea18a5e4c4e609eae767aa08e36db

[7] EchoPixel's interactive mixed reality software platform for improved surgical imaging (Accessed: 2020-08-24) : https://echopixeltech.com/

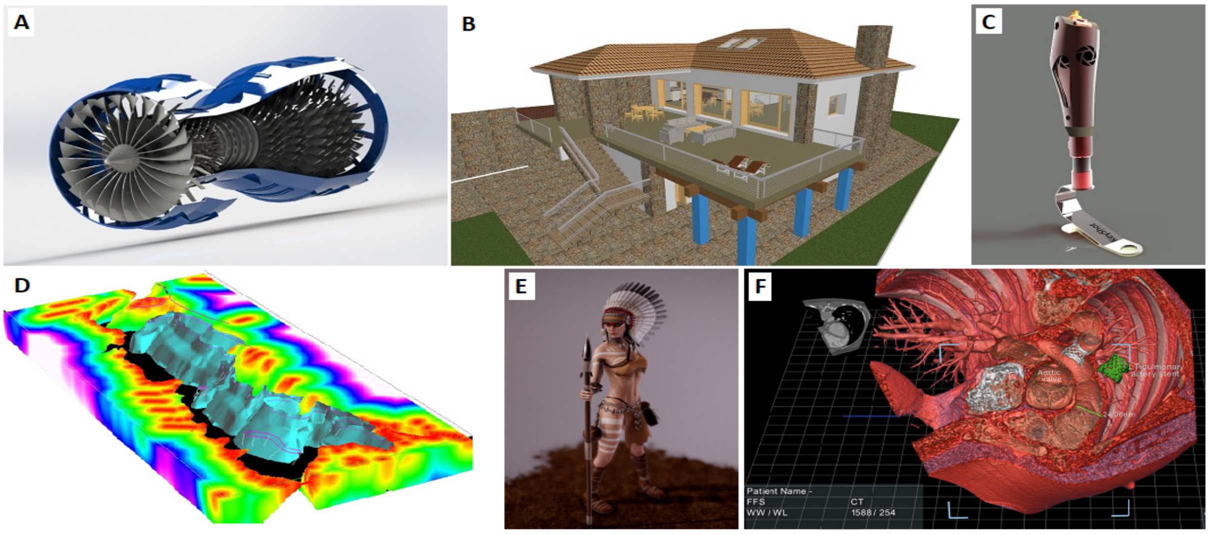
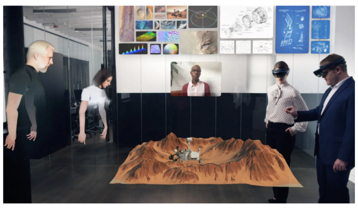


Figure 1.1: Examples of 3D models used in different fields: [A] - 3D Computer Aided Design (CAD) model of a gas turbine engine [2], [B] - 3D CAD of a familiar house [3], [C] - 3D CAD model of a prosthetic leg[4], [D] - 3D geological and mineral potential mapping[5], [E] - 3D model of game character[6], [F] - Organ 3D model constructed from medical images for pre-surgical planning[7].

Key new technology for enabling remote collaboration is Mixed Reality (MR), which encompasses both Virtual Reality (VR) and Augmented Reality (AR) via immersive technology and is growing increasingly capable and accessible every year. We can point out the Spatial [8] startup that developed a "holographic" MR collaboration platform where people can meet in a virtual room though AR, represented by resembling avatars if using a VR headset, or even joining through video stream, as can be seen in Figure 3.1a. Other up to date virtual collaboration platforms are Hubs by Mozilla [9], where people can meet in a shared virtual space represented by an avatar of choice (see Figure 3.1b), or MeetinVR [10] that enables whiteboard interactions, 3D viewing, or podium speeches in a virtual room of choice (see Figure 1.2c).



(a) MR meeting powered by Spatial's collaboration platform.

(b) Different representations used in a VR meeting powered by Mozilla Hubs.

(c) VR meeting powered by MeetinVR collaboration platform.

Figure 1.2: State-of-the-art MR platforms for remote collaboration.

[8]Spatial: https://spatial.io/
[9]Hubs by Mozilla: https://hubs.mozilla.com/
[10]MeetinVR: https://meetinvr.com/

Remote collaboration through MR allows teams to meet in life-sized representations and spontaneously discuss a topic, as if sharing the same office, enabling a high level of co-presence. Indeed, previous research found that for a remote meeting to be closer to a co-located experience, it should rely on real size portrayal of remote people to maintain the sense of ”being there” [4, 5]. Furthermore, people have higher task performance when able to communicate using full-body gestures [6]. Working in a face-to-face setting benefits collaboration as users can see the virtual representation of their partner and the virtual content at the same time. However, the different perspectives of the 3D content resulting from the face-to-face setting induce ambiguities when using pointing gestures and demonstrations. Indeed, to analyze 3D content, a shared perspective is more effective and preferred when compared to an opposing point of view [7, 8]. Unfortunately, in the real world, it is impossible to collaborate face to face while sharing perspective.

In this dissertation, we present MAGIC, a new approach to collaborative remote work that enables two people to engage in collaborative work in a virtual office, adopting a face-to-face configuration, while sharing each other's perspective of the workspace. To enable share of perspective in a face-to-face setting, the real-sized virtual representations of collaborators are subtlety manipulated so that the remote person's gestures make sense to the local observer. With MAGIC, we aim to increase the common understanding of a virtual workspace to increase task performance, without affecting co-presence between users.

1.1 Motivation

The increasing technological complexity present in the Aerospace Engineering sector makes it one of the fields with more intricate 3D models. The design process of aircraft components often requires a panel of experts to analyze all the 3D elements to guarantee no errors are made during their development. Therefore, a rigorous spacial comprehension of the 3D models engineers are working with is crucial.

MR can offer a significant improvement in the collaboration task, as it can improve the shared understanding of the models engineers are working on. In fact, Embraer already takes advantage of a Virtual Reality Center (Centro de Realidade Virtual (CRV)), which enables Embraer's engineers to visualize the aircraft's structure and systems during the development phase, reducing the time spent on the development of any aircraft when compared to traditional 3D software. The CRV allows for Design Review Technical Meetings, where engineers examine the different aircraft 3D models to check for interferences in their integration in the aircraft. In 2018 alone, 467 Design Review sessions took place, involving more than 10,000 users [11].

[11] “How Virtual Reality Speeds Up Aircraft Development (Journal of Wonder - Embraer): ”https://journalofwonder.embraer.com/global/en/82-how-virtual-reality-speeds-up-aircraft-development

However, engineers still face some challenges in this type of virtual encounters. When virtually collaborating face-to-face, we have to deal with opposing points-of-view of the workspace and do not share the same forward-backward orientation. There is no common orientation of right or left, and there can be occlusions of parts of the workspace. This constrains people's ability to use descriptions of relative positions and affects the understanding of where or what the remote person is pointing at [9], leading to communication missteps and causing tasks to be more laborious. Figure 1.3 illustrates the occlusion issue present when people have opposing points-of-view of the workspace.

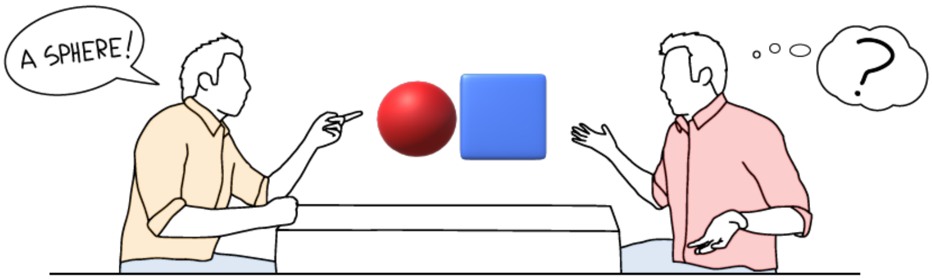


Figure 1.3: Example illustration depicting the occlusion issue present when people have opposing points-of-views (from [9]).

1.2 Research Statement

In this work, we introduce MAGIC, a new technique that allows two collaborators to share the same point-of-view of a specific workspace without compromising the sense of co-presence that is ensured when people interact in a face-to-face arrangement. We intend to make our contribution in validating the assumption that a virtual meeting that involves 3D content is more productive if the participants can stay face to face while sharing the same point of view, without added complexity for local participants.

Gutwin and Greenberg [10] define Workspace Awareness as the *"up-to-the-moment understanding of another person's interaction with a shared workspace"*, which involves knowledge about what other people are doing and what they are about to do. Thus, sharing the same perspective of the workspace increases workspace awareness as it avoids occlusions and reduces ambiguity.

However, supposing the remote collaborator shares the same point-of-view as the local collaborator but is rendered in front of it; his gestures will not match the local space, making it hard to use non-verbal communication mechanisms, especially *deictic gestures*. Deictic gestures refer to gestures that are produced to direct someone's attention towards a specific referent in the environment, with pointing gestures being the main representative of this category [11]. Recent work suggests that the interpretation of deictic gestures can be significantly improved by warping the pointing arm [12, 13]. In MAGIC, we resort to manipulating the remote collaborator's representation to guarantee that his gestures make sense to the local observer.

We can then highlight the research statement of this dissertation as follows:

> *Manipulating remote people's virtual representations to enable a shared point-of-view improves workspace awareness in face-to-face shared 3D workspaces*

As we aimed to show that when a collaborator indicates a specific area of interest in a shared workspace, while standing in a face-to-face setting, the observing partner can identify it better if both share the same perspective, our main hypothesis throughout the dissertation is:

H1: *A face-to-face setting coupled with manipulations that enable perspective sharing improves workspace awareness.*

We hypothesize that manipulations that enable perspective sharing in a face-to-face setting enhance nonverbal communication mechanisms (see section 2.1), as a shared perspective avoids occlusions of the workspace, and enables the use pointing gestures and descriptions of relative positions. This leads to an increase of shared understanding when performing collaboration tasks on virtual 3D spaces.

Additionally, we wanted to confirm that our approach did not distract people from focusing on the main collaborative task; and to assess whether the shared point of view with corrected gestures directly impacts the feeling that the collaborator is present. As such, our remaining hypotheses are:

H2: *Manipulations applied to the remote person's representation are not noticeable by the local collaborator.*

We hypothesize that the manipulations we apply to the remote collaborator's representation are subtle enough to make the local collaborator unaware of them being applied. Unnoticeable manipulations do not require collaborators to learn a new interaction technique, nor they distract them from the collaboration task.

H3: *Manipulations applied to the remote person's representation improve the feeling of co-presence.*

We hypothesize that manipulating remote people's representation does not negatively affect the collaboration task in terms of the sense of co-presence felt between collaborators as we believe collaborators are not aware of them. Furthermore, we believe that an increased awareness of the remote person's actions can increase its presence.

1.3 Results

We conducted a user study to evaluate the impact of our approach on workspace awareness, task performance, and feeling of co-presence. We present the synthesis of the results below:

- **A face-to-face setting coupled with manipulations that enable perspective sharing improves workspace awareness.** Results showed an improvement in people's capacity to use and understand deictic gestures as there was an increase in the accuracy of the matching task by 13% compared to traditional face-to-face. Hence, manipulations that enable remote collaborators to share the point of view while remaining in a face-to-face setting improve workspace awareness. As MAGIC increases task performance, we argue that manipulations do not add complexity to the collaboration task.
- **Manipulations applied to the remote person's representation are not noticeable by the local collaborator.** None of the participants reported noticing anything strange under Condition A or B in the User Preference Questionnaires, which leaves us to assume that participants remained unaware of the experiment's manipulations. Results of the user evaluation showed no statistically significant differences in task completion time between both conditions. Hence, MAGIC does not negatively impact the time collaborators take to complete the shared task, remaining straightforward. Since collaborators do not notice manipulations, there is no need for them to learn a new interaction method, as it appears to them that they are collaborating traditionally.
- **Manipulations applied to the remote person's representation improve the feeling of co-presence.** Results from the User Preference Questionnaires also showed participants felt that their partner was more present in the virtual environment under MAGIC. We believe the manipulation of remote people's representation to enable a shared perspective improved the sense of co-presence since they increased awareness of the remote collaborator's actions in the workspace.

Finally, we can argue that when collaborating face-to-face with 3D content, we can improve workspace awareness and increase task performance by enabling the same point of view for the two collaborators. The manipulations we employ to guarantee that the virtual remote collaborator is correctly rendered in front of the local user are not noticeable; nonetheless, they positively influence the sense of the presence of the remote user.

1.4 Contributions

Bearing in mind the problem of interacting with 3D virtual objects in face-to-face collaboration scenarios, our work offers the following contributions:

- **Design Space**

  A new approach on object-centered collaboration that integrates person-, task- and reference spaces (see Section 2.2), enabling users to interact with 3D content in a face to face disposition with improved awareness of the workspace;

- **Representation Manipulation Techniques**

  Detailed overview of the manipulation techniques we used to enable the same point of view of the workspace for the two users, that include a mirror-reversal of the remote person's representation, followed by a repositioning of the pointing arm and an adjustment of the embodiment's position along the forward-backward axis (see Chapter 3).

- **VR Prototype**

  Implementation details of the virtual office prototype we developed to study our approach, which combines the virtual spaces of two remote collaborators so that they can analyze 3D objects together on the same worktable, in a face-to-face setting, while sharing perspective (see Chapter 4).

- **User Study**

  Reports of the user study we conducted in order to evaluate our research statement hypotheses, which includes details on the design and procedure of the experiment, apparatus and equipment, participants' analysis, and results obtained (see Chapter 5).

1.5 Publications

Though the work developed in this dissertation, we developed a research paper on this topic submitted for the ACM CHI Conference on Human Factors in Computing Systems 2021, to be evaluated by an international panel of experts.

1.6 Organizational Overview

The remaining contents of this dissertation are organized as follows. In Chapter 2, we discuss related work that motivated and influenced our approach, reviewing existent research in the field of Computer Supported Cooperative Work (CSCW). Chapter 3 describes our proof of concept, MAGIC, for interacting with 3D virtual objects in face-to-face collaboration scenarios with perspective sharing. Chapter 4 goes through the MAGIC implementation process, describing the technology and development that allowed us to achieve our prototype. Chapter 5 reports the experiments conducted to evaluate and validate our approach along with the results obtained. Finally, Chapter 6 concludes our developed work and discusses possible future work.

2

Related Work

Contents

For a full understanding of the work presented in this document, it is important to contextualize it. Collaboration can be defined as "the situation of two or more people working together to create or achieve the same thing" [14]. When people do not share the same physical space, collaboration becomes difficult to achieve, as communicating with each other and observe each others' actions is not as straightforward as in co-located encounters. Nevertheless, a large amount of previous research suggests that to enhance remote collaborative experiences, staying aware of others and aware of the workspace are two fundamental factors. Therefore, a high sense of presence - which in Virtual Reality refers to the user's sense of being inside the simulated environment - coupled with good workspace awareness pay a major role in enhancing remote collaboration. We further explore the state-of-the-art approaches that study workspace awareness and the sense of presence in virtual shared workspaces, providing a discussion where we compare them with our approach.

2.1 Workspace Awareness

The concept of awareness has gained relevance in CSCW research over the years. Simply put, awareness can be defined as "knowing what is going on" [15].

Greenberg et al. [16] identified four overlapping types of awareness that people naturally maintain when working in a group:

- **Informal Awareness** – the general knowledge of *"who's around and what they are up to"* [16];
- **Social awareness** - the information that a person acquires about others in a social or conversational context. This is maintained through nonverbal cues and back-channel feedback, and includes whether the other person is paying attention, their emotional state or level of interest;
- **Group-structural awareness** - the knowledge about people's roles and responsibilities, their positions on an issue, their status, and group processes;
- **Workspace Awareness** – the *"up-to-the-moment understanding of another person's interaction with a shared workspace"* [10], which involves knowledge about what other people are doing and what they are about to do - people's interaction with the workspace.

While maintaining awareness of what surrounds us in a real-world workspace is something we take for granted, the same does not happen easily in virtual collaboration systems. It often seems that collaborating remotely is inefficient and sloppy compared to real-world face-to-face interactions. Hence, it has become increasingly evident that maintaining and developing workspace awareness in remote collaboration systems is the key to making these types of encounters less awkward and more appealing to the public [10].

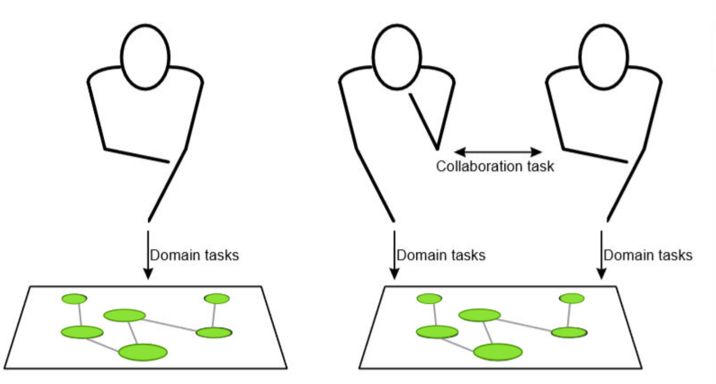


Figure 2.1: Tasks present in individual work (on the left) and collaborative work (on the right) (from [17]).

When an individual works alone in a given workspace, his activities and workspace awareness only involve the workspace and the domain task. However, in the context of collaboration, a new task arises, the collaboration itself; therefore, the workspace awareness will involve the workspace, the domain task, and the collaboration task of communication and decision making (see figure 2.1)

Gutwin and Greenberg [10] also suggest that it is easier to ensure workspace awareness if people can gather information in ways that are familiar to what happens in real environments, with mechanisms such as *consequential communication, feedthrough* and *intentional communication*.

