Facing the challenges of fully immersive virtual and augmented reality for simulation and training

Giovanni Avveduto

Facing the challenges of fully immersive virtual and augmented reality for simulation and training

Perceptual Robotics Laboratory, Tecip

Scuola Superiore Sant'Anna

Author:

Giovanni Avveduto

Supervisor:

Franco Tecchia

Ph.D Course
Emerging Digital Technologies

Academic Year
2016 / 2017

July 14th, 2017

Giovanni Avveduto

ABSTRACT

Immersive Virtual Environments are extremely powerful tools; they allow to build worlds whose only limit is the imagination. They are not only useful to artists letting their creativity run wild, but they can also have a tremendous impact on several everyday activities. Using IVEs we could meet a friend living far away as we're located in the same room, we could relax during our break by swimming with the whales or visiting ancient ruins on the other side of the world or we could practice a dangerous work procedure in a totally controlled and safe environment. This work will try to give an overview of the technological challenges faced when developing fully immersive virtual environments, presenting the state-of-the-art of the enabling technologies. The first part of the thesis presents the solutions adopted to face some of these challenges faced encountered during the development of fully immersive VEs of different types—namely Virtual Reality, Augmented Reality and Mixed Reality systems—exploiting different technologies. In the second part of this work, we will try to evaluate the impact of these technologies on different application fields. An investigation on the use of IVEs as training tools in industrial contexts is conducted. Finally we will explore the effects that IVEs technologies could produce on our social life and social behaviour, and vice versa, how our social habits influence the way we use new technologies and what are the user's expectations.

ACKNOWLEDGMENTS

In the end also this last long journey has finished.

First of all I want to thanks all my family that supported me during my whole life.

Thanks to Ilenia, my love, adventures companion with whom I shared the best moments of my life but who also supported me during the hardest ones.

Many thanks to all the people I had the pleasure to work with during this journey. First of all to Franco, who allowed me to work on exciting and challenging projects together with Marcello.

Thanks to all colleagues with whom I shared disappointments and successes and became my closest friends. Thanks to Raffaello, Cristian, Michele, Gastone, Claudia, Giulia, Camilla, Marco, Daniele, Paolo, the many Alessandro, Gabriele, Juan, Massimo, Sandro and all the others.

A special “thanks” to Henry, who allowed me to make an unforgettable experience both as a researcher but even more as a human being, what a special person. Thanks to all my colleagues and friends in Chapel Hill. I’ve learnt a lot from them and I share with them wonderful memories.

Thanks to all my friend in Pisa and in the rest of the world who cheered up my life during these years.

I hope I have not forgotten anybody in this list of thanks. Anyway, although not said, this thanks are for you.

RINGRAZIAMENTI

Alla fine anche questo ultimo lungo viaggio si è concluso.

Primi fra tutti voglio ringraziare tuta la mia famiglia che mi ha sempre supportato durante la mia intera esistenza.

Grazie a Ilenia, amore mio, compagna di avventure con cui ho condiviso i monmenti piú belli ma che mi ha supportato anche duranti quelli difficili.

Grazie a tutte le persone con cui ho avuto il piacere di lavorare in questo periodo. Grazie a Franco che che mi ha permesso di lavorare a progetti eccitanti e stimolanti insieme con Marcello.

Grazie a tutti i colleghi con i quali ho condiviso questo periodo, alcuni dei quali sono diventati i miei più cari amici. Grazie a Raffaello, Cristian, Michele, Gastone, Claudia, Giulia, Camilla, Marco, Daniele, ai Paolo, ai molteplici Alessandro, Gabriele, Juan, Massimo, Sandro e a tutti gli altri.

Un ringraziamento speciale a Henry, grazie al quale la mia esperienza a Chapel Hill si è rivelata indimenticabile, sia come ricercatore che come essere umano. Una persona davvero speciale. Grazie a tutti i colleghi e gli amici di Chapel Hill da cui ho imparato molto e con cui condivido stupendi ricordi.

Grazie a tutti i miei amici di Pisa e sparsi nel resto il mondo che hanno allietato la mia vita durante questi anni.

Spero di non aver dimenticato nessuno, in ogni caso, anche se non detto, questo grazie è per voi.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ost optical see-through

vst video see-through

ar augmented reality

mr mixed reality

vr virtual reality

ve virtual environment

ive immersive virtual environment

hmd head-mounted display

1

INTRODUCTION

> “There are a few special things about Virtual Reality to keep in mind, the things that make it important. One is that it’s a reality in which anything can be possible, provided it’s part of the external world. **It’s a world without limitation, a world as unlimited as dreams. It’s also a world that’s shared, like the physical world.** It’s as shared and as objectively real as the physical world is, no more, no less. Exactly how shared or real that is, is open to question, but whatever the physical world has Virtual Reality has as well. The thing that’s remarkably beautiful to me about Virtual Reality is that you can make up reality in Virtual Reality and share it with other people. It’s like having a collaborative lucid dream. It’s like having shared hallucinations, except that you can compose them like works of art; you can compose the external world in any way at all as an act of communication.”
>
> Jaron Lanier, *A Vintage Virtual Reality Interview*, first published in 1988 in the Whole Earth Review magazine and reprinted many times in many languages.

The nowadays very widespread term “virtual reality” was coined in 1987 by Jaron Lanier, founder of the visual programming lab (VPL). However, Virtual Reality in the form we would recognize today started in the 1960s although with a totally different hardware availability. Sutherland presented what is widely considered to be the first virtual reality and augmented reality head-mounted display system in 1968 (Sutherland, 1968). The system was a large contrivance suspended from the ceiling—for this reason called “Sword of Damocles”—and too heavy to be comfortably worn by any user.

Thanks to new displays and tracking technologies VR reached the attention of the public during 1980s and 1990s, hailed by many as the beginning of a new era to suddenly disappear from public view for 25 years. During these decades, however, both Virtual and Augmented Reality have been investigated and a huge number of researches have been carried out across a wide range of fields. VR became a commonplace tool in many areas from military to medicine, from industry to business, from marketing to psychotherapy. Recent key advances in displays, tracking and rendering technologies brought again VR and AR to the public and to be considered able to change the world for the better.

The VR pioneer Fred Brooks defines a virtual reality experience as any in which the user is effectively immersed in a responsive virtual world. VR can be defined in terms of a particular collection of technological hardware, including computers, head-mounted displays, interaction devices and headphones. The focus of virtual reality up to present days was thus technological, rather than experiential (Steuer, 1992). However, today VR technologies are mature enough to allow the community to focus more on the contents and on the experiential output of VR applications. Currently VR is used to enjoy traditional contents developed for traditional media technologies in a different way: VR games are actually standard video games enjoyed in a more immersive way maintaining exactly the same constructs, thus not fully exploiting the potential of VR. However this is only the dawn of VR intended as a new ways to think, produce and experience contents, communicate and collaborate. Virtual reality—today more than ever before—can be instead portrayed as a medium, like telephone or television. A paradigm shift in the way we think VR applications is possible and has indeed already started. Think at what is a movie today and what it could be tomorrow using VR: instead of enjoying a movie in a totally passive way, the spectator could become a participant, he could choose his preferred point of view on the set and—why not—be part of the plot. This is only one possible application context where we could assist to a radical paradigm change in the VR contents production and fruition. So, while nowadays Virtual Reality is a term related more to a technological setup, in the future the meaning of VR could instead mean a totally different way we make experiences.

According to Milgram and Kishino (1994), Mixed Reality (MR) is a particular subset of VEs involving the merging of real and virtual worlds somewhere along the “virtuality continuum” which connects completely real environments to completely virtual ones. Probably the best known of these is Augmented Reality, which refers to all cases in which the real environment is augmented by means of virtual contents. The term “augmented reality” is attributed to former Boeing researcher Tom Caudell, who coined the tern in 1990 while working on a head-mounted apparatus that would display plane’s specific schematics through high-tech eye-wear (Lee, 2012).

Figure 1.1: Milgram's reality-–virtuality continuum.

Azuma defines an AR system to have the following properties (Azuma, Baillot, Behringer, Feiner, Julier, and MacIntyre, 2001):

- combines real and virtual objects in a real environment;
- runs interactively, and in real time; and
- registers real and virtual objects with each other in three dimensions.

The goal of a VR application, no matter the specific implementation, is to technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the surrounding real world. In contrast, real world's perception is not completely suppressed in AR, it is instead supplemented with virtual objects that appear to coexist in the same space as the real world (Azuma, 1997). Augmented Reality is very interesting cause it allows to enhance the user's perception of the world by displaying informations that the user cannot directly perceive with his own senses. The user can be helped in performing real-world tasks by providing additional co-located information. Using the computer as a tool to make a task easier is what Brooks calls *intelligence amplification* (Brooks Jr, 1996): AR is an example of that. Sometimes it is necessary for AR applications to remove real objects from the environment. For example, interior designers can use AR simulate the decoration of a room by replacing the old real furnitures with new virtual ones. Some researchers call this kind of task *diminished reality* (Mann and Fung, 2001), but is considered a subset of AR. Some define AR as a special case of VR, others see AR as a more general concept that includes the VR definition. From a technological standpoint I agree with the latter: the requisites of an AR system comprehend all the components exploited by VR systems and needs to solve further more problems like the registration between the real and virtual environments. In theory, an AR system can act as a VR system by replacing the whole real world with augmented contents. Practically, developing a system that work both for AR and VR still provides some technological challenges—like the difficulty of designing an AR display which can perfectly occlude the real world—and drawbacks.