The first source of information is the other person's body: voice, position, posture, movement of head, eyes, arms, and hands provide an abundance of information about what is happening in the workspace. This practice of seeing and hearing other people active in the workspace is called **consequential communication**: *"information transfer that emerges as a consequence of a person's activity within an environment"* [10].

The artifacts in the workspace are a second source of awareness information. When artifacts are manipulated, they give off information, and what would generally be feedback to the person performing the action can also inform others who are watching. This mechanism is called **feedthrough**, and when both the artifact and the person who's using it can be seen, feedthrough is coupled with consequential communication.

The third source of information that is always present in collaboration is conversation and gesture, which makes **intentional communication**. This information gathering happens when people explicitly talk about what they are doing in direct discussions and when overhearing others' conversations or picking others' *verbal shadowing* (the running commentary spoke to no one in particular). Deictic gestures and other visual actions, such as demonstrations, are also used to carry out intentional communication.

Even though these mechanisms can all be achieved in co-located encounters, they are challenging to acquire in remote collaboration due to current technological limitations. The undermining in the perception of intentional communication, specifically failing to understand a pointing gesture's exact target, hinders remote collaborators' capacity to be aware of the collaborative task's context correctly.

In this dissertation, we hypothesize that perception manipulation techniques can improve the knowledge of the actions that are being performed in the workspace, which artifacts are being used and why, enhancing consequential communication and feedthrough. Intentional communication can also be improved by these techniques, as they can make deictic gestures and other visual actions accurate for both the participants.

2.2 Awareness in Shared Workspace

To understand this next section, it is important to define the different types of spaces inherent to the micro-level of communication. Bill Buxton [18] identifies three distinct types of spaces that need to be considered at this level:

- **Person space:** the *"space where one can read the cues about expression, trust, and gaze "* [18]. It is where the voice comes from and where people look when speaking to someone.
- **Task space:** the *"space where the work appears"* [18]. The task space can be a shared space if others can see it, or a private space otherwise. It is the space where a person "changes things" besides only viewing.
- **Reference space:** the *"space within which the remote party can use body language to reference the work "* [18]. Here, the user can employ things like pointing and gesturing.

Most telepresence systems typically rely on a separation between the person space and the task space. However, in [19], Buxton suggested that *"effective telepresence depends on quality sharing of both person and task space"*. Integrating the person and the task spaces is key to enable intentional and consequential communication mechanisms so essential to maintain workspace awareness in the course of a virtual meeting. In 1992, Ishii et al. addressed the problem on how to communicate in a shared environment without separation between the task and the person space by identifying three types of groupware configurations: *"talking in from of a whiteboard"* or the whiteboard metaphor, *"talking over a table"* or the over the table metaphor and *"talking through a glass window"* or the glass window metaphor (Figure 2.2).

In the **whiteboard metaphor**, collaborators are side by side, sharing the common board's orientation. In this configuration, although the two participants share the same reference space, there is a clear separation between the person and the task space making it difficult for the collaborators to simultaneously focus their attention on the task and the other person's visual cues. Therefore, the capacity to recognize consequential and intentional communicative cues is hindered. Also, as the participants share the same space in front of the whiteboard, it is hard to implement a mechanism that can coordinate the

Figure 2.2: Three Metaphors of Seamless Space for Shared Drawing and Face-to-Face Conversations (from [20]).

use of this shared space.

Following the whiteboard metaphor, Kunz et al. [21] opted for rendering the entire upper body of the remote user on top of the displayed shared 2D content with *Collaboard* (Figure 2.3A). By doing so, this approach was able to preserve the meaning of deictic gestures when pointing at displayed shared artifacts. However, CollaBoard makes it difficult for participants to share gaze, limiting consequential awareness. Similarly, Zillner et al. [5] proposed the *3D-Board*, that blended a mirrored front-facing 3D embodiment of the remote user with the digital content of the interactive whiteboard (Figure 2.3B). This creates the impression that the remote user would be standing behind the whiteboard. Their findings suggested that face-to-face telepresence approaches allow greater efficiency when compared to side-by-side configurations. In a different line of thought, Higuchi et al. [22] developed *ImmerseBoard*, that emulates writing side-by-side on a physical whiteboard, or alternatively on a mirror. ImmerseBoard rendered the remote participant in real size on the side of the whiteboard (Figure 2.3C), enabling participants to estimate their remote partners' eye gaze direction along with posture and gestures. Therefore, participants were able to perceive feedthrough, consequential, and intentional communicative cues. However, tilting the virtual task space induced imprecision in perceiving the task space. An extended arm approach (Figure 2.3D) was also studied in ImmerseBoard. However, rendering the person space on top of the task space decreased the workspace's visibility, causing a decrease in feedthrough.

When it comes to the over the **over the table metaphor**, it replicates two people sitting in front of each other, on opposite sides of a table. This metaphor is entirely appropriate for face-to-face communication as the two participants can easily see each other's face and body movements, and there is no conflict between personal spaces. Here, there is an integration of the person and the task spaces. However, a drawing's orientation becomes upside-down for one of the parties, which adds complexity to the collaboration task, reducing workspace awareness.

Using the over the table metaphor, Genest et al. [23] developed the KinectArms toolkit, which cap-

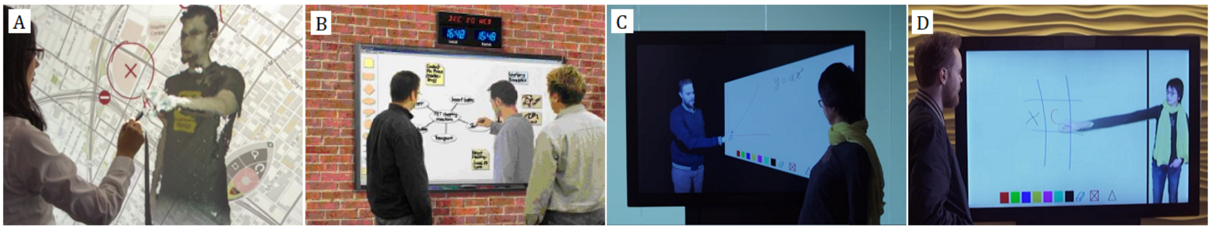


Figure 2.3: Whiteboard metaphor Approaches: [A] - Collaboard [21] , [B] - 3DBoard [5], [C] - ImmerseBoard in tilted task space approach [22], [D] - ImmerseBoard in extended arm approach [22].

tures arm embodiments using depth cameras and displays them in a tabletop (Figure 2.4D). Junuzovic et al. [24] presented the *IllumiShare* approach. Illumishare combined physical and digital objects on a table surface enabling users to collaborate in a common reference space, sharing the same point-of-view. They used IllumiShare to share the task- and reference spaces and Skype Video to share the person space (Figure 2.4B). However, sharing nonverbal communicational cues such as gaze, facial expressions, and body posture was limited to the video stream, with a clear separation between the person- and the task-space. R. S. Sodhi et al. [25] also explore 3D gestures and spatial input in *BeThere*, a self-contained mobile smartphone with integrated depth sensors that tracked the user's fingers and captured the 3D shape of objects in front of the sensor. The local user could capture a real-time 3D model of the task space and transmit it to a remote user who could manipulate a virtual 3D hand to show the local user how to complete a specific task (Figure 2.4A). These approaches focused on rendering representations of remote participants' arms and hands, providing good intentional and feedthrough communication cues. However, they offer limited support for consequential communication mechanisms.

To address the issue of consequential communication, Benko et al. [26] introduced *MirageTable*. In MirageTable, participants shared the same task space interacting with physical objects in a face-to-face display. A depth camera tracked the user's eyes and performed a real-time capture of the shape of any object placed in front of the camera, including the user's body and hands (Figure 2.4C). However, although participants shared the same task space, this approach did not seemingly avoid occlusions. Another work that is considered relevant for our research on shared space awareness is Leithinger et al. 's [27] approach to physical telepresence. Here, two people could manipulate physical shapes in a face-to-face or corner-to-corner formation on the sides of the shared workspace (Figure 2.4D). Again, this approach did not avoid occlusions, limiting intentional communication.

The **glass window metaphor** illustrates collaborators talking through and drawing on a transparent glass window. Similarly to the over table metaphor, there is no conflict on the shared space use since each collaborator's space is isolated from the other by the glass. People can see their partner's face and gestures without shifting their attention from the task space. There is an integration of the person- and task spaces, and the major drawback is that participants do not share a common orientation of "right" and "left" of the drawing space.

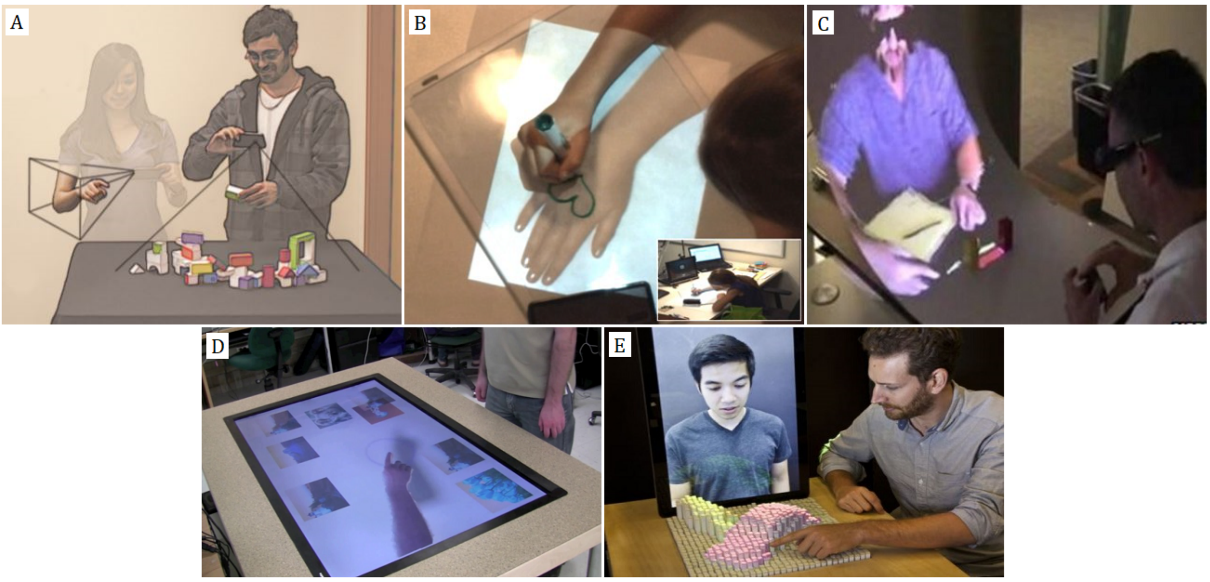


Figure 2.4: Over the table metaphor Approaches: [A] - BeThere [25] , [B] - Illumishare [24], [C] - MirageTable [26], [D] - KinectArms [23], [E] - Leithinger et al. [27].

To explore the glass window metaphor, Ishii et al. developed the Clearboard [20], a transparent display where two participants can engage in collaborative drawing tasks while seeing each other face-to-face (Figure 2.5A). To correct the inaccurate reference system, the authors resorted to mirror-reversing the video image to establish the same point-of-view for both participants, increasing awareness and reducing ambiguity. The transparent display approach has been the subject of research for 2D content collaborative manipulation. With *FacingBoard2*, Li et al. 's [28] reaffirmed Ishii et al. findings and suggested that to maintain awareness in face-to-face interactions, telepresence systems should rely on selective image reversal of text and graphics, displaying different content for each participant, using relaxed What You See Is What I See (WYSIWIS) approaches (Figure 2.5B). In" strict" WYSIWIS, every participant perceives the same task space. We can also refer to *ShadowHands* [29] that captures the gestures of a remote user using a hand tracking system and renders them onto a shared digital workspace. The virtual hand is presented as if the users are separated by a transparent display (Figure 2.5C). Here, however, there is a poor sense of the remote instructor's presence since there is no embodiment of the remote person, which inhibits consequential communication.

2.3 Awareness of Remote People

The sense of presence of remote people in a virtual meeting plays an essential role in people's ability to interact and collaborate.

Systems that enable virtual meetings can be categorized into *audio-conferencing, groupware, video-*

Figure 2.5: Glass window metaphor Approaches: [A] - ClearBoard [20], [B] - FacingBoard2 by Li et al. [28], [C] - ShadowHands [29].

conferencing, telepresence and *collaborative mixed reality systems* [30]. Audio-conferencing technologies include both fixed and mobile telephony services and Internet-based audio tools, while video-conferencing refers to systems that allow multiple people to exchange live audio and video. Groupware applies to computer desktop-based collaboration applications, which usually provide a shared 2D desktop, accessible for a group of people over a network, while conversational interaction is supported via text messages or live audio channels. When it comes to telepresence, it often merges video-conferencing technology with telerobotics, allowing a person to 'teleport' to a remote place. Mixed reality refers to a large class of display technologies that blend physical and virtual worlds.

Traditional systems such as audio- and video-conferencing have maintained a good sense of presence in a social context using only audio and video. However, these technologies neglect some essential aspects of inter-personal nonverbal communication such as gaze and body posture and, when it comes to remote work, these limitations negatively affect how remote people can interact, challenging the ability to maintain a good sense of presence.

Throughout the nineties, research was carried out in the field of remote presence, relying mostly on two-dimensional images from traditional cameras. To simulate a 4-way round table meeting, Sellen [31] created the Hydra system, where a set composed by a camera, a monitor and a speaker were placed in the spot that would otherwise be held by each remote participant. Ichikawa et al. developed Majic [32], a multi-party video-conferencing system that enabled eye contact among people in remote places, with life-sized projected images of participants. Going for a different approach, Morikawa et al. [33] chose to display local and remote people side by side on the same screen, enabling the local person to see its image on the same screen as the remote user, creating an illusion of interaction. These were significant steps in bettering the sense of presence in remote meetings. However, recent developments with commodity depth cameras have enabled three-dimensional representations that permit a more reliable life-size scale portrayal of remote people. Life-sized avatars raised the level of presence in remote encounters by closely mimicking the visual perception we get from a face-to-face encounter in the real world.

In 2009, Johns et al. [34] presented a set of algorithms and an associated display system capable of producing correctly rendered eye contact between a remote participant and a group of local observers. They resorted to transmitting the remote person's scanned face into a 3D auto-stereoscopic display system in real-time. The remote person was not required to use any headgear. Also without the user's need to wear any tracking or viewing apparatus, Maimone et al. [35] used commodity depth cameras to capture the remote person in real-time, resorting to a 3D display for a correct visualization. Beck et al. [36] presented an immersive telepresence system that allowed distant groups of multiple users to meet and explore a shared workspace. A cluster of depth cameras continuously captured participants and physical objects at each site. The captured 3D data was then transferred to the remote location in real-time and displayed within the shared virtual environment. Room2Room [4] resorted to color and depth cameras do capture a 3D image of the remote user and used commodity projectors to project their life-size virtual copy into the local space, creating the illusion of co-presence as well as a shared understanding of verbal and non-verbal cues. In 2016, Orts-Escolano et al. presented Holoportation [37], a significant improvement in augmented and virtual reality telepresence. Holoportation utilizes custom depth cameras to render high-quality, real-time reconstructions of people, furniture, and objects, allowing users wearing virtual or augmented reality displays to see, hear and interact with remote participants in 3D, almost as if they were present in the same physical space. Other studies were carried out to investigate presence in virtual environments resorting to avatar representations. Piumsomboon et al. [38] developed the Mini-Me, an adaptive avatar for enhancing remote collaboration between a local AR user and a remote VR user. The Mini-Me avatar reproduced the remote collaborator's gaze direction and body gestures while it transformed in size and orientation to stay within the AR user's field of view. Figure 2.6 shows the representation of remote people in state of the art approaches described above.

Previous work suggested that a full-body or upper-body representation of the remote people enhances awareness since the verbal communication can be complemented by gestures [39], deictics, [40] gaze direction [41] and eye contact [32]. In this dissertation, we follow state-of-the-art research and use a life-sized upper body representation. However, no depth cameras are used, and, instead, we track the user's movements with an Oculus Rift Set, as explained in Section 4.1.

2.4 Manipulation of Remote People Representations

Contrarily to what happens in the real world, it is possible in virtual environments to manipulate remote people's representations to achieve different goals. Piumsomboon et al. [38] found that manipulating the remote user's representation to guarantee that his gestures were always in the field of view of the local user enables a high sense of presence and increases task performance over an unmodified avatar. They modify the remote user's position and scale, and redirect its gestures and gaze direction through inverse kinematics.

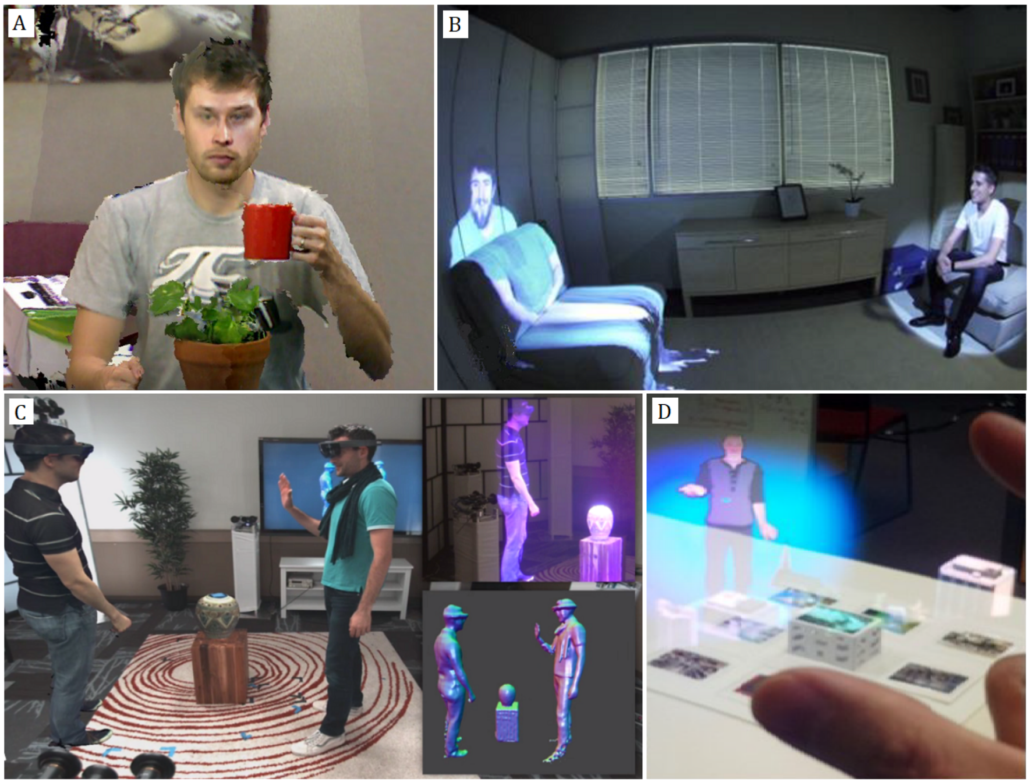


Figure 2.6: State of the art Presence approaches: [A] - Maimone et al. [35], [B] - Room2Room [4], [C] - Holoportation [37], [D] - Mini-Me [38].

With Negative Space, Sousa et al. [9] studied different manipulations of remote people and workspaces, mirroring one or the other to enable a shared point of view for the users. Results suggested remote collaboration benefits more from workspace consistency rather than people's representation fidelity. Hoppe et al. [42] also manipulate virtual people's representation to provide a shared point of view, finding that modifications of the workspace and the user's avatar to induce shared perspective reduces mental load and increases task performance. However, their virtual representations only show head and hands position.

Roth et al. [43] manipulated people's avatars by injecting artificial mimicry in their virtual representations. Their studies compared a negotiation task under a condition where virtual characters replicated only the original behavior with a condition where characters displayed the original behavior plus induced mimicry, finding that most participants did not detect the modifications. Additionally, manipulations did not impact the perception of communication.

Previous work studied manipulations of remote people's representation, concluding they can increase presence and task performance. Zibrek et al. [44] findings suggest, however, that manipulations should rely on realistic behaviors.

In this dissertation, we follow [9] and [42] approaches in exploring remote people's representation manipulations for perspective sharing. Similarly to Piumsomboon et al. [38], we apply inverse kinematics when redirecting the remote avatar's gestures.

2.5 Discussion

Throughout this section, we analyzed several state-of-the-art works. Table 2.1 shows our proposed classification of the most relevant approaches on collaboration in shared workspaces, reviewed in terms of collaboration metaphor, mechanisms enabled for workspace awareness, and type of artifact (2D or 3D) supported in the shared workspace.

From Table 2.1, one can see that the [5], [22], [20] and [28] approaches are successful in maintaining workspace awareness, as they allow all three typed of workspace awareness mechanisms. However, non of them supports three-dimensional artifacts. [26], [27] and [36] support 3D artifacts, ensuring good mechanisms of feedthrough and consequential communication. However, they have a limited capacity to ensure intentional communicational cues as they do not prevent occlusion issues.

We can say that, for all we know, there is no approach for remote face-to-face object-centered col-

Approach	Collaboration Metaphor	Artifact Type	Workspace Awareness		
			Feedthrough	Consequential Communication	Intentional Communication
Collaboard [21]	Whiteboard	2D	Available	Limited	Available
3DBoard [5]	Whiteboard	2D	**Available**	**Available**	**Available**
ImmerseBoard [22]	Whiteboard	2D	**Available**	**Available**	**Available**
KinectArms [23]	Table	2D	Available	N/A	Available
Illumishare [24]	Table	2D	Available	Limited	Limited
BeThere [25]	Table	Physical 3D	Available	N/A	Limited
MirageTable [26]	Table	Virtual & Physical 3D	Available	Available	Limited
Leithinger et al. [27]	Table	Physical 3D	Available	Available	Limited
Hope et al. [42]	Table	2D	Available	Limited	Available
Negative Space [9]	Table	Virtual 3D	Limited	Available	Limited
Beck et al. [36]	GlassWindow	Physical 3D	Available	Available	Limited
ClearBoard [20]	GlassWindow	2D	**Available**	**Available**	**Available**
FacingBoard2 [28]	GlassWindow	2D	**Available**	**Available**	**Available**
Shadowhands [29]	GlassWindow	2D	Available	N/A	Available
MAGIC	Table	**Virtual 3D**	**Available**	**Available**	**Available**

Table 2.1: Classification of the most relevant approaches for remote object centered collaboration in shared workspaces. * N/A indicates the feature is not available.

laboration with 3D content that allows all three types of workspace awareness mechanisms.

Similarly to [26], [27] and [36], we opted to follow the over the table metaphor as we believe we can tackle the challenge of mismatched views of the task space and occlusions by enabling a share of perspective between collaborators. Following the idea that integrating the person and the task spaces is key to enable intentional and consequential communication mechanisms [19], we explore the idea of a face-to-face setting with a shared perspective for improved workspace awareness and task performance. To enable a face-to-face meeting and a shared perspective at the same time, we resort to manipulations of the remote person's representation, as supported by [9] and [42].

3

MAGIC

Contents

Now that we fully understand the context in which our research is inserted, we can present our virtual collaboration approach in shared 3D spaces. MAGIC is a collaboration tool where two people can interact with each other in a virtual environment that allows them to analyze a given 3D object from the same point-of-view without losing the communicational advantages of a face-to-face arrangement. Following the assumptions of Gutwin and Greenberg [10] that peripheral tasks bring additional efforts in maintaining collaboration, we propose to diminish the shift of attention between the person space and the task space by creating a virtual environment where the two are integrated, improving workspace awareness. MAGIC avoids occlusions in the workspace by enabling its users to share perspective at all times while maintaining possible the observation of the remote collaborator's gestures by rendering its virtual replica in front of the local user. We focused this work on collaboration for two people only, as this is the simplest case, although some aspects of our work may extend to multiple people.

We begin this section by presenting a usage scenario for MAGIC, where our approach is used to facilitate the analysis of the 3D model of a jet engine during a collaborative meeting between two aerospace engineers. We further provide a detailed analysis of the impact of our work in workspace awareness, followed by an overview of the manipulation techniques employed to the remote collaborator's representation to enable the sharing of perspective while standing in a face-to-face setting.

3.1 MAGIC Usage Scenario: A Jet Engine Design Review Meeting

Kate and George, two aerospace engineers, are engaged in a remote meeting to discuss a jet engine's development. George is developing the main engine while Kate is in charge of designing a Secondary Air System (SAS), and wants to show George her final design. Since they cannot have a face-to-face meeting, as they work for a multinational company and are currently in different countries, they scheduled a virtual meeting powered by the MAGIC approach. They meet facing each other across a virtual environment that includes life-size depictions of their avatars in front of a worktable. George loads the CAD model of the project he has been working on, and the main system of the engine model appears in the shared workspace between them, above the worktable.

Halfway through the meeting, George points to direct Kate's attention to a change in the turbine configuration he wants to try. Kate notices, however, that the changes in the turbine will conflict with the simulations she ran indicating the best position for the SAS in the main engine.

Despite being face-to-face, both share the same perspective of the engine 3D model, and, using pointing gestures, Kate can indicate the optimal position she obtained for the SAS she is designing so that George takes that in consideration in the finalization of the main engine design. Figure 3.1 illustrates the remote meeting under a traditional face-to-face versus the remote meeting powered by MAGIC.

(a) Face-to-face Traditional Remote Meeting: Kate uses deictic gestures to indicate the optimal position for the SAS she is designing. As she and George have different perspectives of the workspace, the location she is pointing to is occluded. George does not understand the location Kate is referring to, and she needs to explain it through words, which brings added complexity to the collaboration task.

(b) Face-to-face Remote Meeting powered by MAGIC: Kate uses deictic gestures to indicate the optimal position for the SAS she is designing. As both share the same perspective of the model, George understands the location she is pointing to and can keep designing the main engine with a clearer view of the system.

Figure 3.1: Comparison between a Traditional Face-to-Face remote meeting and a MAGIC remote meeting.

3.2 Improving Workspace Awareness

Following Ishii et al. 's [20] metaphors for how to communicate in a shared environment without separation between the task space and the person space, MAGIC assumes an "above-the-table" metaphor with life-size virtual representations of remote people.

In a face-to-face setting, participants can see each other face to face, being able to perceive physical gestures besides maintaining verbal communication, which contributes to a good sense of presence and increases awareness of other people's actions. Since the work happens above the same table, accessible by both participants, we consider to be facing a shared task space with a shared reference space, as it is possible to use deictic references and pointing gestures to interact with the virtual objects in the workspace. This scenario appears to be ideal if it were not for the occlusions that can happen in the workspace, reducing its awareness.

To tackle this problem, MAGIC guarantees that the two participants share the same point of view at all times, preventing misunderstandings. However, to allow for a face-to-face setting with a share of perspective, we cannot simply render the remote collaborator in front of the local user, as this turns the remote collaborator's gestures obsolete. To maintain the same reference space, MAGIC manipulates the remote person's representation so that there is a matching between the location the local observer perceives and the location the remote person is referring to. This is essential for tasks that involve actions such as communicating specific features using deictic gestures.

Therefore, we believe MAGIC greatly contributes to the increase of workspace awareness in object-centered three-dimensional collaboration, as it integrates person, task and reference spaces, enabling collaborators to thoroughly express and perceive consequential communication, feedthrough, and intentional communication.

- *Consequential Communication*

  Since MAGIC adopts a life-sized face-to-face arrangement, the remote collaborator is always visible, in a similar way to what happens in real life. Therefore, the local person can observe his posture, body language, and eye movement at all times, which brings a significant amount of information about what is happening in the workspace. Contrary to what would happen in a situation where two people are collaborating using video-stream, the need to oscillate the focus of attention between the person space and the task space is reduced (requiring fewer eye-movements and fewer body rotations away from the workspace).

- *Feedthrough*

  When the artifacts in the workspace are manipulated, they give feedback to both the person performing the action and the viewer. Thus, sharing the same perspective of the workspace allows object manipulations to remain identical in terms of position and orientation in both the local and remote workspaces, ensuring the feedthrough mechanism is achieved.

- *Intentional Communication*

  Once the remote person's gestures are correctly converted to the local reference space, deictic references and gestural demonstrations can be used safely, guaranteeing that the two collaborators always share the same understanding of the workspace.

3.3 Manipulation of virtual people's representations

To combine the advantages of both being face-to-face and sharing perspective, both users stand on the same location, sharing the same point of view of the 3D model, but see a manipulated version of their partner.

In Figure 3.2, we illustrate the different steps involved in implementing the manipulations that enable our approach.

A) A collaborator points to an area of interest in Location A;

B) In Location B, without any manipulation, the gesture of the remote collaborator highlights a different area of the 3D model;

C) Similarly to Clearboard [20] and 3D-Board [5], we introduce a *mirror-reversal* of the remote person's representation. With the mirroring put in place, the local observer can correctly understand the remote person's interactions in the workspace's left-right axis. There is a common understanding of right and left and, body language and gaze direction match horizontally with the local reference space. However, since we are working in a three-dimensional workspace, the mirroring is insufficient, and depth needs to be corrected.

D) *Re-positioning of the pointing* arm along the forward-backward axis to correct for the depth of the interaction. It is to be noticed that when changing the end position of the pointing finger, all the arm has to move accordingly, creating changes at the level of the upper arm, the forearm, and the wrist. These changes are calculated through an Inverse Kinematics algorithm.

E) *Adjustment of the embodiment's position*. If the remote person stands in a position relative to the worktable that requires his avatar to stretch the arm to match the local space's pointing position, the remote person's avatar is moved further along the forward-backward axis. An arm that is too long could make it weird for the local observer, breaking the immersion sensation.

To evaluate the validity of our approach, we developed a prototype where the previous manipulations could be applied. We describe the prototype's implementation process in the following section.

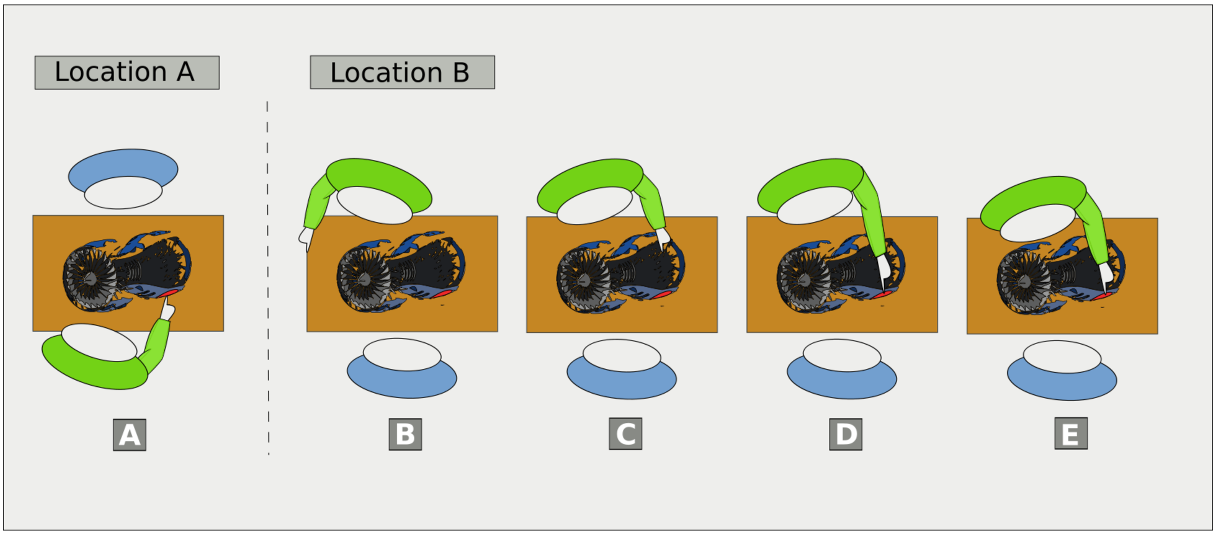


Figure 3.2: Illustration of the different steps involved in implementing our approach.

4

Implementation

Contents

In order to create a virtual environment that would allow us to study the approach idealized in Chapter 3, we developed a prototype using Unity, a cross-platform game engine developed by Unity Technologies[1]. We chose Unity over Unreal Engine [2] because of its fast prototyping ability to develop quality virtual and interactive systems.

Our prototype combines both collaborators' virtual spaces to allow them to work together, sharing the same worktable in a face-to-face formation. The prototype features an abstract avatar to embody each collaborator, shown in Figure 4.1. We choose a far from detailed representation to fit different genders and body-types and focus on nonverbal communication cues.

Our prototype involves three main modules: avatars' animation from user tracking motion capture; transmission of local data to the remote collaborator's space; and manipulation of the remote avatar embodiment in the local space to enable face-to-face perspective sharing.

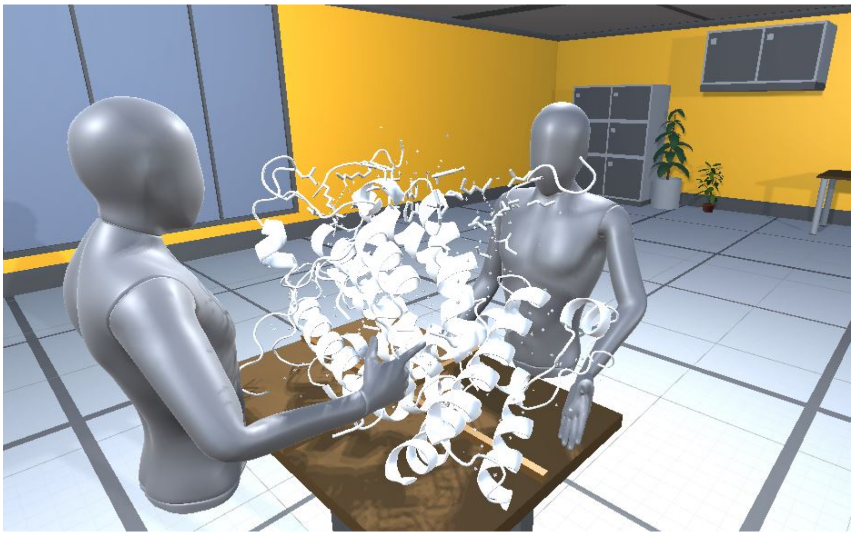


Figure 4.1: Virtual environment from our prototype, featuring two collaborators and the worktable with an abstract model used for the task.