COLLABORATION AND SOCIAL PRESENCE Human beings are social beings. Communication and collaboration are fundamental aspects of our lives. The VR medium is a technology that interfaces people with machines but, at the same time, it is also capable of providing virtual shared experiences to multiple users: participants can conduct face-to-face interactions as if conducted in person. In this way, VEs became a tool to interface people to people rather than interfacing people to machines. Collaborative VEs are seen by many as the future in telecommunications (Bradley, Walker, and McGrath, 1996; McGrath, Oldroyd, and Walker, 1998; Raskar, Welch, Cutts, Lake, Stesin, and Fuchs, 1998), where multiple participants located in different geographical locations are able to interact and collaborate using Virtual Reality. Social interaction in VEs opens several new issues, both technological and methodological, to be faced in order to address the complexity of the interaction between two persons. The technology must take care to faithfully reproduce a smile or a touch from the partner allowing for a high level of social presence. Social presence is measured as the degree of awareness of the other person in a communication interaction both in mediated or unmediated interactions. We are used to communicate and interact with others everyday, therefore frequent communication and interaction with other people within VEs may enhance sense of ones own and other's existence in virtual environment. This is about social connections that a user establishes within a virtual space; the level of social presence influences one's feeling of being in a virtual environment (Hudson and Cairns, 2014; Rettie, 2003). Therefore, social presence is fundamental in person-to-person communication (Short, Williams, and Christie, 1976). Moreover awareness of the others in the VE strengthens the social sense of being together (Heeter, 1992). People could not only like to communicate across great distances, but also want to perform work remotely in a collaborative working environment where they can provide and share information and exchange views to reach a common understanding. VEs allows distant participants to collaborate and interact in ways which are far beyond what is possible with normal teleconferencing tools employing communicative tools otherwise unavailable. For instance, a collaborative virtual prototyping simulator would allow colleagues to design a new machinery using the power of the creative tools offered by IVEs. In telepresence surgery, multiple surgeons can watch an operation from the same vantage point, and perhaps hand off control to another participating surgeon in a particular situation. Collaborative working environments enable new rich forms of communication and co-operation, and for this reason attracted people from a variety of disciplines. One of the main attracting feature of these kind of applications is that they create a virtual shared space that can be seen more like a shared physical location.

SIMULATION AND TRAINING VEs constructed through computational models allow to simulate portion of the world and its behaviour. Using VR it is possible to simulate—more or less faithfully—the physical "reality". However, the real power of VR is not necessarily to crate a faithful reproduction of reality but rather to offer the possibility to step outside of the normal bounds of reality and realize goals in a totally new and unexpected way (Slater and Sanchez-Vives, 2016). In VR it is indeed possible to violate the rules of the physical world like modifying the gravity force, rewinding the time to experience an event happened in the past multiple times or instantly "teleport" distant people to share the same virtual space. The user is brought into an alternate reality, which could be a representation of an actual space that exists elsewhere, or a purely imaginary environment. AR, in some ways, doesn't offer the same flexibility in terms of imagination. AR in fact is strictly linked to the real world surrounding the user that must continue to obey to physics rules. Opportunities provided by Virtual, Mixed and Augmented Reality address similar issues although with a slightly different perspective. For instance, VR can be used to instruct workers on performing procedures involving virtually simulated machinery and environment, without the needing of the actual machinery and without exposing the worker to risk providing a "sandbox" where operations can be safely performed. Flight simulators are one of the first and most popular example of use of VR for training. Pilots can learn and practice how to fly in any conditions in totally safe way. The computational model behind the simulation allows to control various aspects of the flight as well presenting unexpected or dangerous situations. Not only this type of training can be more effective, but it can bring with it cost benefits by avoiding material damage and wastage. It also most importantly keeps people safe. VR allow skills to be learned without risk.

AR may, in turn, result more effective whenever the real context is fundamental. In AR, in fact, the real environment is not substituted by a virtual counterpart, therefore workers can be trained to perform operations on actual machinery providing virtual aids. This approach has the obvious advantage that the worker is trained to operate on a actual part developing dexterity, on the other hand, certain situations or unexpected happenings cannot be easily simulated and the worker is still exposed to safety risks. The choose between VR and AR must depends on the specific application or task that needs to be performed.

In general, one of the most important consequences of living the training experience in a totally virtual context, or keeping the vision on the real context, is related to the body self visual perception. In MR the real context, including own body, is always present. This has of course an impact in training, especially in tasks where manipulation operations, or other types of direct interaction with the body, take place. Avatar representations, in fact, might not correspond exactly to the dimensions or the current posture of the user and might, although slightly, mislead the self perception and limit the effectiveness of the virtual training.

1.1 RESEARCH AIMS AND QUESTIONS

Aim of this work is to face the technical challenges provided by fully Immersive Virtual Environments (IVEs)—Virtual, Augmented and Mixed Realities—by providing suggestions and technological solutions to them (Research Aim 1 **(RA1)**). Virtual and augmented realities provide many challenges in common—like rendering the virtual world, performing an effective user's tracking, achieving realistic sensory stimulation etc.—as well as specific challenges like the registration between the real and the virtual worlds as well as a visually seamless fusion between them specific to AR. This work highlights the challenges faced during the development of four systems: two fully immersive virtual reality, an augmented reality and a fully immersive mixed reality and presents the technological solutions developed. During the development of the aforementioned solutions, particular attention has been paid to the specific context each system has been developed for. One of of the goal of this thesis was the development of immersive systems that could effectively used as simulation and training platforms in industrial contexts. Then we also aimed at conducting perceptual studies such as the investigation of the influences of fully IVEs on different aspect of our life such as communication, work and entertainment.

Nowadays, VR applications are increasingly demanding in terms of computational resources. They may require performances exceeding the computational power that a single workstation even by several orders of magnitude. At the beginning of my Ph.D. I focused on the development of a projector-based fully IVEs and the resulting hardware/software architecture called *XVR Network Renderer* is presented. The system exploits the use of multiple workstation in a “cluster rendering” configuration to allow the rendering of very complex scenarios using low-mid end workstations, and at the same time allowing an easy configuration of complex immersive visualization systems like CAVEs (Cruz-Neira, Sandin, DeFanti, Kenyon, and Hart, 1992). During the development of this system we also analysed the existing solutions for the navigation of VEs. In order to allow a hands-free long-range navigation, we developed a pressure sensitive mat which can be used as a touch surface for feet that we used to implement intuitive navigation metaphors.

Interest in augmented reality has substantially increased in the past few years. The current trend is to adopt optical see-through (OST) displays to seamlessly merge the virtual with the real world. The Microsoft HoloLens – the first fully self-contained AR headset – is already available to developers and other big names are going to release their solutions soon. However, the shortcoming of the current generation of AR headsets is making virtual elements appear solid not transparent. Optical see-through AR displays are in fact typically not capable of masking real-world elements; as a result of this lack, dark elements appear to be transparent causing virtual objects to be often perceived as translucent ghosts foreclosing the adoption these devices in many situations. Another limit given by this lack is the impossibility of enhancing user’s perception in some ways that have no real-world equivalents, like “x-ray vision”, without generating depth conflicts to the perceptual system. Despite of the obvious visual flaw, we want to assess whether if virtual elements were solid, would it improve performance of certain tasks with an AR system (Research Question 1 **(RQ1)**) and which are the scenarios that would benefit most of this capability **(RQ2)**. We present a system exploiting the use stereoscopic projectors to add real-world masking capability to the current generation of AR headsets. The resulting framework allowed us to conduct our experiments and is made freely available to anyone interested in conducting their studies.

Virtual Reality is widely regarded as an extremely promising solution for industrial training as it allows to perform simulated hands-on activities in a controlled and safe environment. However, we ask what is the impact of this systems on actual knowledge transfer.

Current industrial safety training methods generally rely on repetitive classroom-taught lessons integrated by directions given in the physical workplace, followed by on the job training. It would be extremely interesting to investigate if the use of VR systems could be helpful to effectively train users in performing maintenance or mechanical operations and its effects on learning compared to traditional approaches. Many are the questions that we asked ourselves and that led us to investigate the use of VR systems in this field:

- Can a VR simulator be an effective training instrument? **(RQ3)**
- Can a VR simulator be used to train operators on performing manual procedures? **(RQ4)**
- Which level of knowledge transfer can be achieved using these systems both theoretic and practical? **(RQ5)**
- Can a VR system lead to better results compared to traditional training approaches? **(RQ6)**

Studies aimed at answer these questions were conducted: two MR/VR training simulators were developed and evaluations of the outcomes of the training methodologies are presented.