4.1 Tracking and Avatar Animation

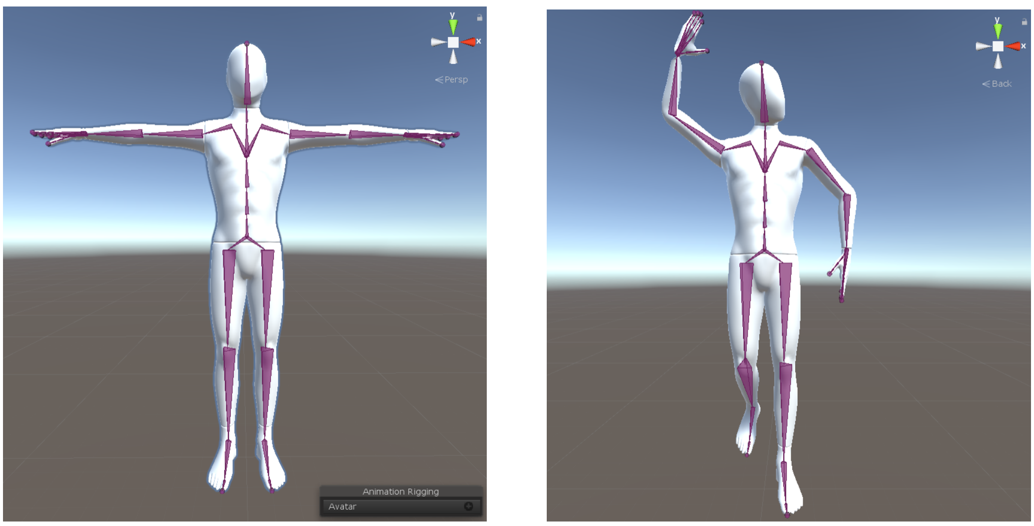
In order to allow users to see the virtual space and track their movements, we opted to use the Oculus Rift VR hardware package, composed of an Head Mounted Display (HMD), two touch controllers and two optical sensors (see Figure 4.2). We chose the Oculus Rift Set over the HTC Vive Set as both offered the same features - resolution of 1080×1200 per eye, 110-degree field of view, 90Hz refresh rate, and built-in motion sensors that work with external sensors to track movement - with the Rift Set being less expensive and more comfortable to use.

[1]Unity 2019.2.12: https://unity.com/releases/2019-2
[2]Unreal Engine: https://www.unrealengine.com/en-US/

We used the inbuilt Constellation Tracking System [45] from the Oculus sensors to track the users' heads and hands. A series of infrared LEDs embedded in the HMD and the touch controllers are monitored by the two optical sensors, which allows our prototype to keep track of the user's head and hands' position and rotation at all times and establishes a connection between the virtual and the real worlds. This information was used to animate an abstract avatar model, which acted as the user's embodiment once he entered the virtual space. The avatar[3] used to represent the users' embodiment was provided by *Mixamo*, a 3D computer graphics technology company that develops high-quality 3D characters. The avatar model allows access to its skeleton model (see Figure 4.3a), and all the joints that compose it can be manipulated through their rotation in order to achieve any position desired (see Figure 4.3b).

The matching between the VR hardware and the avatar's skeleton is made in the *VRRig* class. There we calculate the difference in position between the HMD and the avatar's body position (*headBodyOff-*

Figure 4.2: Oculus Rift Hardware: A - HMD; B - Touch Controllers; C - Position Trackers.



(a) Original Mixamo pose. **(b)** Manipulated pose.

Figure 4.3: Manipulation of the avatar joints.

[3]Mixamo Mannequin Character: https://www.mixamo.com/#/?page=1&query=mannequin&type=Character

set) so that we can make the body follow the HMD position, and attribute the HMD and touch controllers' position and rotation to the skeleton's head and hands.

To access the position and orientation properties of the VR hardware, we imported the *Oculus Integration Package* [4], that acts as a source for core VR features, components, scripts, and plugins that ease the Oculus app development process in Unity. This package contains different prefabs, including the *OVRCameraRig*, which provides the main interface to the VR hardware. In its turn, the *OVRCameraRig* contains a "tracking space" Game Object where we can access the position and orientation of the centre of the HMD (*CenterEyeAnchor*) and of the touch controllers (*RightHandAnchor* and *LeftHandAnchor*).

Following Zibrek et al. [44] findings suggesting manipulations should rely on realistic behaviors, we animated the avatars in a way that replicated the users' movement as closely as possible. Therefore, the avatars' arms had to move accordingly with his hands and head. For this, we used the *Chain IK constraint* available in the *UNITY Animation Rigging Package* [5]. This constraint implements the Forward And Backward Reaching Inverse Kinematic (FABRIK) solver [46, 47], an inverse kinematics algorithm that allows determining an appropriate set of configurations for different joints so that the end-effector moves to the desired position as smoothly and accurately as possible. Thus, it is possible to restrict the arm joints' position (shoulder, elbow, and wrist) so that the avatar's hand is rendered in the Touch Controller position with an adequate arm movement accompanying the change in position. We opted to do an upper body representation since there is no information on how the legs and feet are moving. Figure 4.4 shows how the avatar mimics different user positions based only on his head and hands position.

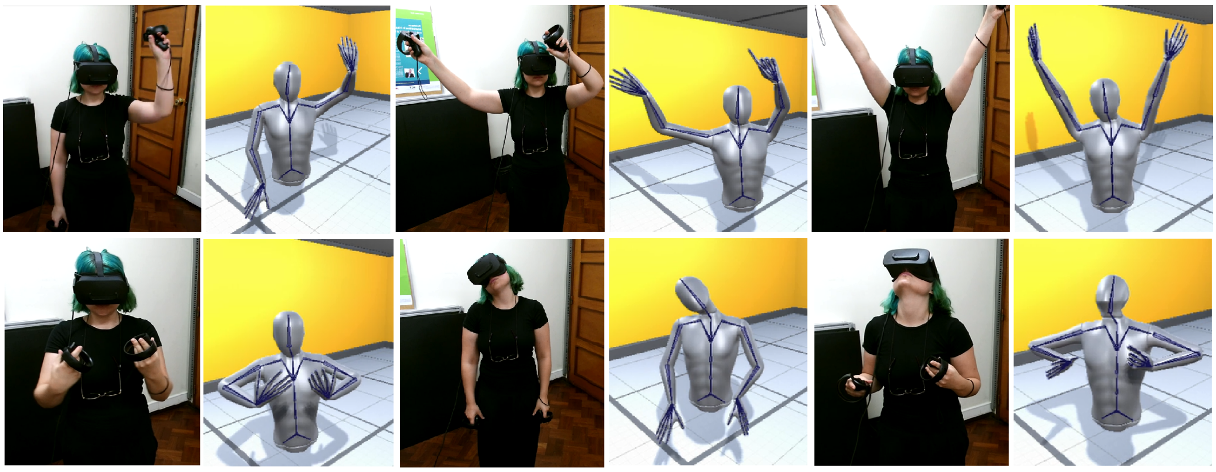


Figure 4.4: Illustration of the avatar representation corresponding to different user positions, based only on the user's head and hands positions with employment of the inverse kinematics algorithm.

[4] Oculus Integration Package [oculusintegration.unitypackage]: https://developer.oculus.com/downloads/package/unity-integration/

[5] Animation Rigging Package [com.unity.animation.rigging]: https://docs.unity3d.com/Packages/com.unity.animation.rigging@0.2/manual/index.html

4.2 Data Transmission

For the local collaborator to be rendered in front of the remote collaborator, data about the local user's VR hardware information had to be transmitted to the remote space.

To transmit the data, we studied different communication protocols - Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) - to find the one that better suited our needs, opting for the use of UDP.

Both protocols have different ways of handling data packet transmission over the network: while TCP establishes a connection and transmits packets in a reliable, ordered, and error-checked manner, UDP is a simpler protocol that does not need to establish a connection and does not guarantee data integrity. The main differences between TCP and UDP can be found in table C.1 from Appendix C.

Comparing both protocols, and bearing in mind our application is very time-sensitive, a TCP communication protocol was not appropriate since it induces latency by guaranteeing the sending of all information packets, stopping the streaming of data for error handling and packet retransmission in case of loss. In real-time applications such as our own, the loss of one packet is not problematic as it will be replaced as soon as the next arrives.

The message we sent through our UDP network included information on the different body parts of the local person's avatar that were needed to animate the person's avatar in the remote space: head, right hand, left hand, right fingertip and left fingertip. To codify the message, we use different characters as message separators:

- ' $\#$' for the different body parts (1 to 5 in the order stated above)
- '/' for the different properties of each body part (position and orientation)
- ':' for the property coordinates in the different axis (p_x, p_y, p_z for the position; r_x, r_y, r_z, r_w for the orientation)

The structure of the message can be seen in Figure 4.5:

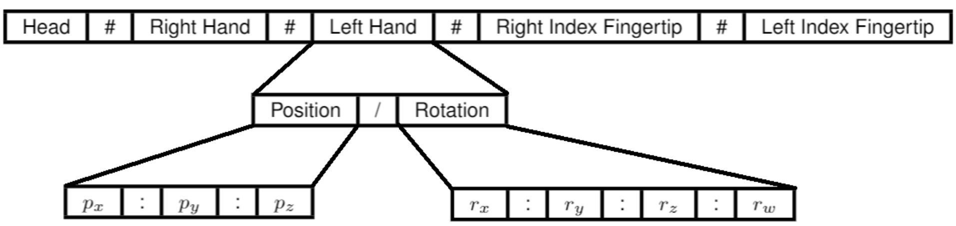


Figure 4.5: Structure of the message sent through the network.

The *SendAvatar* class accesses the information on the five body parts we need to send to the remote space, proceeds to the codification of the message, and instantiates a *UdpBroadcast* object that will be responsible for creating an endpoint at port 7001 for sending the message to the other machine. The *UdpListener* class listens at the same port (port 7001) and proceeds to decode the received message and save the information in the respective variables in the remote space. Once this information is ready, we can proceed to its manipulation in order to guarantee that the remote person is correctly rendered in the remote space, in front of the local user.

4.3 Movement Manipulation

Once the information required to create a representation of the remote collaborator in the local space is received, it has to be manipulated so that the remote collaborator can be represented in front of the local user in a way that guarantees that the gestures seen by the local observer are representative of what the remote collaborator is doing in his own local space.

Let us refer to the remote collaborator's representation as to the remote avatar. From now on, let us focus on the remote collaborator and his avatar in the local collaborator's space. The manipulation of the remote avatar is divided into three phases, as stated in section 3.3, and happens in the *Workspace-Transformation* class.

- *Mirroring*

  The first step of the manipulation is the mirroring of the remote person's avatar. The remote workspace is obtained from the local workspace by doing a 180º rotation, followed by an inversion along the x axis using the scale properties of Unity. Then, we start the process of treating the remote data coming from the network. To achieve a mirrored avatar, the remote head is matched directly to the avatar without further manipulation, while we attribute the properties from the right hand (and index fingertip) to the left avatar hand, and vice versa.

- *Repositioning the pointing arm*

  To guarantee the remote avatar's hands point to the same location as the remote person's hands in his local space, we reposition the pointing arm. The transformation vector, $\vec{t}$, is calculated from the difference in position from the remote collaborator's fingertips in his local space and the remote avatar's fingertips after the mirroring process (Equation (4.1)). Once $\vec{t}$ is calculated, we can apply it to the remote avatar's wrist (Equation (4.2)), and obtain the avatar's hand in the correct position, as depicted in Figure 4.6.

$$\vec{t} = rf_p - af_p \tag{4.1}$$

$$w_p = w_{p_0} + \vec{t} \tag{4.2}$$

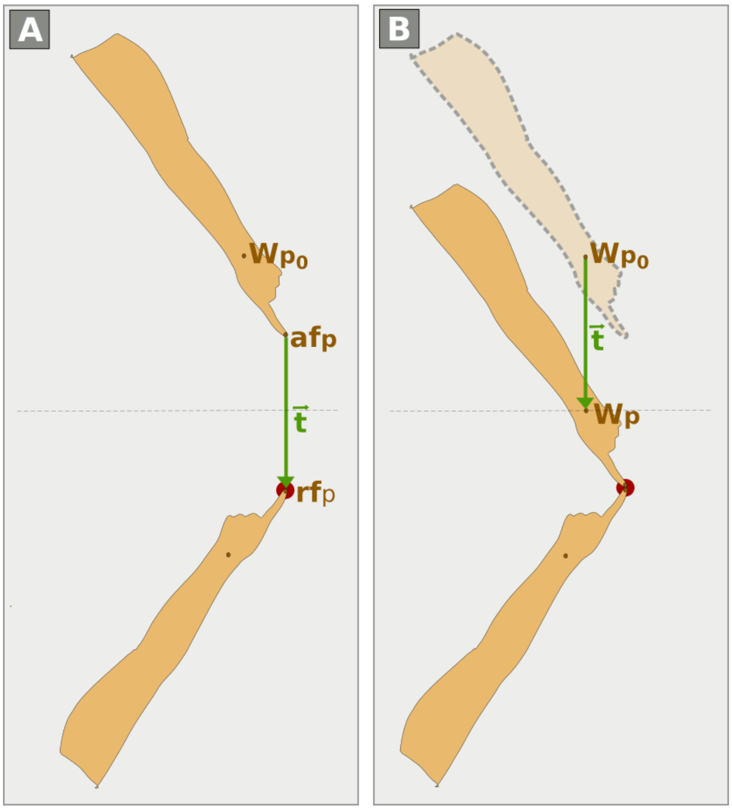


Figure 4.6: Repositioning of the pointing arm: arm position before [A] and after [B] manipulation. Transformation vector $\vec{t}$ is applied to the remote avatar's wrist so that the pointing arm matches the local reference space

Where rf_p and af_p are the positions of the remote collaborator and remote avatar index fingertips, respectively, and w_{p_0} and w_p are the positions of the avatar's wrist before and after the transformation. All positions are three-dimensional. Note that only the hand that is pointing suffers this transformation.

- *Adjust embodiment's position*

  At this point, we have a mirrored remote avatar, with hands that match the local reference space. However, there are scenarios where the avatar's arm would have to stretch beyond its length in order for the remote avatar to reach its collaborator's intended position. To avoid such a stretch, the avatar's position is adjusted in depth (only the z component). We considered two limit points:

  – The remote collaborator is pointing to a position near his edge of the table: in this scenario, the remote avatar should be rendered the closest to his table's edge to avoid stretching.

  – The remote collaborator is pointing to a position the farthest across his edge of the table: in this scenario, the remote avatar should keep a minimum distance from the table edge. It would be awkward to render the remote avatar close to the edge of the table and to have to bend his arm to an extreme for him to reach the desired position when, in real life, if a person wants to point to something near their edge of the table, they tend to step back and point in a comfortable position for the arm.

Since the remote avatar's position is controlled by it's head position, we correct its head, and the body will move accordingly. As we wanted to make the avatar's position change based on the remote collaborator's index fingertip position, we developed a linear model where the head position is computed as a function of the fingertip position. Then, the transformation of the avatar's position according to the location the remote collaborator's pointing finger follows the linear equation (4.3):

$$ah_{p_z} = m \cdot rf_{p_z} + b \tag{4.3}$$

Where ah_{p_z} is the avatar's head position along the z axis, and rf_{p_z} is the remote collaborator's index fingertip position along the z axis, and $m = -0.823529412$ and $b = -0.26$. We determined m and b based on the limit points considered.

Once the manipulation is complete, the remote avatar can be rendered in front of the local collaborator's avatar without losing the ability to interact with the workspace in a way that keeps awareness of the task for both participants.

4.4 Architecture Overview

The combination of the three modules (Avatar animation, Data Transmission and Movement Manipulation) enables the creation of a functional prototype were two users located in different spaces can interact face to face with three-dimensional objects while sharing the perspective of the workspace. Figure 4.7 presents the overall architecture of the system:

As soon as our prototype was ready to use, we moved on to designing an experimental protocol that would allow us to evaluate our approach.

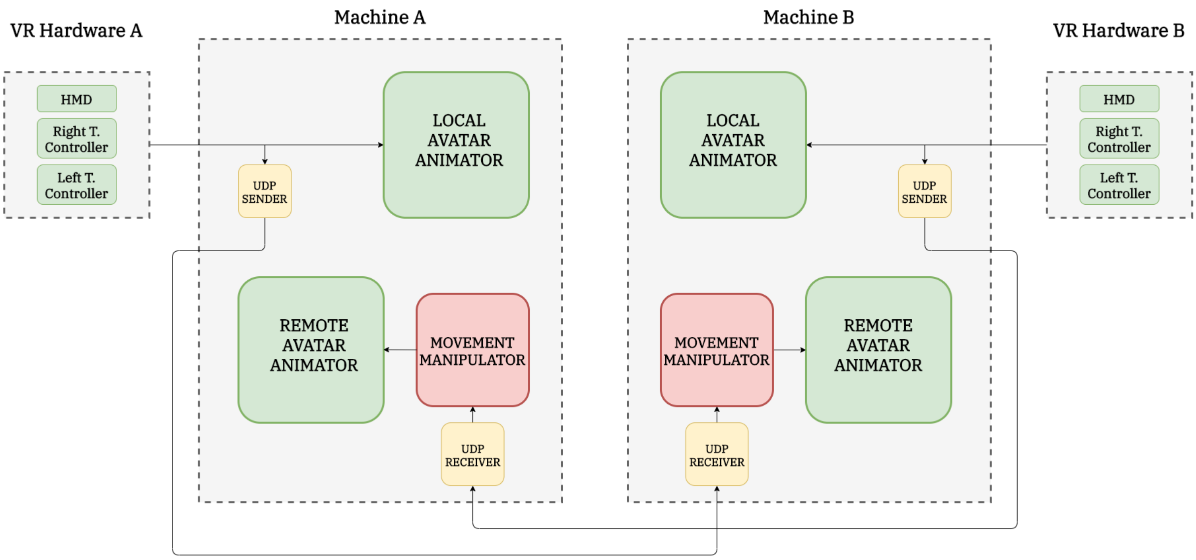


Figure 4.7: Architecture Overview.

5

User Evaluation

Contents

To determine if our approach improves face-to-face collaboration in 3D object-centered remote collaboration, we conducted a user study with 10 participants.

The user study's main goal was to check whether people can maintain a shared understanding of the workspace while interacting in a face-to-face formation and sharing the same perspective of the workspace at the same time.