The last part of the thesis investigates the influences of the new technologies on collaborative VEs and how these reflect on different aspects of our social life. Exploring how new technological solutions can be adopted in different social scenarios can empower classic methodologies or open new perspectives that were not possible before (like enabling virtually co-located collaboration). At the same time, especially in a social context, it is important to evaluate also how a technology is used and fits with the needs of the users. Social factors can deeply impact on user experience and are often used to increase enjoyment and retention. Games are perfectly suited to promote socio/emotion-relational skills like collaboration or competition. The great technological improvements happening today in the entertainment world are reshaping the way we interact with the game's VE opening new unexplored solutions to game designers. In particular Natural User Interfaces (NUIs) are changing the way we act and communicate inside games. Social interaction in games is heavily influenced by the technological solutions adopted to communicate. The implications of this technological revolution on the social interaction in entertaining applications have not been yet extensively addressed.

- Can NUIs and immersive displays increase social presence in games and user engagement more than classic mediated interfaces? **(RQ7)**
- Can NUIs and immersive displays reversed the wrong correlation between flow and social presence that typically happens with classic interfaces? **(RQ8)**
- Does the adoption of novel technologies social games introduces new challenges for game designers? **(RQ9)**

User studies were conducted to answer these questions in the context of a multi-player game exploiting different game mechanics different visualization systems and interaction interfaces.

1.2 CONTRIBUTIONS

In this section a list of contributions of this thesis are presented. Between brackets the research aims/questions answered by each contribution are reported. The first group of contributions referred to the technological aspect can be summarized as following:

C1: An analysis of the challenges provided by IVEs and a survey of the enabling technologies (Chapters 2 and 3). **(RA1)**

C2: The development of an intuitive input device—a pressure sensing carpet—for the navigation of VEs (Appendix A). **(RA1)**

C3: The development of an efficient cluster rendering architecture specifically optimized for multi-display Immersive visualization systems like CAVEs (Chapter 4). **(RA1)**

C4: The development of a state-of-the-art HMD-based IVR setup exploiting custom solutions for the user's tracking and used as platform to develop immersive training simulators (Section 5.1). **(RA1)**

C5: The development of a novel HMD-based Mixed Reality system exploiting the 3D capture of user's hands that are embedded in the VE. The user can see and use his own hand and naturally interact with the virtual objects (Section 5.2). **(RA1)**

C6: The evaluation of the importance of providing to AR headsets the capability of selectively block out the real environment (Section 6.3). **(RQ1, RQ2)**

C7: The development of an open framework exploiting fine lighting control through the adoption of stereoscopic projectors to provide real-world occlusion capability to commodity AR headsets (Appendix B). **(RA1)**

A second group of contribution is referred to the impact of the adoption of VEs technologies on industrial and entertainment application fields:

C8: In the context of industrial training scenarios, the knowledge transfer achievable using VR/MR systems compared to traditional training approaches are investigated. The investigation is conducted both when teaching theoretical notions and when training workers on performing maintenance or safety procedures (Chapter 7). **(RQ3, RQ4, RQ5, RQ6)**

C9: The user's involvement while performing training sessions using a Virtual Reality system against traditional classroom-taught lessons is investigated (Section 7.2). **(RQ3, RQ5, RQ6)**

C10: The comparison of the level of engagement and social presence reached with a classic interfaces like Keyboard & Mouse with a novel natural interface combined with an immersive visualization system is performed (Chapter 8). **(RQ7, RQ8)**

C11: The comparison of two popular games mechanics—competitive vs collaborative—in terms of player engagement and social presence when playing with an immersive natural interface is investigated (Chapter 8). **(RQ9)**

1.3 THESIS STRUCTURE

This thesis is organised as follows:

Chapter 1. This chapter introduced the Immersive Virtual Environments context and the research questions that we tried to address in this work. A summary of the contributions of this work was presented.

Chapter 2. The challenges faced when developing fully immersive VEs are presented as well as a brief survey on the application fields where these technologies have been adopted.

Chapter 3. An overview of the enabling technologies and the state-of-the-art of IVEs is presented.

Chapter 4. This chapter introduces projector-based IVE systems and present a cluster rendering architecture for high demanding rendering applications we developed. The system allows to easily build immersive visualization systems like the 18-projectors CAVE system we present.

Chapter 5. HMD-based IVE systems are introduced and a state-of-the-art VR system and a novel Mixed Reality system that we developed are presented.

Chapter 6. This chapter introduces the technological challenges of optical see-through displays. Two studies to evaluate the importance of providing see-through AR headset with real-world occlusion capabilities are presented. A solution exploiting projectors-based fine lighting control is presented together with an open framework we developed.

Chapter 7. The effectiveness of fully IVEs for industrial training purposes is investigated, and a comparison with the traditional training approaches discussed.

Chapter 8. The influences of the new technologies on Collaborative Virtual Environments and how these reflects on different aspect of our social life are evaluated.

Chapter 9. In the last chapter, the conclusions of this work are presented and discussed.

2

VR/AR CHALLENGES

The aim of virtual reality is to immerse the user in a simulated world that can be both autonomous and responsive to its actions. From a technical standpoint the goal of VR is to provide computer-generated stimuli to our senses replacing the real world sensory perception. If the substitution is effective, our inner model of the surrounding is inferred from the provided sensory stream rather than from the perception of the real world. Hence, the consciousness of real world became consciousness of the virtual one, despite that the user know for sure that the virtual world is not real. Differently from VR, aim of AR is to enhance the user's perception by providing additional information that he cannot directly perceive with his own sense. A seamless merging between the virtual and the real worlds would allow the user to perceive the virtual objects as real improving the realism of the experience. For virtual objects to appear convincing they must match the real environment both in terms of visual appearance (lighting conditions, reflections and so on) and must interact with the real environment as they would do in real life (realistically collide with the real environment, being subject to the gravity force and so on). VR technology is probably, at present, more mature than the AR, as the technological demands for AR are much higher (Azuma, Baillot, Behringer, Feiner, Julier, and MacIntyre, 2001; Azuma, 1997; Van Krevelen and Poelman, 2010), which might be the main reason why the field of AR is taking longer to mature. Most notably, in order to provide a truly convincing experience to users, an AR system must brilliantly solve the challenges arising when trying to display virtual object inside the real world. A constant research activity can be seen around topics such as: display technologies (Zhou, Duh, and Billinghurst, 2008), accurate tracking of the user's movements (You and Neumann, 2001; Zhou, Duh, and Billinghurst, 2008), the registration of the virtual contents with the real world (Bajura and Neumann, 1995; You and Neumann, 2001), and the minimization of the latency between the estimation of the user's pose and the presentation of the contents to the user's eyes. Most of this challenges must be faced in VR systems too. The amount and the level of each challenge depends on the level of accuracy that each sense require to be effectively stimulated. The level of realism and plausibility and responsiveness of the VE can greatly vary, and is dependent from the technological solutions adopted as well as from the interaction metaphors adopted. Some applications could not require high-level of realism and could therefore stimulate only one or few senses rather than providing a fully immersive experience.

Four are the key elements that shape each virtual reality experience: a **virtual world**, **immersion**, **sensory feedback** (responding to user input), and **interactivity** (Sherman and Craig, 2002).

2.1 VIRTUAL WORLD

There is not a unique definition of virtual world, however, we try to express the key concepts important for our purpose. A virtual world is the content of a given medium (Sherman and Craig, 2002). It may exist solely in the mind of its originator or be shared with others. A virtual world can exist even without being displayed much like play or film scripts exist independently of specific instances of their performances. Such scripts do in fact describe virtual worlds. In VEs we typically refer to a virtual world as the description of objects within a simulation, that in fact form a world not existing in reality. Such virtual worlds can be experienced through a multitude of mediums. Initially, virtual worlds were limited to text and document sharing such as in chat rooms and through conferencing systems. With the advancement of 2D and 3D graphics rendering technologies, a virtual world became a collection of virtual objects geometrically and physically described that can be experience using bi-dimensional displays as well as being experienced via virtual reality. Virtual worlds can be completely imaginary worlds made of non-existent objects as well being a faithful reproduction of an actual place that obey to same rules that the real world is subjected to. As Jaron Lanier said: "*It's a world without limitation, a world as unlimited as dreams. It's also a world that's shared, like the physical world*".

Today, virtual environments are very popular and describe both fantasy worlds—like in most video games—or worlds very similar to reality—like in simulators—. Beyond the creativity effort, the creation of a virtual world provides a series of challenges dealing with the description of the world as well as the modelling of the virtual objects and of their behaviours. Such descriptions must then be experienced through a medium, so, one of the challenge that engaged researchers in the past decades dealt with the simulation and the rendering techniques. The visual rendering is a complex operation usually requiring very high-performance machines and often exceeding even the performances achievable by available workstations even by several orders of magnitude. The complexity, in fact, increases with the level of visual fidelity called photo-realism at it's maximum expression.

2.2 IMMERSION AND PRESENCE

An immersive virtual environment is one in which the user's senses are enveloped with computer-generated stimuli. The VE is perceived by the user through natural *sensorimotor contingencies* (O'Regan and Noë, 2001). The theory's central idea is that vision is a mode of exploration of the world that is mediated by knowledge of what is call sensorimotor contingencies accounting for the phenomenal character of perceptual experience.

Immersion also describes the technical capability of the system to provide perceptual illusions to the user. The capabilities to naturally explore and perceive the virtual environment by walking and looking around are enabled by the sensing technologies. The better are the adopted technologies in terms of resolution, field of view, latency, sounds localization etc., the more immersive would result the user experience. It is therefore possible to classify systems according to the level of immersion the can provide to the user. A system A is more immersive than the system B if using the system A it is possible to simulate the perception afforded by the system B but not vice versa. Total immersion is what everyone making a VR system or application is aiming towards, making the virtual experience so real that we lose the connection with the real world and forget about the computer, headgear and accessories that are delivering the experience.

The state of being mentally immersed is often referred to as having a *sense of presence* within an environment (Slater, Frisoli, Tecchia, Guger, Lotto, Steed, Pfurtscheller, Leeb, Reiner, and Sanchez-Vives, 2007). Presence is a key factor when defining a virtual reality experience in terms of human experience. The presence is the subjective illusion of "being there" in the virtual environment despite of the certainty of being in another place different from what is presented on the displays. Sense of presence can also include the illusion of being physically immersed with own body into a medium, and is usually referred as "physical immersion". Physical immersion can be reached by providing synthetic stimuli—at different levels—to the body's senses; this does not imply all senses or that the entire body is immersed. The sense of presence is determined by many perceptual factors including inputs from sensory channels, as well as mental processes that elaborate incoming sensory data with current concerns and past experiences (Gibson, 1966). Perception is in fact an active process that combine sensory inputs with prior experience, expectations, and beliefs based on our previously existing model of the world.

According to Slater (2009) the sense of being there in a real place is referred as *place illusion*, and can happen even in a static unresponsive environment. Differently, the *plausibility illusion* refers to the illusion that what is apparently happening is really happening despite you know for sure that it is not. Key components of the plausibility illusion is that the environment responds to you as you would expect, the events that happens in the VE respond to you, are correlated with your actions and refer to your personality. It is important to realise that plausibility illusion does not require physical realism, however, if the situation being simulated has correspondence in reality, then it must meet the minima expectations as to how that reality works. For this reason, achieving plausibility illusion is an harder task than obtaining place illusion, cause it requires an higher knowledge of the laws that lie behind the simulated reality. Place and plausibility illusions can occur even in low-level systems. This happens because the brain "fills-in" details of the environment that are missing. After few seconds we are walking into an environment, we think that we know it. This happens because we focus on few key points building an inner model of the environment. In response to a situation, our brain infer a full model that is used to react to the situation starting from the perceptual model. Sutherland presenting his system capable of showing very simple wire frame scenes wrote: "Observers capable of stereo vision uniformly remark on the realism of the resulting images". VR provides enough cues for our perceptual system to infer a full model of the surroundings even if the visual stimuli are very simplistic, this is how VR works (Stark, 1995). In unmediated perception, presence is taken for granted and is related to the user's perception of the real world. In a mediated perception, instead, the user senses simultaneously the real world where he actually is and the virtual environment through the medium. The sense of presence related to the VEs is referred by many as *telepresence*, to underline the mediated nature of the virtual experience (Biocca and Levy, 2013; Sheridan, 1992).

In AR systems, we are interested in how to develop systems where the user loses the sense of mediation, and begins to respond to being immersed in a blended physical/virtual as if it was a “single world”. AR elicits a different sense of presence: “It is here” presence referred to virtual objects, or “You are here” referred to other users virtually co-located into our mixed world (Lombard and Ditton, 1997). Therefore presence in AR systems are totally related to the quality achieved by the seamless fusion of the two worlds. A seamless fusion must take into consideration visual and auditory aspects, but not less important an effective interaction between the virtual objects and the real world.

2.3 SENSORY FEEDBACK

Sensory feedback is an essential ingredient to virtual reality. The more sensory channels are effectively stimulated, the more the participant is immersed in the VE and the more impressive would be the experience or the more effective would be the learning in a training simulator. In order for the participant to be fully immersed in the scenario all the senses have to be properly stimulated. However, due to technological limitations, the level of accuracy of the feedback provided to each sensory channel can greatly vary. Sensory feedbacks must be consistent and synchronized for the user to perceive a coherent and predictable world. For instance in a roller-coaster simulator, the visual feedback must be synchronized to the aural one as well as to the fan which simulates the wind and the moving platform which simulates the variable direction and magnitude of the gravity force. Sensory conflicts resulting from fundamental limitations of the immersion system are important and typically not easily addressed. For example solving the visual/vestibular conflict experienced while a user is running inside a VE is not so easy as it would require a motion platform, and still it won’t be effective enough to avoid any cyber-sickness disturb. Similarly, if is possible to provide kinesthetic feedback, for instance by using exoskeletons (one hand only up to full body), however it is still difficult to provide haptic feedback with a good level of realism the let to feel the roughness of the surface in a very realistic way.

The most solicited sense in a VR experience is the visual one. For this reason, to provide a convincing experience the visual stream’s properties should match as much as possible the characteristics of the human sight in terms of field-of-view, resolution, dynamic range, latency and stereoscopy. Aural feedback is nonetheless important, cause the localization of the sounds helps the participant to understand the flow of the events. Hearing is arguably more relevant than vision to a person's "sense of space" and human beings react more quickly to audio cues than to visual cues. In order create truly immersive Virtual Reality experiences, accurate environmental sounds and physically-based audio rendering must be provided. However, providing effective aural feedback is not an easy task cause it means providing a feedback that simulate the actual physics laws, specifically sound should get louder as the object the produce it gets close to the user and the reverberation/echo effects must be accurately simulated according to the geometry of the VE and the materials of what it is made of. An effective binaural sound lend to a powerful sense of presence to a virtual world.

Tactile inputs such as omni-directional treadmills allow users to feel as though they're actually walking through a simulation, rather than sitting in a chair or on a couch (Darken, Cockayne, and Carmein, 1997; Wang, Bauernfeind, and Sugar, 2003). Haptic gloves open up the world of force feedback by allowing the user to pick up and feel virtual objects in a natural way. Being able to touch a virtual object, feel its size, shape, stiffness or roughness is a fundamental part of what makes an experience real. Researches found evidence of the importance of the role of tactile sensations in a VR experience. Some experiments asked participant to conduct a virtual exploration of virtual objects using a virtual hand with while exploring the real objects with the real hand at the same time. Adding physical objects that provide tactile feedback for actions turned out to increase the sense of presence in VEs (Hoffman, Hollander, Schroder, Rousseau, and Furness, 1998; Lok, Naik, Whitton, and Brooks, 2003). Unfortunately, existing gloves are still somewhat too bulky and heavy being made of a large number of actuators and sensors and cannot render the surfaces characteristics in a very realistic way. Recently however, haptic gloves are becoming smaller, lighter, and easier to use and control, and in a while they could become more common as human–machine interfaces (Blake and Gurocak, 2009).

Smell and taste senses, today, received less attention from the VR community, however they could really matter in some virtual experiences. Imagine the importance of this two senses in cooking simulator. Some devices to provide such feedback have been however developed. In recent years, different strategies to stimulate smell, touch or taste have been considered in order to enhance the VR experience. Dinh, Walker, Hodges, Song, and Kobayashi (1999) evaluated the effects of smell, touch, sight and hearing on the sense of presence, providing stimuli associated with specific objects in the VE like scent of coffee, sensation of air on the participants' skin by using a real fan or the feeling of the sun on the skin by using a light lamp.

Producing an effective sensory output for all senses involves, as thoroughly discussed, a clear understanding of the environment as well as of the user. Knowing the position of the user's head (nominally the eyes, ears and nose location) is fundamental to correctly calculate the visual, aural and olfactory stimuly, while the estimation of the whole body pose is important to provide haptic/kinesthetic feedback and for natural interaction. For this reason, a typical VR system will track the head and at least one hand or an object held by the hand of the participant, but many systems may track many of the major body joints.