In this chapter, we offer a detailed overview of our experiment's design, the procedure followed, and the evaluation metrics used to evaluate our User Interface (UI). Furthermore, we report the experiment's results along with a discussion of our findings.

5.1 Design

The user evaluation aimed to show that when two remote collaborators work together with three-dimensional models in a face to face setting, if a collaborator indicates a specific area of interest in the shared workspace, the observing collaborator can identify it better if both of them share the same point of view. Additionally, we wanted to find out if the manipulations employed to the remote collaborator's representation to enable the shared perspective affected the feeling of co-presence between users. We carried out a user study where pairs of participants were asked to complete a collaborative "matching" task on an abstract 3D model (see Figure 5.1), under two different conditions:

- *Face-to-face with our approach* (**MAGIC**) – Participants were on opposite sides of the model but shared the same point of view of the workspace.
- *Veridical face-to-face* (**Veridical**) – Participants were on opposite sides of the model with opposing points of view of the workspace, as in real life face-to-face interactions.

In the MAGIC condition, participants stood face-to-face, however, both had the same point of view of the workspace, with their representations manipulated in order to guarantee a common understanding. The Veridical condition worked as a control condition, where we reproduced a typical face-to-face

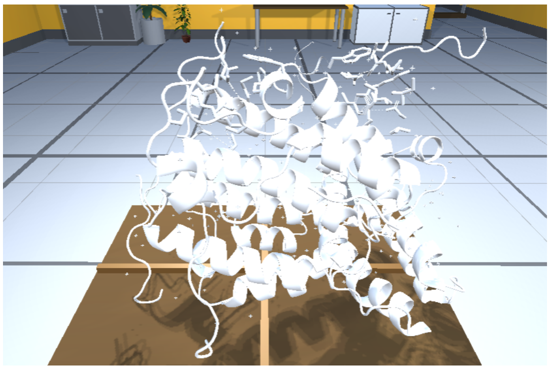


Figure 5.1: Abstract 3D model used for the purpose of prototype evaluation.

interaction (opposing points of view of the workspace for each participant). These two conditions were designed in order to allow us to compare MAGIC interactions with traditional face-to-face collaboration.

Each pair of participants experienced under both conditions, while they remained unaware of the condition under which they were performing.

Our main goal was to evaluate intentional communication. However, participants could use verbal communication as an additional collaboration tool, for collaboration to be as close to real-life interactions as possible. Since our experimental design intentionally employed an abstract 3D rendering, we hypothesize that talking wouldn't affect the evaluation as describing the targets through words was very difficult for participants. We observed that participants performed the task using deictic gestures and used verbal cues to keep track of the partner's understanding of performed actions. With sentences like "This is the area I want to show you", "Are you understanding where I'm pointing to?" or "Yes, I get what you are drawing my attention to!".

In the beginning of the experiment, each participant was assigned a different role from the two following: *Demonstrator* and *Interpreter*:

- The Demonstrator can see a red highlighted area of interest in the workspace. His 'job' is to communicate that area to the Interpreter by outlining it with his pointing finger.
- The Interpreter cannot see the highlight. Following the Demonstrator's gestures only, the Interpreter uses one of his pointing fingers to outline the interpreted area.

With one participant assuming the role of Demonstrator and the other assuming the Interpreter's role, the two participants were asked to execute a series of matching tasks, starting in a given condition (MAGIC or Veridical).

As the experiment began, a starting canvas (see Figure 5.2a) appeared on top of the empty worktable. To make the workspace appear and start each matching task, the Demonstrator was asked to press the Index Trigger Button from one the Oculus Touch Controller (see Figure 5.2b).

Once the workspace appeared, a small part of the model on top of the table would change color to red, and the user had to communicate that red target area to his partner. To do so, the Demonstrator was asked to outline it with his finger while pressing the A or X Buttons from the Touch Controllers, depending on the hand he chose to use, which would leave a green trail in the area outlined (see Figure 5.6c). The outline stopped once the button was released. The controls are shown in Figure 5.3.

To highlight different parts of the abstract model, we developed a shader[1] that features a "target sphere" that, when applied to an object, has the ability to change the parts of the object it intersects with to red (Figure 5.4). We saved different positions of the "sphere target" in a text file and gradually accessed it to move the "sphere target" and highlight different parts of the model.

[1] *"Small script that contains the algorithms for calculating the color of each pixel rendered, based on the lighting input and the Material configuration"* (from `https://docs.unity3d.com/Manual/Shaders.html`).

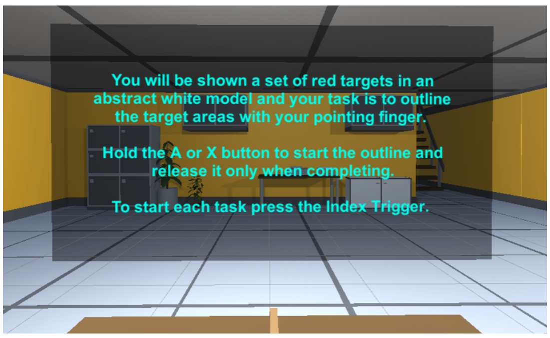
The Demonstrator was shown a set of 16 targets to communicate to his partner, one a time. After completing the 16 matching tasks, roles were switched, and the former Interpreter assumed the role of Demonstrator, communicating a new set of 16 targets to his partner.

5.2 Procedure

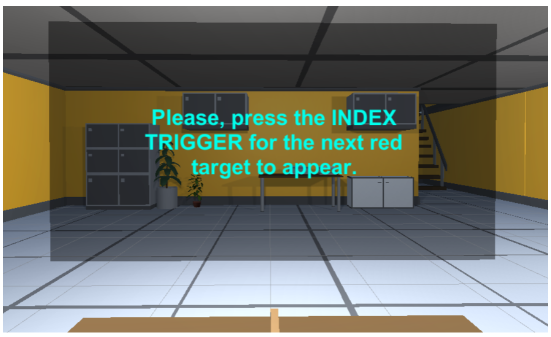
All evaluation sessions followed the same structure and lasted for about 50 minutes. Apart from the initial COVID-19 safety measures (See Section 5.3), participants started by fulfilling a profile questionnaire and a consent form. They were then introduced to the evaluation, where we explained conditions, tasks, and roles. Participants were able to pose any questions they might have on the course of the experiment. Before begging the experiment, participants took part in a training session that consisted of the execution of four practice matching tasks for familiarization with the environment and hardware. After, participants performed under both conditions. For each condition, upon completing the first 16 tasks, participants switched roles and executed the task again with a different set of 16 targets. After completion, they were asked to answer a questionnaire regarding that condition.

We can summarize the procedure as follows:

- **Initial COVID-19 safety measures:**
  - Hands and Face cleaning procedures;
  - Surgical mask placing;
- **Introduction:**
  - Experiment contextualization;
  - Question and answer session;
  - Filling of the Consent Form and Covid-19 Form;



(a) Canvas shown in the beginning of the session.



(b) Canvas shown between each matching task.

Figure 5.2: Different canvas shown during the session.

Figure 5.3: Oculus Touch Controllers' Buttons used by the application.

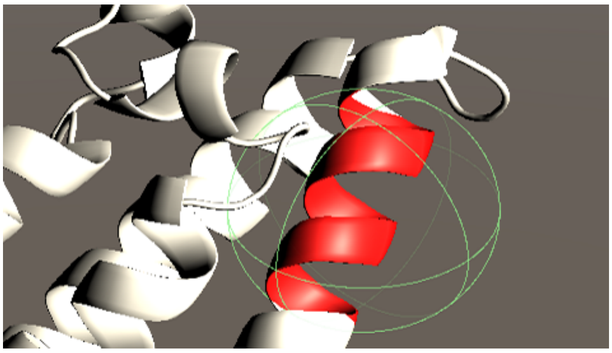


Figure 5.4: Shader's "Sphere Target" changing the color of the abstract model's points that it intersects with to red.

  - Filling of the Profile Questionnaire;

- **Training session:**
  - Teaching of how to put the HMDs and how to handle the controllers;
  - Execution of four practise matching tasks for familiarization with the environment;

- **Beginning of experiment:**
  - One participant assumes the role of Interpreter and the other of Demonstrator and execute a series of 32 matching tasks in a given condition - Condition 1 (switching roles after the 16^{th} matching task);
  - Participants fill the User Preference Questionnaire for condition 1;
  - Participants repeat the procedure under the second condition;
  - Participants fill the User Preference Questionnaire for condition 2;

- **Disinfection of Hardware:**
  - Controllers and HMD
  - Desktop mice and keyboards
  - Pens

As subjects performed a similar task in more than one condition, there was a possibility of *carry-over effects* - the lingering effects of a previous experimental condition that are affecting the current experimental condition. We created four different sets of sixteen targets - T1, T2, T3 and T4 - and counterbalanced them to avoid carryover effects. Each set of targets comprised similar targets that differed enough to guarantee that participants would not interpret them by memory from the precedent task, but were similar enough to ensure consistency of tasks between participants and conditions. We also counterbalanced the order of conditions between the two possible conditions (MAGIC and Veridical). Table 5.1 reports the order of tasks and tests assigned to each participant. Note that it was determined that participants A and B would always start as Demonstrator and Interpreter, respectively.

Group of Participants	ID A	ID B	1^{st} Condition	1^{st} Set of Targets	2^{st} Set of Targets	2^{st} Condition	1^{st} Set of Targets	2^{st} Set of Targets
1	1	2	V	T1	T2	M	T3	T4
2	3	4	M	T3	T4	V	T2	T1
3	5	6	V	T4	T1	M	T3	T2
4	7	8	M	T2	T3	V	T1	T4
5	9	10	V	T1	T2	M	T4	T3

Table 5.1: Distribution of Roles, Conditions and Sets of Targets for each pair of participants. M and V refer to the MAGIC and Veridical conditions, respectively.

5.3 COVID-19 Safety Procedures

In response to the current global circumstances caused by the Covid-19 (Coronavirus) outbreak, extra safety measures had to be undertaken to minimize the risk of contracting and spreading the virus during the experimental sessions. Taking in consideration advice issued by the World Health Organization (WHO), the Academy of International Extended Reality (AIXR) specifically created a set of basic guidelines for safety in the use of Virtual Augmented Reality hardware[2]that were followed during our user evaluation, along with some extra measures. The measures adopted can be consulted below:

1. **Follow rigorous hand cleaning procedures**: Before each user entered the experiment room and put on the VR devices, they were required to wash their hands with water and soap for 20 seconds. Additionally, participants were also asked to wash their faces with water and soap before putting on the HMD. Also, all staff had 70% alcohol-based hand sanitizer for use at all times.

[2] "Covid-19 Safety for Virtual & Augmented Reality AIXR Guidelines" (Accessed: 2020-07-09): https://aixr.org/press/articles/covid-19-safety-for-virtual-augmented-reality-aixr-guidelines/?fbclid=IwAR3m-WnzAkIYymDX2DRrK2ah08yHrR9F04KaaGELgxlndEc1K2SVlr-tEUo

2. **Mandatory use of mask**: After the hand and face cleaning procedures, every participant was given a surgical mask to use throughout the experiment.

3. **Disinfection of hand controllers and headset**: Between each utilization by different users, the Oculus Touch Controllers and HMD were wiped down using 70 % alcohol-based wipes.

4. **Disinfection of Desktop mice and keyboards**: For the purpose of questionnaire filling, a Desktop was used by each participant during the sessions. Between each utilization, the Desktop mice and keyboards were wiped down using 70 % alcohol-based wipes.

5. **It is the large gathering of people in general that increases risk**: A strict schedule was planned to guarantee that the participants would interact only with their evaluation partner and the experiment supervisor. Therefore, there was a maximum of three people in the experiment room, which guaranteed that social distancing was possible and respected the Instituto Superior Técnico (IST) 's room occupancy policy of 5 people for that room.

5.4 Metrics

When it comes to the evaluation metrics for our UI, we focused on the general metric of usability. A UI is usable *"when the user can reach its goals; when the goal tasks can be done better, easier, or faster than with another system; and when users do not get frustrated or uncomfortable"* [48]. We considered two types of metrics for 3D UIs to measure the usability of MAGIC: task performance metrics and user preference metrics.

User task performance refers to the quality of performance of specific tasks in the interface, and, typically, speed (efficiency) and accuracy are the most important task performance metrics. Therefore, we used the following two metrics to evaluate task performance:

- Completion time, measured for each task, as the elapsed time between the moment when the Demonstrator presses the Index Trigger button from one the Oculus Touch Controller making the workspace appear and the moment when the Interpreter concludes the outlining of the target area (by releasing the button chosen);
- Accuracy of the pointing task, measured as the percentage of intersection between the two areas of interest outlined by the participants. A detailed overview of how the percentage of intersection was measured can be found in Section 5.4.1);

User preference generally refers to the user's subjective perception of the interface and is often measured via questionnaires or interviews, and can either be qualitative or quantitative [48]. To evaluate user preference, we developed a User Preference Questionnaire, to be answered after completion of matching task under each condition. This questionnaire contained a set of statements regarding the feeling of co-presence between participants, attentional allocation, and perceived message understanding, to be answered on a 6-point Likert-scale. We chose a 6 point Likert scale since it does not possess a 'neutral' option and 'forces' choice. We also included an open question regarding whether participants could identify any strange behaviors to evaluate if participants noticed the manipulation we were employing to their remote partner's avatar. Additionally, we asked participants any suggestions they had to better the experiment plus any comment they would like to share with us. The User Preference Questionnaire we developed can be found in Appendix B.

Note that, during the experiment, we recorded other parameters for replicability: the set of points the participants used to outline each area of interest and body position (HMD and Touch controllers position) of both participants in each update.

5.4.1 Accuracy of the matching task

To measure the matching task's accuracy, we opted to calculate the intersection between the two matching zones of interest: the one the Demonstrator outlined and the one the Interpreter perceived. To find the percentage of intersection between the two zones, we opted to generate, for each participant, the three-dimensional solid that forms from the convex set comprising all the points that constitute the outline of the zone of interest. These solids were calculated using a QuickHull Algorithm [3].

The QuickHull algorithm is a method for computing the convex hull of a finite set of points in n-dimensional space, in our case, the 3-dimensional space. The convex hull of a set of points is the smallest convex set that contains it. The algorithm uses a divide and conquer approach and can be broken down into the steps that follow. Firstly, an initial hull is created by picking four (n + 1) points from the point cloud set that do not share a plane. The initial hull consists then of just a simple tetrahedron made up of those four points from the point cloud. The point cloud points are then divided into two parts, the "open set" and the "closed set". The open set consists of all the points outside of the tetrahedron, and the closed set consists of all the points already inside the tetrahedron. The initial hull is then grown until all the points in the point cloud are inside the closed set, and the open set is empty, at which point the algorithm is finished, and we have the Convex Hull (CH) for the entire set.

The QuickHull algorithm returns a list of vertices, triangles, and normals that can be directly converted into Unity mesh. A mesh consists of a set of triangles arranged in 3D space to create the impression of a solid object in our scene. Figure 5.5 illustrates the transformation of a given set of points in the respective CH.

[3] Quickhull Algorithm for Generating 3D Convex Hulls - an Implementation in Unity, Oskar Sigvardsson (Accessed: 2020-05-15): https://github.com/OskarSigvardsson/unity-quickhull

From the set of points that constituted the outline of the zone of interest for each participant, two CHs were created for each matching task repetition, the Demonstrator's CH and the Interpreter's CH, to which we will refer to as the Demonstrator's Solid, S_{Dem}, and the Interpreter's Solid, S_{Int}.

Once we obtained S_{Dem} and S_{Int}, we compute the intersection between these two meshes. To obtain the list of points that result from the intersection of S_{Dem} and S_{Int}, we implemented a Constructive Solid Geometry (CSG) Library[4]. CSG is a technique used in solid modeling that uses Boolean operations such as union and intersection to combine 3D solids.

The intersection function from the CSG library gives us a list of all the points that are common to both S_{Dem} and S_{Int} and, from this set of points, we can once again use the QuickHull algorithm to compute the Intersection mesh, S_I. Figure 5.6 shows the S_I obtained from the intersection of two solids S_{Dem} and S_{Int}:

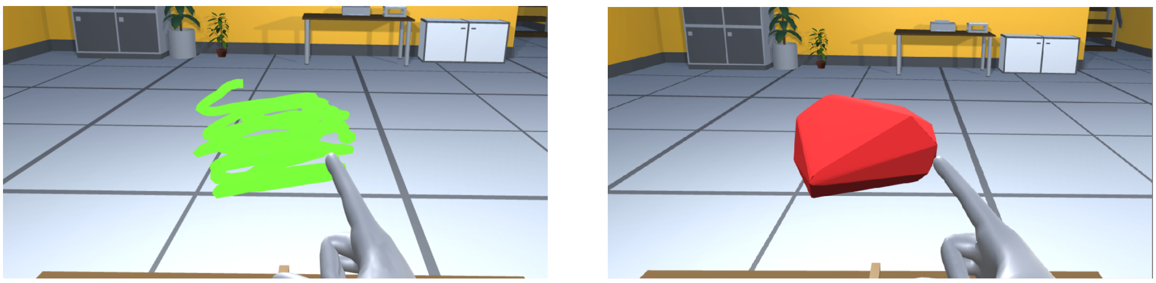
Once we have S_I , we can calculate its volume and compare it with S_{Dem} in order to compute a percentage of intersection, $I_{\%}$:

$$I_{\%} = \frac{V_{S_I}}{V_{S_{Dem}}} \cdot 100 \tag{5.1}$$

The idea behind calculating the volume of these meshes, V_{S_I} and $V_{S_{Dem}}$, is to calculate the volume for each tetrahedron of the mesh - based on each triangle that composes the mesh and topped off at the origin - and then add them up [49]. The base expression to calculate the volume of a tetrahedron, V_t, is:

$$V_t = \frac{1}{3}(A_{base}) \cdot h \tag{5.2}$$

Now, looking at Figure 5.7, where A, B and C are three vertices of a mesh triangle, and considering the vectors that form from the origin to those vertices, $\vec{A}$, $\vec{B}$ and $\vec{C}$, we see that the area of the base,

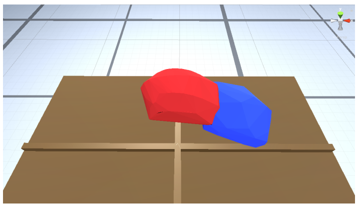


(a) Trail left by the user's finger when pressing the A or X Buttons.