2.3.1 *Latency*

An extremely important challenge faced when dealing with any AR/VR system is the minimization of the delays between the user's actions and the perception of the environment's reactions. Welch, Blackmon, Liu, Mellers, and Stark (1996) and Meehan, Razzaque, Whitton, and Brooks (2003) investigated the effects that delay in visual feedback have on presence in virtual environments. They found that a delay in visual feedback decreased the sense of presence, therefore keeping the "motion-to-photon" latency is extremely important for VR. Different tasks have varying requirements on the accuracy, speed, and latency of the tracking system's reports. Researchers estimate that the lag between when we turn our head and when the VR environment changes need to be kept under $90ms$, and preferably under $50ms$. If the lag is too high, the VR system induces a "swimming" feeling, and might make the participant disoriented and hamper the quality of interactivity. Lag can also be a problem for any motion tracking inputs such as controllers that measure our hand and arm movements. Latency minimization is also very important in augmented reality because the viewer has the instantaneous movement of the real world to compare against the augmented reality overlay. In a perfectly working AR system, a graphical overlay appearing on top of a physical object must stay locked to the object even when the user rotate his head so that it appears to be part of the real world.

Many are the sources of delays that contribute to the motion-to-photon latency:

- **Sensing delays**. The MEMS sensors may be bandwidth-limited and do not instantaneously report orientation or position changes. Similarly, camera-based sensors may exhibit delay between when the camera sensor receives light from the tracked object and when that frame is processed.
- **Processing delays**. Sensor data is often combined using some kind of sensor-fusion algorithm, and executing this algorithm can add latency between when the data is received and when the algorithm outputs the answer.
- **Data smoothing**. Sensor data is sometimes noisy and to avoid erroneous jitter, software or hardware-based low-pass algorithms are executed.
- **Transmission delays**. If orientation sensing is done on a device which is connected via USB to a workstation, there is some time between data collection by the host processor and the time data transferred over USB.
- **Rendering delays**. Rendering a complex scene takes some time for the graphics workstation and the resulting frame needs to be sent to the display.
- **Display lag delays**. Display lag is a phenomenon associated with some types of LCDs. It refers to latency, or lag measured between the time there is a signal input, and the time it takes the input to display on the screen.
- **Frame rate delays**. A 90Hz display shows an image every 11 milliseconds. Information that is not precisely current to when a particular pixel is drawn may need to wait until the next time that pixel is drawn on the display.

Some of these delays are very small, but unfortunately all of them add up. One common method to reduce the apparent motion-to-photon latency is using predictive tracking. Since there is some delay between the movement itself and when the information about that movement ends up on the screen, using an estimated future orientation and position as the data used in updating the display could shorten the perceived latency. The prediction of user's movements is unlikely to be 100% accurate every time. However, the more you know about the user's behaviour the more your prediction can be accurate. For instance, when doing head-tracking, understand how fast the human head can possibly rotate and the common rotation speeds are, can improve the tracking model. The most know prediction algorithms are dead reckoning, Kalman predictors and alpha-beta-gamma. Predictive tracking could help in reducing the apparent latency, however to avoid inaccurate results the prediction time must be kept low. The minimization of all the latency sources is still the preferable way to follow.

2.4 INTERACTION

The participant views the virtual environment from a first person perspective point of view. Unlike traditional media like television, VR makes one of his strengths in allowing the user to drive his experience by freely moving across the environment and interact with it. The "user" from being a "spectator" become a "participant" to the experience. Giving the ability to the user to freely move from place to place inside a virtual environment is a kind of interactivity. The point of view of the user in real-life is linked to the user's movements, equally the virtual environment is expected to move and rotate accordingly. A poorly designed human-computer interaction in VEs can prevent immersion to a great extent and broke the plausibility illusion. For this reason providing an effective for the user to naturally navigate the environment as he would do in real life is very important. Providing a joystick to the user to move inside a virtual room instead of allowing the user to naturally explore it by walking, moving and rotating his head could result in a poor experience with limited sense of presence, as well as to result physically disturbing cause of the disparity between the sight and the proprioception (aka motion sickness or cybersickness) (Whitton, 2003; Whitton, Cohn, Feasel, Zimmons, Razzaque, Poulton, McLeod, and Brooks, 2005).

Many VR experiences consist in static worlds that cannot be changed by the participant but only navigated; however, many more are dynamic and do allow modification. When the user can modify the environment then the simulation becomes even more interactive. In a VR simulator, the user performs the task while immersed in a virtual world that responds to his actions. VR interaction strives for a high level of fidelity between the virtual action and the corresponding real action being simulated. One of the reasons VR simulators are so costly is that the developer must model the behaviours of all the objects important for the simulation—whether autonomous or responding to user's actions—as well as to model the physics of the world. The most recurring questions when designing a *reality simulator* or a more limited *realistic simulator* is "How this object would behave in real life?" or "Is this object important to the simulation?". If the developer doesn't design a behaviour it won't be there, however, good design principles suggest not to add embellishments only because it is possible in order to avoid to distract the user from the task. As for the navigation tasks, providing a natural interface to interact with the virtual objects results in a higher sense of presence as well as more effective training (Brondi, Alem, Avveduto, Faita, Carrozzino, Tecchia, and Bergamasco, 2015): think about rotating a valve or opening a jar by using own hands (by means of tracked gloves) rather than using a tracked wand with a button to push. Immersive VR systems can become an important tool for training, simulation, and education when providing the high-fidelity interactions, especially to simulate tasks that are dangerous, expensive, or infeasible to recreate in real life.

NATURAL INTERACTION In traditional graphical applications interactions were typically achieved through abstract commands, metaphors used to specify coordinates and rotation angles. However in VR and AR is desirable for many applications to designing the interaction using real-world metaphors. Users may reach out a hand, grab an object, and move it around the virtual environment using natural, physical motions. Ideally, a participant should be able to interact with the VE by natural body motions and the VE system would understand and react to user actions. This has the obvious advantage that we all know how we can perform this type of control due to our everyday experience. The more familiar and realistic is the interaction with the VE the better will be the user experience as well as the higher will be the sense of presence and the self-awareness achieved (Brondi, Alem, Avveduto, Faita, Carrozzino, Tecchia, and Bergamasco, 2015). Natural interaction is not possible or efficient for many applications, however it is still one of the goals of the VR. Human-computer interaction in VEs can be strikingly different than traditional 2D or 3D interaction. Some virtual actions does not have no real action correlate, for example selection and deletion of virtual objects, so designing interaction metaphors that result to be natural is not an easy task.

2.5 APPLICATIONS OVERVIEW

There are many uses of VR and AR which range from academic research through to engineering, design, medicine and entertainment. This section presents a very brief overview—far from being an exhaustive enumeration—of the limitless applications enabled by VEs.

Virtual reality has been first adopted by the military for training purposes. This is particularly useful for training soldiers for combat situations or other dangerous settings where they have to learn how to react in an appropriate manner (Livingston, Rosenblum, Julier, Brown, Baillot, Swan, Gabbard, and Hix, 2002). One of the first and well-known training application are the flight simulators.

Examples in the industrial field can be found in the aerospace industry (De Sa and Zachmann, 1999), in the automotive industry (Li, Khoo, and Tor, 2003), in logistics (Bergamasco, Perotti, Avizzano, Angerilli, Carrozzino, and Ruffaldi, 2005) and, in general, in the sector of maintenance (Magee, Zhu, Ratnalingam, Gardner, and Kessel, 2007). VR has the potential to revolutionise the product and environmental design industry. It is possible to simulate and render all characteristics relevant to physical mock-ups generating digital "products" and allow the audience to interact with them to collect their reactions. Perhaps most exciting is the potential to rapidly iterate this product led by the audience's feedback from their hands-on experience. In complex manufacturing a considerable amount of resources is focused on training workers and developing new skills. Many projects involving the use of VR and AR to increase the effectiveness of manufacturing processes and reducing the investment required have been carried out (Cardoso, Prado, Lima, and Lamounier, 2017; Gonzalez-Franco, Pizarro, Cermeron, Li, Thorn, Hutabarat, Tiwari, and Bermell-Garcia, 2017) VR applications for safety have been already used in some fields. Van Wyk and De Villiers (2009) studied how VR applications could help miners to improve their safety in South African mines, using all the peculiarity of the natural environment. An AR prototype to support military mechanics conducting routine maintenance tasks inside an armored vehicle turret was developed (Henderson and Feiner, 2009). Within this categorization, assembly tasks have received the most attention. Caudell and Mizell (1992) proposed an AR prototype to assist in assembling aircraft wire bundles.

Many studies involved the use of VR or AR in the healthcare field. A popular use of this technology is in robotic surgery. This is where surgery is performed by means of a robotic device controlled by a human surgeon. VR has been also been used in the field of remote tele-surgery in which surgery is performed by the surgeon at a separate location to the patient. Fuchs, Livingston, Raskar, Colucci, Keller, State, Crawford, Rademacher, Drake, and Meyer (1998) developed an AR visualization system to assist with laparoscopic surgical procedures, and other examples that uses Mixed Reality to superimpose digital diagnostic images directly on the patient body exists (Lapeer, Chen, Gonzalez, Linney, and Alusi, 2008; Magee, Zhu, Ratnalingam, Gardner, and Kessel, 2007). VR and AR have also been proficiently used in the treatments of some phobias. Parsons and Rizzo (2008) used the virtual reality exposure to treat anxiety and specific phobias, and Juan, Baños, Botella, Pérez, Alcaníiz, and Monserrat (2006) for the treatment of acrophobia. PsyTech is creating an Anxiety Management Virtual Reality Platform for exposure therapy, creating a space for the player to go at their own pace within a secure environment[1].

Entertainment will likely be one of the first and strongest examples of the change virtual reality will bring to the industry, and gaming is one of the most obvious uses.

The potential of VEs for supporting education is widely recognized. Several programs to introduce students and teachers to the technology have been established. Professionals and researchers have striven to apply AR to classroom-based learning within subjects like chemistry, mathematics, biology, physics, astronomy, and to adopt it into augmented books and student guides (Chang, Morreale, and Medicherla, 2010; Fjeld and Voegtli, 2002; Freitas and Campos, 2008; Lee, 2012). It's now possible for museum spaces and schools to teleport students to specific moments in history, to allow them to experience being executed by a guillotine, take tours of space or even explore the depths of the ocean. Those interested in natural history will also soon be able to watch a VR nature documentary narrated by David Attenborough[2]

VR is being increasingly used in the field of scientific visualisation. This field is based upon using computer graphics to express complex ideas

---

[1] `http://psychologicaltechnologies.com/`

[2] `http://www.attenboroughsreef.com/vr_dive.php`.

and scientific concepts, for example molecular models or statistical results (Bryson, 1996; Cruz-Neira, Leigh, Papka, Barnes, Cohen, Das, Engelmann, Hudson, Roy, and Siegel, 1993). Hamdi, Ferreira, Sharma, and Mavroidis (2008) simulated bio-nano environments in VR, to allow the design and characterization—through physical simulation and 3D visualization—of the behaviour of protein-based components and structures.

AR has been used as an interactive tool in cultural heritage sites by showing visitors the original images of the sites and information about historical episodes happened in the places (Vlahakis, Karigiannis, Tsotros, Gounaris, Almeida, Stricker, Gleue, Christou, Carlucci, and Ioannidis, 2001). VR can be an extremely powerful tool for immersing people into new worlds and places, to see things they have never seen before. One example is a guided tour of Stonehenge visited using VR[3]. The VR demo allows you to set the time of day you take the tour, letting you experience the monument during sunset or even at night, giving this prehistoric monument a sense of beauty that even actual visitors have said can be lost in real-life.

---

[3] `http://www.voyagervr.com/stonehengevr/`

3

ENABLING TECHNOLOGIES

In this chapter we present an overview of the enabling technologies adopted for VR/AR devices and their applications. Conceptually, a minimal immersive VR system places a participant into a surrounding 3D world—overlaid to the real world in the AR case—that is delivered to a display system by a computer. The participant's head is tracked so that visual and auditory updates depend on head-position and orientation. The computer graphics of the system delivers perspective-projected images individually to each eye so that the user can perceive the scenario with the correct parallax. The goal of the hardware is to create what appears to be a life size, 3D virtual environment without the boundaries we usually associate with TV or computer screens. Finally the participant should be able to interact with the virtual world by means of tracked wands or cyber-gloves.

3.1 IMMERSIVE VR SYSTEMS

Ivan Sutherland, one of the originators of 3D computer graphics, was the first person to conceive and build an immersive VE system (Sutherland, 1968). There are two typical implementation of IVE systems: head-mounted displays (HMDs) and CAVE Automatic Virtual Environments (CAVEs). A complete taxonomy of VR systems was proposed by Muhanna (2015) and shown in figure 3.1 but only the two most common implementation are presented; presenting all the systems would have required alone a dedicated article of the same length of this dissertation.

An HMD is head-worn helmet provided with two small screens located a short distance from the user's eyes which delivers two computer-generated images, one for each eye. The two rendering of the VE are computed with respect to the position of each eye with appropriate perspective and together form a stereo pair providing strong depth perception. Between the two small displays and the eyes are placed some optics. The optics are typically magnifying lenses which have a double function: allows to focus on the display otherwise too close to the eye to result in focus and allows to expand the field of view of the headset. The displays are integral to the headset which exploits a mechanism to continually capture the position

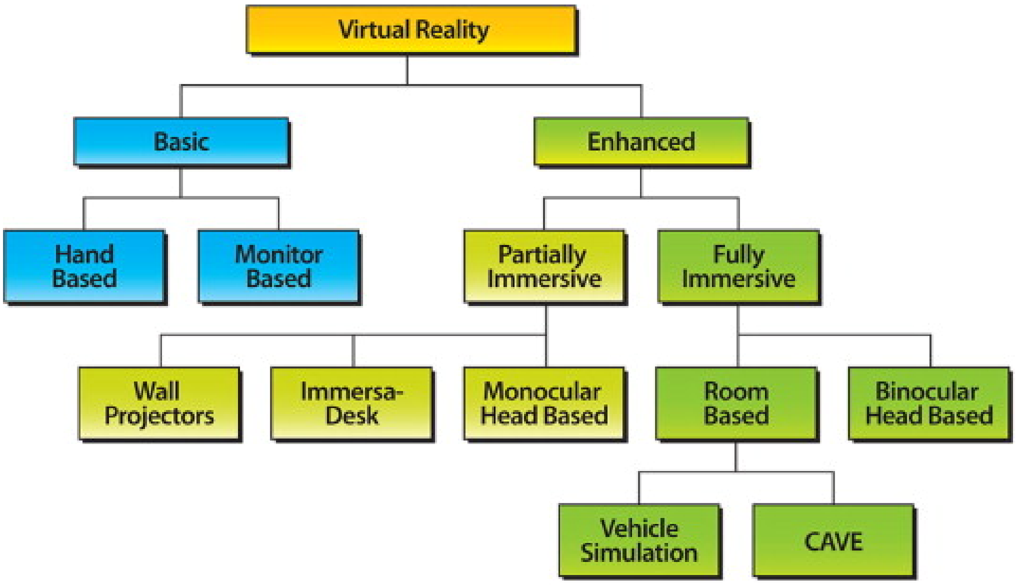


Figure 3.1: A taxonomy of virtual reality systems.

and orientation of the user's head, and—assuming that the eyes are looking straight ahead—the gaze direction. In this way the images can be computed to the actual pose of the "participant" inside the virtual world. Two fully IVE systems—a Virtual and a Mixed Reality systems—were developed and are presented in chapter 5

Another popular design of a fully immersive VR system is the CAVE™ developed by Cruz-Neira, Sandin, DeFanti, Kenyon, and Hart (1992). CAVE-like systems are rooms where users are surrounded by large screens onto which a nearly continuous virtual scene is projected in stereoscopy. Here, images are back-projected onto the walls of an approximately 3m cubed room, and usually front projected onto the floor by a projector mounted on the ceiling. Lightweight shutter glasses are needed to separate the stereo images for the two eyes. One perspective for each wall is calculated according even to head-tracking data. More than one person can be in the Cave simultaneously but the perspective is correct only for the head-tracked user or the one closest to the sweet spot where the views are calculated from. Compared to HMDs, CAVEs are far more expensive due to the cost of the stereo projectors, of the physical structure and of the projection screens (usually made of expensive materials with special diffusive properties) and of course not portable. Totally immersive 6-faces CAVEs ,like the HyPI-6, exists even if they are even more expensive due to the needing of back-project also the floor [1]. CAVEs are typically tailor-made, designed according

---

[1] `http://www.iao.fraunhofer.de/lang-en/presse-und-medien/277.html`

to the available space and top the desired specifications. CAVE became popular -even if not a mass product- in 1990s when HMDs due to the inadequate resolution of the available displays were not able to match VR requirements. On the contrary, CAVE could count on higher-resolution projectors and could exploit even more than a single projector for each wall. A 4 walls high-resolution CAVE exploiting 18 projectors and a cluster rendering architecture is presented in chapter 4.

3.2 AR SYSTEMS

VR systems use technology to replace reality and create an immersive environment. In contrast, the main goal of an Augmented Reality system is to enhance reality with digital content in a non-immersive way. To be fully immersive a VR system must have a wide field of view and the 3D graphics must be as realistic as possible. Contrary, an AR display can be non-immersive, have a small field of view and use minimal graphics. For example, an AR navigation application can work well even showing very simple maps and arrow graphics. Despite of the differences, most of the technologies that works for the VR also adapt to AR.

A taxonomy of VR systems was proposed by Billinghurst, Clark, and Lee (2015) and shown in figure 3.2 but only few common implementation are presented.