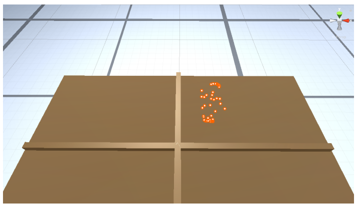
(b) Convex Hull generated by the QuickHull algorithm.

Figure 5.5: Transformation of a given set of points in the respective CH.

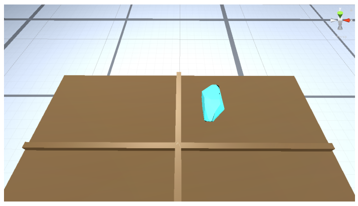
[4]Constructive Solid Geometry (CSG) for Unity in C#, Andrew Perry (2020-07-22): https://github.com/omgwtfgames/csg.cs/blob/master/Assets/CSG/Plugins/CSG/CSG.cs



(a) S_{Dem} in red and S_{Int} in blue, obtained from the employment of the QuickHull algorithm to the set of points used by the Demonstrator and the Interpreter's to outline the areas of interest.



(b) Intersection points returned by the CSG algorithm.



(c) S_I obtained from the QuickHull algorithm applyed to the list of points returned by the CSG algorithm.

Figure 5.6: S_I obtained from the intersection of two solids S_{Dem} and S_{Int}.

A_{base}, is half the area of the parallelogram found by taking half the cross product of the two vectors that represent the edges of the base, $\vec{A}$ and $\vec{B}$.

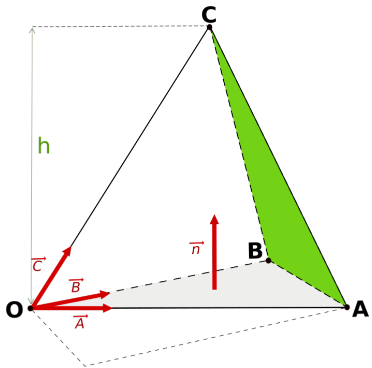


Figure 5.7: Volume of the tetrahedron that forms from a mesh triangle(green face) to the origin.

We can calculate the height, h, by doing the dot product between one of the vectors along the side of the tetrahedron, $\vec{C}$ and the perpendicular relative to the base - the triangle's normal, $\vec{n}$ - that results from the cross product of $\vec{A}$ and $\vec{B}$. From Equation 5.2, the tetrahedron's volume, V_t, expands to:

$$V_t = \frac{1}{3} \cdot \frac{1}{2} (\vec{A} \times \vec{B}) \cdot \vec{C} \tag{5.3}$$

$$V_t = \frac{1}{6} \mid \vec{A} \mid \mid \vec{B} \mid \vec{n} \cdot \vec{C} \tag{5.4}$$

Since these volumes are the result of the dot product $\vec{n} \cdot \vec{C}$, they are signed, and can be negative or positive depending on whether the $\vec{n}$ of the triangle in question is pointing in the direction of the origin or otherwise. Faces pointing out add to the total volume as $\vec{n} \cdot \vec{C} > 0$, and faces pointing in subtract to the total, as $\vec{n} \cdot \vec{C} < 0$. What is left is only the volume inside the object. To get the mesh's total volume, we go through each triangle, compute its signed volume, and add them up.

5.5 Evaluation Communication Protocol

For our experiment to be successful, we had to guarantee that information on the experiment's current stage was known both by the Demonstrator and the Interpreter's Machines. When the Demonstrator presses the Index Trigger making the workspace appear, information has to be sent to the Interpreter's machine so that the workspace appears for the Interpreter at the same time. Also, when the Interpreter finishes outlining the interpreted area, we needed to guarantee that all the points he used in his outline are sent to the Demonstrators machine in order to compute the Interpreter's CH and calculate its percentage of intersection with the Demonstrator's CH. Finally, the Demonstrator's Machine has to send information to the Interpreter's Machine when the 1st part of the session under a given condition ends (after conclusion of the first 16 matching tasks) and roles switch - the Demonstrator assumes the role of Interpreter and vice versa. At that point the Demonstrator's Machine becomes the Interpreter's Machine and inversely.

Since this data is crucial for the experiment to keep running matching task after matching task, we needed a reliable, ordered, and error-checked delivery of data. On the contrary of what happened with the streaming of the controller's data described in Section 4.2, we were no longer interested in speed, but in guaranteeing data integrity and reliability. Therefore, a new network relying on a TCP communication protocol was implemented.

As usual in a TCP based network, the server first creates a socket bound to a specific IP address, in our case, the IP address of Machine A, which will always assume the role of server. The socket listens for a connection at port 9001, and waits until a client makes a connection request. Once it receives a request, the server accepts the connection, and data transmission can begin. The client, Machine B, behaves similarly to the server. First, a socket is created and is bound to an IP address in a similar fashion to the server. The client attempts to establish a connection with the server's IP address at the port the server is listening to.

Both the server class, *TCPServer*, and the client class, *TCPClient*, have sending and receiving methods since the information has to flow both ways.

The Demonstrator's Machine, A or B depending on the session state, sends two types of message in the form of a string:

- The value of the boolean *showWorkspace*, which indicates if the Interpreter's Machine should display the workspace (when *true*), or not (when *false*). The message is composed of a prefix that codifies the type of information, 'update', followed by the character separator '#' and the boolean value.
- The inverse value of the boolean *localIsDemonstrator*, which stays *true* while the Demonstrator is completing the experiment'part where he assumes that role, and changes to *false* when the

Demonstrator finishes his 16 tasks; the moment when the Interpreter becomes Demonstrator. The message is composed of a prefix that codifies the type of information, 'demonstrator', followed by the character separator '#' and the inverse of the boolean value. Therefore, the Interpreter assumes Demonstrator's role when his machine receives the message' Demonstrator #true'.

The Interpreter's Machine only sends the Demonstrator a string containing the set of trail points used by the Interpreter to outline the area of interest. The message is composed of the prefix 'points', followed by the character separator '' and the set of points in the same structure as Figure 4.5.

Figure 5.8 illustrates the flow of events during each matching task of the experiment:

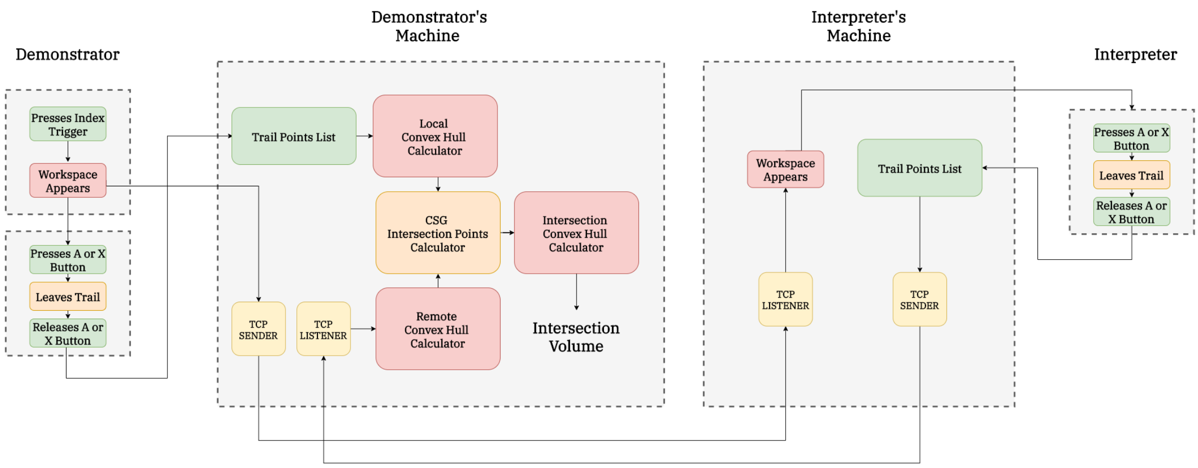
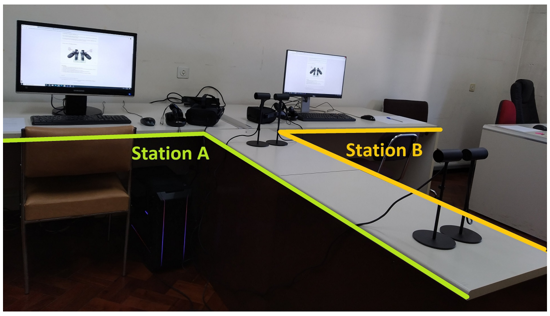


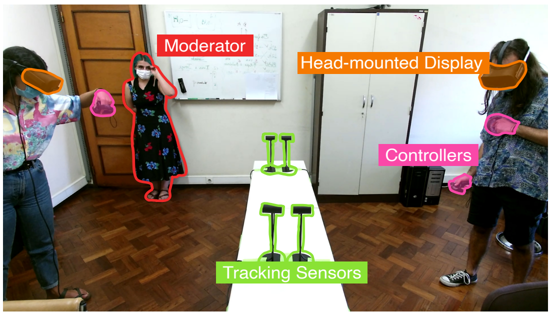
Figure 5.8: Flow of events during each matching task.

5.6 Participants

Our subject group included five pairs of participants (10 total, 3 female). Participants' ages ranged from 23 to 24 years (M = 23,3; SD = 0,67). All participants declared having University Education and knowing their partner. Only 1 participant reported using his left hand as a dominant hand. Two participants indicated using video-conferencing platforms, such as Skype, Google Hangouts, or Zoom, at least once a day, seven at least once a week, and one reported daily use. As for virtual reality environments, three participants reported they never used virtual reality environments, and the other seven participants reported rarely using this type of environment. Data on the participants' profiles was acquired through the User Profile Questionnaire available in Appendix A.



(a) Stations A and B in the experiment room.



(b) Occupation of the room during an experimental session: each participant in his station, monitored by the experiment supervisor. .

Figure 5.9: Physical Setup.

5.7 Apparatus and equipment

Our experiment took place in INESC-ID, Alameda - IST. The room where the evaluation took place had no contact with the exterior, occupied by the two participants and the experiment supervisor only (Figure 5.9b). The physical setup consisted of two separate stations, A and B. Each station was composed of a desktop running the participant's application and an Oculus Rift set (Oculus HMD, two Touch Controllers, and two position trackers), as displayed in 5.9a. A physical table could have been used for haptics purposes; however, we opted to simplify the physical set up to the maximum.

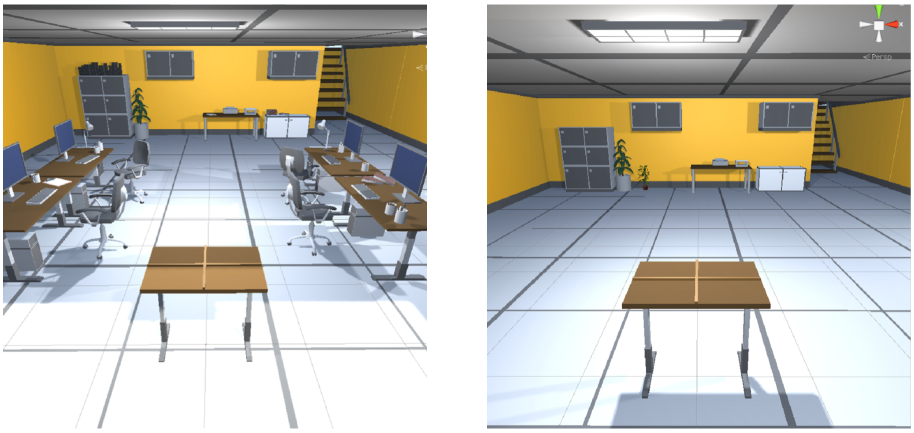
The virtual environment where the two users could engage in collaboration during the evaluation was developed using various elements from the *Snaps Prototype — Office Package*[5] and implemented a simple office room with a worktable in its center.

In a first approach, we designed an office with several workbenches, chairs, cabinets, and props such as computers, books, and lamps distributed around the room. However, it came to our attention that, during the experimental sessions, all the small details could increase distractions and, therefore, we opted to remove most of the unnecessary elements so that the users would only focus on the worktable during the user study. Both the initial and final designs can be seen in Figure 5.10.

5.8 Results

Throughout the evaluation experiments, we collected information on Task Performance using two different metrics - completion time and accuracy - and gathered User Preferences information from questionnaires filled up after the execution of the tasks correspondent to each condition (as described in Section 5.4).

[5]Snaps Prototype — Office Package V1.2: https://assetstore.unity.com/packages/3d/environments/snaps-prototype-office-137490



(a) Initial virtual environment. **(b)** Final virtual environment.

Figure 5.10: Different environments developed for MAGIC testing.

To analyze the gathered data on task performance under both different conditions, we first used the Shapiro-Wilk test [6] to check for data normality. The null-hypothesis of this test, H_0, is that the population is normally distributed. In order to determine statistical significance, a cutoff value, alpha, is decided by the researcher, usually takes on a value of .05 or less, which corresponds to a 5% (or less) chance of obtaining a result like the one that was observed if H_0 was true. Choosing an alpha level of .05, a data set rejects the null hypothesis that data comes from a normally distributed population if the p-value is less than .05, while a p-value greater than .05, points that the population is normally distributed.

In case the gathered data has a normal distribution, a Paired T-Test [7] can be employed in order to determine if statistical significance was observed. In a Paired T-Test, H_0 assumes that the true mean difference between the paired samples is zero, which explains all observable differences by random variation. Conversely, the alternative hypothesis assumes that the true mean difference between the paired samples is not equal to zero. The result is also statistically significant when p ¡ alpha, as a low p-value indicates decreased support for the null hypothesis.

When data cannot be assumed to be normally distributed, which usually happens with non-parametric data, a Wilcoxon signed-rank test can be employed to find statistical differences.

Our statistical tests were made using *IBM SPSS Statistics* [8], provided by IST. All the tests ran with SPSS can be found in Appendix D.

[6] Shapiro-Wilk Test, "Testing for normality using SPSS Statistics" (accessed: 2020-08-03): https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php

[7] Dependent T-Test, "Testing for normality using SPSS Statistics" (accessed: 2020-08-03): https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php

[8] SPSS Statistics: https://www.ibm.com/products/spss-statistics

5.8.1 Accuracy

We measured the matching task's accuracy under each different condition as the percentage of intersection between the two areas of interest outlined by the participants.

Running a Shapiro-Wilk test for the medium percentage of intersection obtained by each participant, we obtained p values of p = .104 and p = .107 for the Veridical and MAGIC conditions, respectively, which indicated that our accuracy results were normally distributed.

Afterwards, we employed a paired t-test which indicated that the percentage of intersection was higher for our approach (M = 46.6, SD = 19.6) than for the Veridical condition (M = 33.8, SD = 13.8), t(9) = -2.31, p = .047 .

As we obtained $p = .047 < .05$, we proved the statistical significance of the results and refuted the hypothesis that observable differences were caused by random variation.

Results for accuracy metric are presented in Figure 5.11.

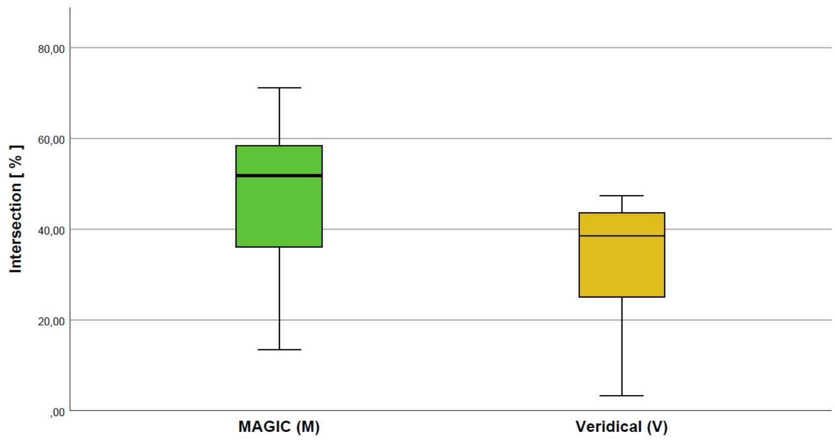


Figure 5.11: Tasks' accuracy as intersection percentage for each condition.

5.8.2 Completion Time

Similarly to the accuracy data, we ran a Shapiro-Wilk test for the medium completion time obtained by each participant, under each different condition to check our distribution for normality. We obtained the values of p = .889 and p = .213 for the Veridical and Approach conditions, respectively, which indicated that completion time data was also normally distributed.

Subsequently, we employed a paired t-test, with results t(9) = -2.00, p = .076. This test showed that there were no statistically significant differences in task completion time using our approach (M = 28.0, SD = 15.9) or Veridical condition (M = 19.8, SD = 5.8), as $p = .076 > .05$.

Results for the completion time metric are presented in Figure 5.12.