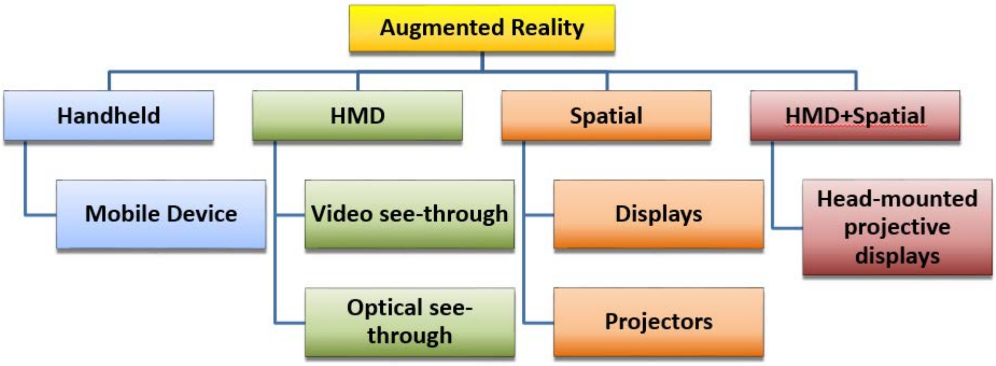


Figure 3.2: A taxonomy of augmented reality systems.

Video based AR displays use digital processes to combine virtual view images with video of the real world view. This type of display first capture the real world using a video camera system, so that the image of the real environment is augmented with the rendered image of the virtual scene. One of the common problems in composition of virtual and real world images is incorrect occlusion between real and virtual objects due to virtual scene image being overlaid on top of the real world image. Video based AR displays can easily solve this problem by introducing depth information from the real world, performing depth tests between virtual and real elements to show only what can be really seen. This technique are used by many AR systems, however the most interesting—in my opinion—are hand-held devices and video see-through (VST) head mounted display. Fuchs, Livingston, Raskar, Colucci, Keller, State, Crawford, Rademacher, Drake, and Meyer (1998) conducted early research on VST-HMDs and developed a system to help surgeons to perform laparoscopic procedures. Hand-held devices like smartphones are nowadays pervasive, it makes them a promising to bring low-cost AR to masses.

Another category shares the same form factor of the video see-through HMDs but instead uses optical elements to allow the user to directly perceive the real world—not captured using a camera—but augmenting his view with artificial contents. Optical see-through (OST) HMDs are becoming the most advanced and interesting form of AR. Microsoft recently released the first self-contained wearable AR OST-HMD—the HoloLens—and many big companies are going to release their devices very soon. A more in depth analysis of this kind of devices will be given in the chapter 6.

The last category I want to mention is called "spatial AR" (Bimber and Raskar, 2005). It consists in augmenting the real world by projecting virtual contents on top of it by using projectors. The projection could be both computed from a fixed point of view as well as dynamically calculated to follow a moving user's.

3.3 VISUALIZATION

Because of the pervasive, dominant role of vision in human perception, visual stimuli are for sure the most important component of the VEs computer-based illusion. For a visual stimuli to became effective the perception of three-dimensional depth is important. There are two types of depth perception cued depending on whether they are apparent when on or both eyes are used: monocular or binocular. The former provides only weak depth perception, and is based on perspective, motion parallax, relative size of known objects, highlights and shadows cues. The latter instead provide a strong depth perception and is based on the convergence angle and the objects disparity between eyes. The computer-generated images must then respect all these cues to provide a convincing visual 3D experience. All VR systems, indeed, provide a different image for each eye calculated from a slightly different point of view (according to the user's inter-pupillary distance). Image resolution is another important aspect to achieve photorealism. The human eye has a visual acuity of about 1 arc minute, meaning that it can resolve up to "$60 pixels$" per degree at the fovea. However currently HMD display are still far from that value (HTC Vive $\sim 11 arcmin$). The visual acuity is not uniformly distributed among all the retina; the center part is the most sensitive, so providing uniform high-resolution images result to be a waste of computational power. For this reason, thanks to the recently advancement in eye-tracking technologies a method called "foveated rendering" is becoming more and more popular. This techniques consist in rendering different portions of the image with different resolutions. The part of the image where the center of the retina is focused on can be rendered at maximum resolution, while the peripheral parts with lower resolution, allowing to save precious computational power with minimal or no impact on user's visual perception. As screens get better and better, we will get increasingly closer to eye-limiting resolution in the headset and thus closer to photo-realistic experiences. To provide a fully immersive experience another fundamental aspect is to completely cover the user's field of view with synthetic images. Humans have a slightly over 200° horizontal field of view, but only about 114° is binocular[2]. Some VR systems are able to fully cover human's binocular view (like CAVEs) while others are able to provide a wide field of view even if not total (like HMDs). Most of the nowadays available HMDs—like Oculus and Vive—have an average horizontal FOV of 90° but only about 75% is binocular, while others—like StarVR—are pushing this limit up to $150° - 210°$ of which 100° are binocular[3]. Finally another important factor to take into consideration is the refresh rate: VEs are experienced in an interactive way, so the environment must change at an interactive speed. The lower limit for an experience to be perceived as interactive is fixed to 30 Hz by many (Brooks Jr, Marcus Brown, Burbeck, Durlach, Ellis, Lackner, Robinett, Srinivasan, Sutherland, and Urban, 1992). However, the current generation of HMDs is capable of 90 Hz-120 Hz to improve the motion perception as well as for minimizing the overall latency which is fundamental for any VR experience. Another important feature

---

[2] `https://vr-lens-lab.com/field-of-view-for-virtual-reality-headsets/`

[3] `http://www.starvr.com/`

that VR/AR OLED and LCD displays introduced in the last years is the low-persistence. A full-persistence display has its pixels lit all the time, showing the correct scene orientation for one point in time. In an interactive experience typically the user continues to move their head, so the scene orientation is out of date until the next frame can be drawn. The low-persistence technique lights the pixels only when the scene orientation is correct and goes dark immediately thereafter. Thanks to a high refresh rate, this happens so quickly that the user sees one continuous image. The end result is significantly reduced motion blur and potentially less motion sickness.

All these technologies are valid both for VR and AR systems. VR technology is probably, at present, more mature than the AR, as the technological demands for AR are much higher (Azuma, Baillot, Behringer, Feiner, Julier, and MacIntyre, 2001; Azuma, 1997; Van Krevelen and Poelman, 2010), which might be the main reason why the field of AR is taking longer to mature. An additional challenge only related to AR displays is the seamless fusion between the virtual and the real worlds. AR systems are mostly based on head-mounted displays. Two are the classes of AR systems: video see-through, or optical see-through. Video see-through HMDs use cameras to acquire the real world from points of view as close possible to the user's eyes positions. The real world acquisition is digitally fused with the computer-generated virtual world and presented to the user on a standard display, so in this way can be considered as a VR system displaying AR contents. Optical see-through displays instead allow user to see the augmented overlaid on top of the real environment which is directly perceived with own senses. Each approach has its strengths and weaknesses: in the video see-through case the perception is mediated and the quality of the experience strictly depends on the technology adopted, while in the unmediated case the perception of the environment is maximum but the quality of the augmented contents are related to the display/optics technology which is more complex and has its weakness.

Transparent light-additive displays used in commodity AR headsets lack the capability of selectively mask the real environment, have a significant shortcoming: the augmented contents are affected by the real world lighting conditions. For instance, black areas in the virtual objects appear to be transparent, resulting in synthetic objects to appear as translucent poorly contrasted "ghosts". Furthermore, differently from what is possible with video see-through headsets, it is not possible to mask or replace real objects unless the overlay is far brighter than the real environment. The current trend is to adopt optical see-through displays like in the case of the Microsoft's Hololens. In chapter 6 we analyse more in details the problem and propose a solution.

3.4 TRACKING

Tracking is the determination of an object's position and orientation. Common objects to track include the participant's head, participant's limbs, and interaction devices (such as gloves and wands). Most tracking systems have sensors or markers attached to the objects and other external devices that track and report the position and orientation of the sensors. Commercial tracking systems employ one or a combination of optical, mechanical, magnetic, acoustic, inertial and GPS approaches. Each method has different advantages with respect to cost, speed, accuracy, robustness, working volume, scalability, wirelessness, and size. A brief description of the characteristics of the main tracking technologies can be summarized as follows:

- **Optical tracking** come in two variants: inside-out or outside-in. The first approach consists the cameras on the HMD looking at external LEDs or markers placed on the environment in a known configuration(3rdTech Hiball[4]). The second approach puts the cameras fixed in the environment, and move the LEDs or markers onto the tracked body in a well known spatial configuration. In both cases, the projections of the LEDs/markers on the cameras image planes contain enough information to uniquely identify the position and orientation of the tracked body. Optitrack[5] systems and the WorldViz PPT[6] are commercial products exploiting this approach. Optical trackers in general have high update rates (60 Hz-240 Hz) and the short lags (4 to $20ms$). The resolution range from few millimetres up to a submillimetre accuracy in the most performing models with a typical 0.1° accuracy in orientation. However, they suffer from the line of sight problem: for some positions some LEDs/markers are not visible leading to performance degradation. as possible these causes of uncertainty.

---

[4] Welch, Bishop, Vicci, Brumback, Keller, and Colucci, 2001.