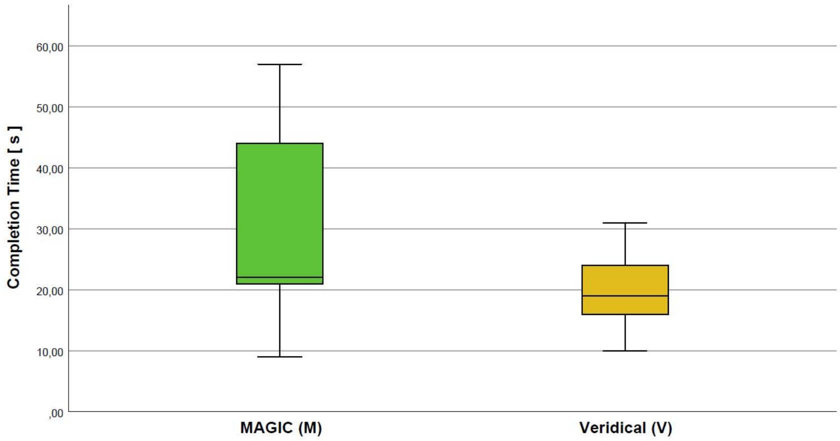


Figure 5.12: Tasks' completion time for each condition.

5.8.3 User Preferences

When it comes to User Preferences, our questionnaires contained a set of statements to be answered on a 6-point Likert-scale, from Strongly Disagree (1) to Strongly Agree (6). For Likert Scales, it follows from definition that data can not be normally distributed as its values are discrete. Therefore, no normality test was employed.

The questionnaires and associated results are summarized in Figure 5.13. Questionnaires showed no statistically significant differences between the two approaches in terms of Attentional Allocation and Perceived Message Understanding, although a tendency for a better message understanding under MAGIC. However, five participants reported in their observations that targets were easier to understand when performing the experiment under MAGIC, and two participants pointed out they had to move less in order to understand what their partner was pointing to.

In terms of feeling of co-presence, the Wilcoxon signed-rank test reported a significant increase in the feeling of presence of the remote partner in the virtual environment under MAGIC (Statement 1.2.), where $Z = -2.236$, p = .025.

Statements	Condition	Median (IQR)
1. Co-Presence		
1.1 I felt present in the virtual environment.	*M*	6 (0.75)
	V	6 (1)
1.2 I felt that my partner was present in the virtual environment. *	*M*	6 (0)
	V	5 (0.75)
1.3 I felt that my presence was evident to my partner.	*M*	6 (1)
	V	5 (1.75)
2. Attentional Allocation		
2.1 I remained focused on my partner throughout our interaction.	*M*	6 (1)
	V	6 (1)
2.2 I remained focused on the workspace throughout our interaction.	*M*	6 (1)
	V	6 (1)
2.3 I felt that my partner could remain focused on me throughout our interaction.	*M*	5.5 (1)
	V	5 (2)
2.4 I felt that my partner could remain focused on the workspace throughout our interaction.	*M*	5.5 (1)
	V	5.5 (1.75)
3. Perceived Message Understanding		
3.1 I could understand my partner's actions.	*M*	5 (0.75)
	V	5 (0.75)
3.2 I felt that my partner could understand my actions.	*M*	5 (0)
	V	5 (1)
3.3 It was easy to understand what was happening in the workspace.	*M*	5 (0.75)
	V	5 (2)
3.4 I felt that I was pointing to where I wanted to point.	*M*	5.5 (1.75)
	V	5 (1)
3.5 It was easy to outline an area.	*M*	5 (0.75)
	V	4 (1)
3.6 It was easy to understand the area outlined by my partner.	*M*	5 (0.75)
	V	4 (1.75)

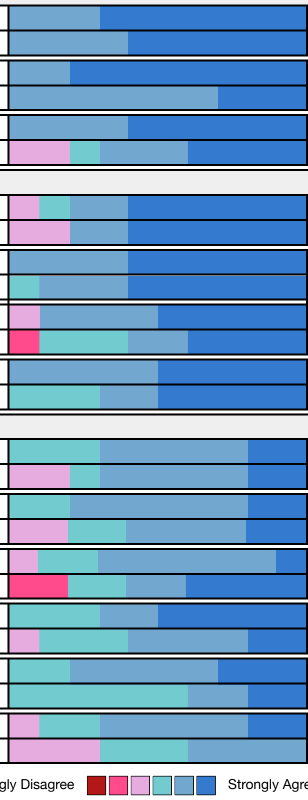


Figure 5.13: Results from the user preferences questionnaire for both MAGIC(M) and Veridical(V) conditions. * indicates statistical significance.

5.9 Discussion

In the beginning of this dissertation, we stated that we wanted to prove the following hypotheses:

H1: *A face-to-face setting coupled with manipulations that enable perspective sharing improves workspace awareness.*

H2: *Manipulations applied to the remote person's representation are not noticeable by the local collaborator.*

H3: *Manipulations applied to the remote person's representation improve the feeling of co-presence.*

To evaluate these initial hypotheses, we conducted the user study gathering data on the matching tasks' accuracy and completion time under our approach and a veridical control condition. We employed questionnaires to assess co-presence felt between participants under each condition, along with additional information on attentional allocation and perceived message understanding.

We repeated the same experiment procedure under two conditions: face-to-face with our approach, where we manipulated the remote participant's representation to enable perspective sharing, and veridical face-to-face, where participants were face-to-face but had opposing points-of-view of the workspace, as in real life face-to-face interactions.

Results presented in Section 5.8.1 showed a better percentage of intersection between the area of interest outlined by both participants when sharing perspective, delivering a better understanding of the collaboration tasks. Indeed, our approach improved the accuracy by 12.8% when compared to the Veridical condition. Through the form of observations in the User Preference Questionnaire, participants also reported a better understanding of the target location under our approach, with some pointing out they had to move less to understand where their partner was pointing to. This increase in accuracy under MAGIC makes sense, considering that when both participants share the same point-of-view of the model, there is a better understanding of the pointing gesture, as people can perceive the model in the exact same way as their partner. The number of occlusions is also reduced comparing to when participants see opposite sides of the model, which explains why participants had to move less in order to understand what their partner was referring to. It is to be noticed however that the questionnaire results did not show any statistically significant differences in Perceived Message Understanding statements between both conditions. This could be due to a ”false” sensation of understanding that was more prominent under the veridical condition, i.e. people thought they perceived their partner's actions correctly when in fact they failed the matching task. Therefore, the accuracy drooped under the veridical condition even though participants felt a similar personal performance.

As both conditions employed a life-sized representation of the remote avatar in a face-to-face configuration, consequential communication was ensured under both conditions. Nevertheless, when sharing perspective, there was an increase of feedthrough mechanisms, as the manipulations of artifacts present in the workspace were visible to the local observer at all times, contrarily to what happened in the veridical condition. The increase of accuracy of the pointing task also pointed out an enhancement in intentional communication, as people showed a better capacity to use and understand deictic gestures. Evaluating workspace awareness in terms of the previous three communication mechanisms, we can argue that our approach improved workspace awareness, proving our first hypothesis, **H1**.

We also hypothesized that the manipulations applied to the remote person's avatar were not noticeable by the local collaborator. We directly asked participants in the User Preference Questionnaire if they noticed any strange behaviors when performing the tasks. Here, none of the participants reported noticing anything strange under conditions MAGIC or Veridical, which leaves us to assume that participants remained unaware of the experiment's manipulations. Since participants did not notice manipulations, collaboration through MAGIC remains straightforward, with no added complexity to the task. We believe participants did not notice any manipulations as our approach manipulates people's gestures in a subtle way that tries to depict human movements as faithfully as possible. Furthermore, in terms of task's efficiency, results presented in Section 5.8.2 showed no statistically significant differences in task completion time under both conditions. Hence, MAGIC does not negatively impact the time collaborators take to complete the shared task. Reaffirming that collaborators do not notice manipulations, there is no need for participants to learn a new interaction method, as it appears to them that they are collaborating traditionally, which doesn't impact time performance. Thus, we proved **H2**.

When it comes to the sense of co-presence enabled by our approach, answers from the User Preference Questionnaire showed no statistically significant differences in the sense of presence felt by the local collaborator itself. However, participants felt that their partner was more present in the virtual environment under MAGIC. We believe MAGIC improves the sense of co-presence since it increases awareness of the remote collaborator's actions in the workspace. As people have a better understanding of what actions their partner is performing, they feel their partner is more present in the workspace. Hence, we validate our last hypothesis, **H3**.

We could argue in the way of limitations, that the experiments focused on objects commensurable with participant's apparent sizes. An open question remains on how to explore the trade-offs between apparent size and level of detail. Too large objects or participants shown at different scales might lead to perceptible Mr. Reed [9]-like distortions. Whether these could have a positive or negative impact on collaboration is an open research question. Also, our experiments assumed static participant positions around the shared workspace. Further research might be needed to see whether these results scale to moving (or more than two) participants. However, scaling this approach to multiple face-to-face participants might prove easier postulated than proved.

Furthermore, verbal communication is crucial in the course of collaboration. However, our approach does not ensure verbal communication between participants that are located in separate physical spaces. During our experimental sessions, participants were able to talk to each other since they shared the physical space, which proved helpful to the collaboration task and ensured a good sense of presence and a more natural interaction. Therefore, future work should complement MAGIC with a viable communication channel. It should also be noted that MAGIC does not work with distal pointing, since the manipulated pointing gestures only reference the correct place when the user touches the desired area. It could be interesting to extend our approach to enable distal pointing.

As results show improved accuracy without disadvantageous effects on completion time, we can argue that MAGIC increases task performance. Additionally, we confirmed that the manipulations we employ to the representation of the remote collaborator are not noticeable and do even improve the sense of presence of the remote collaborator in the workspace. We can then complete this discussion by saying that we validated our research statement as all our hypotheses confirmed that *"Manipulating remote people's virtual representations to enable a shared point-of-view improves workspace awareness in face-to-face shared 3D workspaces "*.

[9] https://www.marvel.com/characters/mister-fantastic

6

Conclusions and Future Work

In today's society, remote work is increasingly common and plays an essential role in projects where geographically separated people need to perform joint tasks. Video-conference and telepresence technologies allow verbal and visual communication between collaborators. However, they fail in other aspects of nonverbal communication such as gaze, deictic gestures, and other nonverbal cues, so important when carrying out intentional communication. These problems reduce the sense of presence in remote meetings, resulting in inefficient communication and impairing collaborative work. With Virtual Reality becoming increasingly capable and accessible, it is possible to create a remote meeting close to the co-located experience. In virtual meetings, people can meet in a face to face formation with life-sized embodiments, which promotes the sense of presence and workspace awareness. However, face-to-face virtual encounters still face problems in enabling specific remote tasks, such as analyzing three-dimensional models. As collaborators do not share the same forward-backward orientation and there is no common orientation of right or left, taking advantage of non-verbal communication is still difficult. Also, contrary points-of-view paired with occlusions that result from people standing in front of each other can lead to serious communication missteps.

In this dissertation, we introduced MAGIC, a design space for VR environments that enables people to engage in 3D object-centered face-to-face collaboration in a shared workspace while sharing the same perspective of the workspace. The remote participant's virtual representation is subtly transformed so that gestures performed in the remote workspace are virtually correct in the local workspace. MAGIC also integrates person space, task space, and reference space, minimizing the need for meeting participants to frequently switch attention between the remote collaborator's person space and the workspace.

We conducted a user study with 10 participants to find if MAGIC could improve workspace awareness and task performance by comparing the performance of the users when doing a pointing matching task with MAGIC versus a veridical baseline condition (face to face with opposing points of view). In terms of task performance, results showed that MAGIC improved the accuracy of tasks using pointing gestures without affecting the task completion time. Results also suggested that the remote person's representation manipulations were not noticeable by the participants, improving the sense of co-presence. Hence, we validated our research statement.

We believe that MAGIC opens a door for interesting future work, some related to our approach's nature and other related to a more technical aspect. The evaluation prototype was designed to use the Oculus Rift set, which led to the use of abstract avatars to represent the users. Using depth cameras to capture the user's real image to be rendered in the virtual environment could be employed to study the real sense of presence induced by our approach. The user study also focused on the accuracy of pointing tasks since other deictic references could not be accessed using our hardware only. It would be interesting to focus on multi-modal deictic references by combining detailed gaze with gestures. Future work could also target the extension of our approach to a multi-user set-up. Finally, our approach was designed in order to manipulate the remote user movements in imperceptible ways. However, we believe that perceptible exaggerated distortions should be studied to ascertain whether they have the potential to either improve or hinder collaboration.

Throughout history, VR research has focused on faithfully reproducing reality. However, we firmly believe that manipulating perception, both in subtle and not so subtle ways, can be used to greater advantage in the future.

Bibliography

[1] C. L. Ventola, "Medical applications for 3D printing: current and projected uses," *Pharmacy and Therapeutics*, vol. 39, no. 10, p. 704, 2014.

[2] M. Sousa, D. Mendes, S. Paulo, N. Matela, J. Jorge, and D. S. Lopes, "VRRRRoom: Virtual reality for radiologists in the reading room," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages=4057–4062, year=2017*.

[3] J. D. Shirk, D. D. Thiel, E. M. Wallen, J. M. Linehan, W. M. White, K. K. Badani, and J. R. Porter, "Effect of 3-dimensional virtual reality models for surgical planning of robotic-assisted partial nephrectomy on surgical outcomes: a randomized clinical trial," *JAMA network open*, vol. 2, no. 9, pp. e1 911 598–e1 911 598, 2019.

[4] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson, "Room2room: Enabling life-size telepresence in a projected augmented reality environment," in *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing, pages=1716–1725, year=2016*.

[5] J. Zillner, C. Rhemann, S. Izadi, and M. Haller, "3D-board: a whole-body remote collaborative whiteboard," in *Proceedings of the 27th annual ACM symposium on User interface software and technology, pages=471–479, year=2014*.

[6] T. Dodds, B. Mohler, and H. Bülthoff, "A communication task in hmd virtual environments: Speaker and listener movement improves communication," in *23rd Annual Conference on Computer Animation and Social Agents (CASA 2010), pages=1–4, year=2010*.

[7] M. Feick, T. Mok, A. Tang, L. Oehlberg, and E. Sharlin, "Perspective on and Re-Orientation of Physical Proxies in Object-Focused Remote Collaboration," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pages=1–13, year=2018*.

[8] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "Vishnu: virtual immersive support for HelpiNg users an interaction paradigm for collaborative remote guiding in mixed reality," in *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE), pages=9–12, year=2016, organization=IEEE*.

[9] M. Sousa, D. Mendes, R. K. d. Anjos, D. S. Lopes, and J. Jorge, "Negative Space: Workspace Awareness in 3D Face-to-Face Remote Collaboration," in *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, pages=1–2, year=2019*.

[10] C. Gutwin and S. Greenberg, "A descriptive framework of workspace awareness for real-time groupware," *Computer Supported Cooperative Work (CSCW)*, vol. 11, no. 3-4, pp. 411–446, 2002.

[11] H. Cochet and J. Vauclair, "Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences," *Laterality: Asymmetries of Body, Brain and Cognition*, vol. 19, no. 3, pp. 278–301, 2014.

[12] S. Mayer, J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, and N. Henze, "Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages=1–14, year=2020*.

[13] M. Sousa, R. K. dos Anjos, D. Mendes, M. Billinghurst, and J. Jorge, "Warping Deixis: Distorting Gestures to Enhance Collaboration," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pages=1–12, year=2019*.

[14] Collaboration., *The Cambridge Dictionary - English*. Cambridge University Press.

[15] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," in *Situational awareness, pages=32–64, year=2017, publisher=Routledge*.

[16] S. Greenberg, C. Gutwin, and A. Cockburn, "Awareness through fisheye views in relaxed-WYSIWIS groupware," in *Graphics interface, volume=96, pages=28–38, year=1996*.

[17] M. J. Adams, Y. J. Tenney, and R. W. Pew, "Situation awareness and the cognitive management of complex systems," *Human factors*, vol. 37, no. 1, pp. 85–104, 1995.

[18] B. Buxton, "Mediaspace–Meaningspace–Meetingspace," in *Media space 20+ years of mediated life, organization=Springer, pages=217–231*, 2009.

[19] W. Buxton, "Telepresence: Integrating shared task and person spaces," in *Proceedings of graphics interface, volume=92, number=1992, pages=123–129, year=1992, organization=Canadian Information Processing Society Toronto, Canada*.

[20] H. Ishii and M. 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, pages=525–532, year=1992*.

[21] A. Kunz, T. Nescher, and M. Kuchler, "Collaboard: a novel interactive electronic whiteboard for remote collaboration with people on content," in *2010 International Conference on Cyberworlds, pages=430–437, year=2010, organization=IEEE*.

[22] K. Higuchi, Y. Chen, P. A. Chou, Z. Zhang, and Z. Liu, "Immerseboard: Immersive telepresence experience using a digital whiteboard," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages=2383–2392, year=2015*.

[23] A. M. Genest, C. Gutwin, A. Tang, M. Kalyn, and Z. Ivkovic, "KinectArms: a toolkit for capturing and displaying arm embodiments in distributed tabletop groupware," in *Proceedings of the 2013 conference on Computer supported cooperative work, pages=157–166, year=2013*.

[24] S. Junuzovic, K. Inkpen, T. Blank, and A. Gupta, "IllumiShare: sharing any surface," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages=1919–1928, year=2012*.

[25] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "BeThere: 3D mobile collaboration with spatial input," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages=179–188, year=2013*.

[26] H. Benko, R. Jota, and A. Wilson, "MirageTable: freehand interaction on a projected augmented reality tabletop," in *Proceedings of the SIGCHI conference on human factors in computing systems, pages=199–208, year=2012*.

[27] D. Leithinger, S. Follmer, A. Olwal, and H. Ishii, "Physical telepresence: shape capture and display for embodied, computer-mediated remote collaboration," in *Proceedings of the 27th annual ACM symposium on User interface software and technology, pages=461–470, year=2014*.

[28] J. Li, S. Greenberg, E. Sharlin, and J. Jorge, "Interactive two-sided transparent displays: designing for collaboration," in *Proceedings of the 2014 conference on Designing interactive systems, pages=395–404, year=2014*.

[29] E. Wood, J. Taylor, J. Fogarty, A. Fitzgibbon, and J. Shotton, "Shadowhands: High-fidelity remote hand gesture visualization using a hand tracker," in *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, pages=77–84, year=2016*.

[30] R. Wolff, D. Roberts, A. Steed, and O. Otto, "A review of telecollaboration technologies with respect to closely coupled collaboration," *International Journal of Computer Applications in Technology*, vol. 29, pp. 11–26, 01 2007.

[31] A. J. Sellen, "Speech patterns in video-mediated conversations," in *Proceedings of the SIGCHI conference on Human factors in computing systems, pages=49–59, year=1992*.

[32] Y. Ichikawa, K.-i. Okada, G. Jeong, S. Tanaka, and Y. Matsushita, "MAJIC videoconferencing system: experiments, evaluation and improvement," in *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work ECSCW'95, pages=279–292, year=1995, organization=Springer*.

[33] O. Morikawa and T. Maesako, "HyperMirror: toward pleasant-to-use video mediated communication system," in *Proceedings of the 1998 ACM conference on Computer supported cooperative work, pages=149–158, year=1998*.

[34] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec, "Achieving eye contact in a one-to-many 3D video teleconferencing system," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, pp. 1–8, 2009.

[35] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in *2011 10th IEEE International Symposium on Mixed and Augmented Reality, pages=137–146, year=2011, organization=IEEE*.

[36] S. Beck, A. Kunert, A. Kulik, and B. Froehlich, "Immersive group-to-group telepresence," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 616–625, 2013.

[37] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou *et al.*, "Holoportation: Virtual 3D teleportation in real-time," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pages=741–754, year=2016*.

[38] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-me: An adaptive avatar for mixed reality remote collaboration," in *Proceedings of the 2018 CHI conference on human factors in computing systems, pages=1–13, year=2018*.

[39] M. M. Bekker, J. S. Olson, and G. M. Olson, "Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design," in *Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques, 49 pages=157–166, year=1995*.

[40] A. Genest and C. Gutwin, "Characterizing deixis over surfaces to improve remote embodiments," in *ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work, 24-28 September 2011, Aarhus Denmark, pages=253–272, year=2011, organization=Springer*.

[41] D. Nguyen and J. Canny, "MultiView: spatially faithful group video conferencing," in *Proceedings of the SIGCHI conference on human factors in computing systems, pages=799–808, year=2005*.

[42] A. H. Hoppe, F. van de Camp, and R. Stiefelhagen, in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), title=Personal Perspective: Using Modified World Views to Overcome Real-Life Limitations in Virtual Reality, year=2018, volume=, number=, pages=577-578,*.

[43] D. Roth, D. Mal, C. F. Purps, P. Kullmann, and M. E. Latoschik, "Injecting Nonverbal Mimicry with Hybrid Avatar-Agent Technologies: A Naïve Approach," in *Proceedings of the Symposium on Spatial User Interaction, pages=69–73, year=2018*.

[44] K. Zibrek, E. Kokkinara, and R. McDonnell, "The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role?" *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1681–1690, 2018.

[45] A. Melim, "Increasing fidelity with constellation-tracked controllers," https://developer.oculus.com/blog/increasing-fidelity-with-constellation-tracked-controllers/, accessed: 2020-05-05.

[46] A. Aristidou and J. Lasenby, "FABRIK: A fast, iterative solver for the inverse kinematics problem," *Graph. Models*, vol. 73, no. 5, pp. 243–260, Sep. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.gmod.2011.05.003

[47] A. Aristidou, Y. Chrysanthou, and J. Lasenby, "Extending FABRIK with Model Constraints," *Comput. Animat. Virtual Worlds*, vol. 27, no. 1, pp. 35–57, Jan. 2016. [Online]. Available: https://doi.org/10.1002/cav.1630

[48] J. J. LaViola, E. Kruijff, R. P. McMahan, D. A. Bowman, and I. Poupyrev, ""3D user interfaces: Theory and practice "." Boston: Addison-Wesley, 2017, ch. 11, pp. 357–360.

[49] Cha Zhang and Tsuhan Chen, in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205), title=Efficient feature extraction for 2D/3D objects in mesh representation, year=2001, volume=3, number=, pages=935-938 vol.3,*.

A

Profile Questionnaire

Shared Vision - Enabling the Share of Perspective in Face-to-Face Collaboration Scenarios

Object centered collaboration where two users discuss and analyze three dimensional models in a shared virtual workspace.

*Obrigatório

Task

Collaborative "matching" task on abstract 3D model. Participants assume the roles of Demonstrator and Interpreter:

- The Demonstrator is able to see a highlighted area of interest in the workspace. It is his 'job' to communicate that area to the Interpreter, by outlining it with his pointing finger.
- The Interpreter cannot see the highlight. Following the Demonstrator's gestures, the Interpreter uses his pointing fingers to outline the interpreted area.

*Each matching task starts when the Demonstrator presses the Index Trigger button from one the Oculus Touch Controller making the workspace appear.

*To outline an area, the user must press the A or X button from the Touch Controllers, depending on the hand he chooses to use. The outline stops when the button is released.

Oculus Touch Controllers' Buttons

Procedure

I. Participants fill the Profile Questionnaire and Consent Form;
II. One participant assumes the role of Interpreter and the other of Demonstrator and execute a series of 16 matching tasks under a first Condition.
III. Participants switch roles and perform another set of 16 tasks.
IV. Participants fill the User Preference Questionnaire for the Condition evaluated first.
V. Participants repeat the steps II and III for the other Condition.
VI. Participants fill the User Preference Questionnaire for the second Condition evaluated.

Profile Questionnaire

Please answer some questions to help the team understand your user profile.

1. 4ID *

To be filled by the team

____

2. Please specify your gender *

Marcar apenas uma oval.

( ) Female
( ) Male
( ) Other

3. Please specify your age *

____

4. Please specify your educational level *

Marcar apenas uma oval.

( ) 9th grade or inferior
( ) High School or Equivalent
( ) University education

5. Please specify your dominant hand *

Marcar apenas uma oval.

- Right
- Left

6. Did you already know your partner? *

Marcar apenas uma oval.

- Yes
- No

7. How often do you use videoconference platforms? (ex: Skype, Google Hangouts, Zoom) *

Marcar apenas uma oval.

- Never
- Rarely
- At least once a month
- At least once a week
- At least once a day

8. How often do you use Virtual Reality environments? *

Marcar apenas uma oval.

- Never
- Rarely
- At least once a month
- At least once a week
- At least once a day

B

User Preference Questionnaire

Shared Vision - Enabling the Share of Perspective in Face-to-Face Collaboration Scenarios

*Obrigatório

1. ID *

   To be filled by the team

   ____

2. Condition *

   To be filled by the team

   Marcar apenas uma oval.

   ( ) A
   ( ) B

User Preferences Questionnaire

At this point you have completed part of the experiment. Please answer these questions about how you felt during the completion of the tasks asked.

1. Co-Presence

3. 1.1. I felt present in the virtual environment. *

   Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	( )	( )	( )	( )	( )	( )	Strongly agree

4. 1.2. I felt that my partner was present in the virtual environment. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

5. 1.3. I felt that my presence was evident to my partner. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

2. Attentional Allocation

6. 2.1. I remained focused on my partner throughout our interaction. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

7. 2.2. I remained focused on the workspace throughout our interaction. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

8. 2.3. I felt that my partner could remain focused on me throughout our interaction. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

9. 2.4. I felt that my partner could remain focused on the workspace throughout our interaction. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

3. Perceived Message Understanding

10. 3.1. I could understand my partner's actions. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

11. 3.2. I felt that my partner could understand my actions. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree	◯	◯	◯	◯	◯	◯	Strongly agree

12. 3.3. It was easy to understand what was happening in the workspace. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree							Strongly agree

13. 3.4. I felt that I was pointing to where I wanted to point. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree							Strongly agree

14. 3.5. It was easy to outline an area. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree							Strongly agree

15. 3.6. It was easy to understand the area outlined by my partner. *

Marcar apenas uma oval.

	1	2	3	4	5	6	
Strongly disagree							Strongly agree

16. Did you notice any strange behavior during the session?

17. Observations and Suggestions:

C

Communication Protocols

	TCP	UDP
Handshake	Connection-oriented protocol: the two devices must establish a connection before transmission	Connection-less protocol: no connections needed before transmission
Reliability	Re-transmittion of data in case of failure	No guaranteed delivery of data, some packets can be lost
Ordering	Data streamed, guaranteeing packets arrive in right order	Packets sent independently from each other, meaning packets can arrive in any order
Error checking	Checks packets for integrity, and recovers corrupted packets	Checks packets for integrity, but discards them in case of corruption
Speed	Slower transfer speed due to intensive error checking and handling	Faster transfer speed as re-transmission and packet recuperation is not done

Table C.1: Comparison between TCP and UDP.

D

SPSS Results

D.1 SPSS Analysis For Accuracy Results

Participant	Average Intersection [%]	
	Veridical	Approach
1	29.87	50.39
2	43.35	58.43
3	22.54	13.51
4	37.93	13.74
5	3.27	36.02
6	39.17	63.34
7	25.09	50.17
8	47.39	55.98
9	43.60	53.19
10	45.58	71.11

Table D.1: Average % Intersection obtained by each participant in the set of 16 tasks, for each condition.

Descritivos

			Estatística	Erro Erro
Veridical	Média		33,7783	4,35442
	95% Intervalo de Confiança para Média	Limite inferior	23,9279	
		Limite superior	43,6287	
	5% da média aparada		34,7170	
	Mediana		38,5521	
	Variância		189,609	
	Erro Desvio		13,76987	
	Mínimo		3,27	
	Máximo		47,39	
	Intervalo		44,11	
	Amplitude interquartil		19,64	
	Assimetria		-1,293	,687
	Curtose		1,491	1,334
Approach	Média		46,5875	6,20559
	95% Intervalo de Confiança para Média	Limite inferior	32,5495	
		Limite superior	60,6255	
	5% da média aparada		47,0627	
	Mediana		51,7933	
	Variância		385,094	
	Erro Desvio		19,62380	
	Mínimo		13,51	
	Máximo		71,11	
	Intervalo		57,59	
	Amplitude interquartil		29,21	
	Assimetria		-,928	,687
	Curtose		-,155	1,334

Figure D.1: Descriptive Statistics for intersection percentage for each condition.

Testes de Normalidade

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estatística	df	Sig.	Estatística	df	Sig.
Veridical	,219	10	,193	,871	10	,104
Approach	,272	10	,034	,872	10	,107

a. Correlação de Significância de Lilliefors

Figure D.2: Normality Test Results for intersection percentage.

Teste-T

Estatísticas de amostras emparelhadas

		Média	N	Erro Desvio	Erro padrão da média
Par 1	Veridical	33,7783	10	13,76987	4,35442
	Approach	46,5875	10	19,62380	6,20559

Correlações de amostras emparelhadas

		N	Correlação	Sig.
Par 1	Veridical & Approach	10	,492	,148

Teste de amostras emparelhadas

		Diferenças emparelhadas					t	df	Sig. (2 extremidades)
		Média	Erro Desvio	Erro padrão da média	95% Intervalo de Confiança da Diferença: Inferior	Superior			
Par 1	Veridical - Approach	-12,80921	17,56638	5,55498	-25,37544	-,24298	-2,306	9	,047

Figure D.3: Paired T-Test Test Results for intersection percentage.

D.2 SPSS Analysis For Completion Time Results

Participant	Average Completion Time [s]	
	Veridical	Approach
1	16	28
2	19	44
3	25	21
4	21	21
5	16	09
6	10	11
7	24	57
8	31	46
9	19	23
10	17	21

Table D.2: Average Completion Time [s] obtained by each participant in the set of 16 tasks for each condition.

Descritivos			Estatística	Erro Erro
Veridical	Média		19,8000	1,84270
	95% Intervalo de Confiança para Média	Limite inferior	15,6315	
		Limite superior	23,9685	
	5% da média aparada		19,7222	
	Mediana		19,0000	
	Variância		33,956	
	Erro Desvio		5,82714	
	Mínimo		10,00	
	Máximo		31,00	
	Intervalo		21,00	
	Amplitude interquartil		8,25	
	Assimetria		,384	,687
	Curtose		,674	1,334
Approach	Média		28,0000	5,03101
	95% Intervalo de Confiança para Média	Limite inferior	16,6191	
		Limite superior	39,3809	
	5% da média aparada		27,4444	
	Mediana		22,0000	
	Variância		253,111	
	Erro Desvio		15,90947	
	Mínimo		9,00	
	Máximo		57,00	
	Intervalo		48,00	
	Amplitude interquartil		26,25	
	Assimetria		,706	,687
	Curtose		-,521	1,334

Figure D.4: Descriptive Statistics for completion time for each condition.

Testes de Normalidade

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estatística	df	Sig.	Estatística	df	Sig.
Veridical	,157	10	,200*	,970	10	,888
Approach	,223	10	,171	,899	10	,213

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Figure D.5: Normality Test Results for completion time.

Teste-T

Estatísticas de amostras emparelhadas

		Média	N	Erro Desvio	Erro padrão da média
Par 1	Veridical	19,8000	10	5,82714	1,84270
	Approach	28,0000	10	15,90947	5,03101

Correlações de amostras emparelhadas

		N	Correlação	Sig.
Par 1	Veridical & Approach	10	,645	,044

Teste de amostras emparelhadas

		Diferenças emparelhadas					t	df	Sig. (2 extremidades)
					95% Intervalo de Confiança da Diferença				
		Média	Erro Desvio	Erro padrão da média	Inferior	Superior			
Par 1	Veridical - Approach	-8,20000	12,94261	4,09281	-17,45858	1,05858	-2,004	9	,076

Figure D.6: Paired T-Test Test Results for completion time.

D.3 SPSS Analysis For Questionnaire Results

Estatística Descritiva

	N	Média	Desvio Padrão	Mínimo	Máximo	Percentis 25o.	Percentis 50º (Mediana)	Percentis 75º
A1	10	5,7000	,48305	5,00	6,00	5,0000	6,0000	6,0000
A2	10	5,8000	,42164	5,00	6,00	5,7500	6,0000	6,0000
A3	10	5,6000	,51640	5,00	6,00	5,0000	6,0000	6,0000
A4	10	5,3000	1,05935	3,00	6,00	4,7500	6,0000	6,0000
A5	10	5,6000	,51640	5,00	6,00	5,0000	6,0000	6,0000
A6	10	5,3000	,94868	3,00	6,00	5,0000	5,5000	6,0000
A7	10	5,5000	,52705	5,00	6,00	5,0000	5,5000	6,0000
A8	10	4,9000	,73786	4,00	6,00	4,0000	5,0000	5,2500
A9	10	5,0000	,66667	4,00	6,00	4,7500	5,0000	5,2500
A10	10	4,7000	,82327	3,00	6,00	4,0000	5,0000	5,0000
A11	10	5,2000	,91894	4,00	6,00	4,0000	5,5000	6,0000
A12	10	5,1000	,73786	4,00	6,00	4,7500	5,0000	6,0000
A13	10	4,8000	,91894	3,00	6,00	4,0000	5,0000	5,2500
B1	10	5,6000	,51640	5,00	6,00	5,0000	6,0000	6,0000
B2	10	5,3000	,48305	5,00	6,00	5,0000	5,0000	6,0000
B3	10	4,9000	1,19722	3,00	6,00	3,7500	5,0000	6,0000
B4	10	5,2000	1,22927	3,00	6,00	4,5000	6,0000	6,0000
B5	10	5,5000	,70711	4,00	6,00	5,0000	6,0000	6,0000
B6	10	4,8000	1,31656	2,00	6,00	4,0000	5,0000	6,0000
B7	10	5,2000	,91894	4,00	6,00	4,0000	5,5000	6,0000
B8	10	4,7000	1,05935	3,00	6,00	3,7500	5,0000	5,2500
B9	10	4,6000	1,07497	3,00	6,00	3,7500	5,0000	5,2500
B10	10	4,6000	1,57762	2,00	6,00	3,5000	5,0000	6,0000
B11	10	4,7000	,94868	3,00	6,00	4,0000	5,0000	5,2500
B12	10	4,6000	,84327	4,00	6,00	4,0000	4,0000	5,2500
B13	10	4,1000	,87560	3,00	5,00	3,0000	4,0000	5,0000

Figure D.7: Descriptive Statistics for questionnaires' results under for each condition: A1 to A13 - MAGIC (Condition A); B1 to B13 - Veridical (Condition B)

Estatísticas de teste[a]

	B1 - A1	B2 - A2	B3 - A3	B4 - A4	B5 - A5	B6 - A6	B7 - A7	B8 - A8	B9 - A9	B10 - A10	B11 - A11	B12 - A12	B13 - A13
Z	-1,000[b]	-2,236[b]	-1,633[b]	-,577[b]	-,333[b]	-1,667[b]	-1,000[b]	-,574[b]	-1,190[b]	-,180[b]	-1,667[b]	-1,179[b]	-1,469[b]
Significância Sig. (bilateral)	,317	,025	,102	,564	,739	,096	,317	,566	,234	,857	,096	,238	,142

a. Teste de Classificações Assinadas por Wilcoxon

b. Com base em postos positivos.

Figure D.8: Wilcoxon Test Results for questionnaires