[5] `http://optitrack.com/hardware/`

[6] `http://www.worldviz.com/virtual-reality-motion-tracking/`

A particular case of inside-out tracking adopt computer-vision algorithms to extract visual features from the environment that are used to compute the pose of the head inside the real environment instead of looking at LEDs or markers. This approach does not require the instrumentation of the environment, while result much more computational demanding as well as less robust. A particularly effective implementation of this technique can be found in the Hololens and will be adopted by some VR platforms in the near future.

- **Inertial Tracking** Inertial tracking use data from accelerometers and gyroscopes. Accelerometers measure linear acceleration while gyroscopes the angular velocity. The output of the accelerometer could be integrated to find the velocity and then integrated again to find the position relative to some initial point. Angular velocity can be integrated as well to determine angular position relatively to the initial point. Modern inertial measurement units systems (IMU) are based on MEMS technology allows to track the orientation with very high update rates (typically $1kHz$ and minimal latency ($<1ms$). Because these systems measure relative positions instead of absolute positions they can suffer from accumulated errors and therefore are subject to drift. Most IMUs, nowadays, integrate a magnetometer which is used to compensate the orientation drift resulting from the integration.

- **Magnetic Tracking** Electromagnetic tracking devices function by measuring the intensity of the magnetic fields generated by sending current through three small wire coils, oriented perpendicular to one another. These coils should be put in a small housing mounted on a moving target which position is necessary to track. The current has the effect of making each wire into an electromagnet while the current is flowing through it. By sequentially activating each of the wires, and measuring the magnetic fields generated on each of three other perpendicular wire coils, it is possible to determine the position and orientation of the sending unit. The system works poorly near any electrically conductive material, such as metal objects and devices, that can affect an electromagnetic field. Another disadvantage to these tracking devices is that the working volume tends to be rather small. Magnetic tracking has been implemented by Polhemus[7], Ascension

---

[7] `http://polhemus.com/motion-tracking/all-trackers/fastrak`

trakSTAR™[8]) and in Razor Hydra by Sixense. Their systems provide low latency ($\sim 5ms$), high update rate (60 Hz-120 Hz) and a pre4cision of few millimetres depending on the distance of the sensor from the sending unit.

- **Acoustic Tracking** Acoustic tracking devices use ultrasonic (high-frequency) sound waves for measuring the position and orientation of the target object. There are two ways of doing this: so-called time-of-flight tracking and phase-coherence tracking. Time-of-flight tracking works by measuring the amount of time that it takes for sound emitted by transmitters on the target to reach sensors located at fixed positions in the environment. The transmitters emit sounds at known times, and only one is active at a time. By measuring when the sounds arrive at the various sensors the system can calculate the distance from each sensor. Sensors are arranged in a known configuration it is possible to uniquely determine the position and orientation of the emitter. Time-of-flight trackers typically suffer from a low update rate, brought about by the low speed of sound in air. Phase coherence tracking works by measuring the difference in phase between sound waves emitted by a transmitter on the target and those emitted by a transmitter at some reference point. Like optical systems, acoustic ones requires a direct line of sight between emitters and receivers. Since phase coherence tracking works by periodic updates of position tracking devices are subject to error accumulation over time.

- **Hybrid systems** Because every technology has its pros and cons, most systems use more than one technology. A system based on relative position changes like the inertial system needs periodic calibration against a system with absolute position measurement. Systems combining two or more technologies are called hybrid positioning systems. Intersense IS-900[9], for instance, exploit a combination of the very high speed and low latency inertial tracking but compensating for the accumulation errors by using acoustic tracking. Similarly most cutting-edge VR HMD—like Oculus and Vive—exploit the very fast but subject to drift inertial tracking combined with a very accurate but with lower update speed optical tracking.

---

[8] `https://www.ascension-tech.com/products/product-history/`

[9] `http://www.intersense.com/pages/20/14`

All these technologies can be used to track user's body as well as arbitrary objects. An effective head-tracking is the most important part for a VR experience and present very strict requisites in terms of accuracy, speed and latency. An inaccurate or delayed tracking will lead to a poor VR experience causing cyber-sickness to the users. However the precision requisites of VR systems are lower than AR ones. In an AR application, the tracking must be as accurate as possible to create the illusion that the virtual content is fixed in the real world. In an optical see-through AR display it it very easy for the human eye to perceive a mismatch between real and virtual elements of even a few millimetres.

Tracking the participant's limbs allows the VR system to a virtual representation of the user within the virtual environment increasing the participant's sense of presence. The accuracy and speed requirements for limb tracking are typically lower than that of head tracking. Many full-body motion capture suits are commercially available, most of them uses a inertial or optical tracking technologies. Finally object tracking, usually accomplished by attaching a sensor, allows a virtual model of an object to be registered with a physical real object. Studies reported an increased sense of presence when manipulating virtual objects having a physical counterpart in the reality that match in terms of shape and appearance (Lok, Naik, Whitton, and Brooks, 2003). Since humans use their hands for many interaction tasks, tracking and obtaining inputs from a hand-based controller was a natural evolution for VR controllers. Furthermore, tracked gloves can also report fingers pose and gestures, allowing for a natural interaction with the virtual objects similar to what happens in real life.

Eye-tracking is the final piece of the VR tracking technologies. Only few commercial headset exploit this technology like the FOVE[10]. A combination of infra-red emitters and cameras monitor's allows to accurately estimate the gaze direction inside the VE. The main advantages consists in the possibility to make depth of field more realistic, in standard VR headsets in fact, everything is in pin-sharp focus while what happens in reality is that objects lying on the focus plane are in focus while the background and the foreground appears blurred. Eye tracking allows to simulate this effect with good approximation. Other advantages of knowing the gaze direction consists in allowing in-game characters to more precisely react to where you're looking and improving graphical performances by using the technique called foveated-rendering.

---

[10] `https://www.getfove.com/`

3.5 INTERACTION

Interacting with own hands is probably the most difficult task for VE. The difficulty is in constantly capturing the exact pose of the user's hands and at the same time simulate their interaction with the virtual objects in a physically credible way. Many researcher worked on the development Natural User interfaces (NUIs) for VEs, and the use of sensorized gloves able to provide finger-tracking to VR applications are a viable option (Bowman, Wingrave, Campbell, and Ly, 2001; Buchmann, Violich, Billinghurst, and Cockburn, 2004; Maggioni, 1993; Popescu, Burdea, and Bouzit, 1999; Popescu, Burdea, Bouzit, and Hentz, 2000). A tracked glove reports position and pose information of the participant's hand to the VR system. They can also report hand gestures that can be associated with virtual actions such as grasping, selecting, translation, and rotation. Tracked gloves provide many different kinds of inputs and most importantly, are very natural to use. Leap Motion[11] is a small sensor that can be attached to the front of the HMD. Using two monochromatic IR cameras and three infrared LEDs, the device observes a roughly hemispherical area, to a distance of about 1 meter. Despite its tiny size, it enables real use of your hands in VR by allowing 10-fingers tracking and interaction with objects as you would in real life. Depth cameras are another kind of devices used in the development of natural interactions. Placed in a way that the whole body can bee seen, they allows for a simple—even if not very accurate—estimation of the user's body pose, which can be used for a body interaction with the VE. However, due to the limited resolution, the finger tracking is not achievable. Section 5.2 presents our natural interaction solution exploiting a depth camera to interact with virtual objects while immersed in a VE. Other input methods can include anything from hooking a controller or a joystick, voice controls or smart gloves. A compromise to get ease of use and proper feedback is to engineer a specific device to interface with the VR system. For example, Bajura, Fuchs, and Ohbuchi (1992) developed an ultrasound AR surgery system attaching a tracking sensor to an ultrasound wand. This enabled the AR system to provide a natural interface for training and simulation. However, this required developing software and manufacturing specific cables to communicate between the ultrasound machine and a PC. Creating these specific devices is time consuming and the resulting tools are usable for a limited set of tasks.

---

[11] `https://www.leapmotion.com/product/vr`

Navigation is another important interaction task of VEs, and is generally split into two subtasks: short range navigation takes place by direct tracked movements—user can move by naturally walking inside the VE—whilst long-range navigation happens by means of interaction devices (Slater, Usoh, and Steed, 1995). Navigation metaphors are needed in order to map raw data coming from such devices into opportune movements in the VE. This means that different navigation strategies can be implemented on top of the same device, or that the same navigation strategy can be implemented on top of different devices. A multitude of them has been proposed in the literature, many of them addressing the problems linked to the use of 2D input devices for a 6-DOF task such as navigation. In Ware and Osborne (1990), a set of basic metaphors are presented, each adaptable to different interaction devices. In general there is not a universal metaphor fitting in every application or applicable to every device, as each of them has specific advantages. Many common navigation tools require at least one hand to be operated thus limiting the possible bi-manual interaction with the environment. As this can be undesirable, hands-free, body based navigation tracking the user body can be instead used to navigate the environment. In this category fall approaches based on the use of treadmills, such as the torus treadmill (Iwata, 1999) or the CyberWalk (Souman, Giordano, Schwaiger, Frissen, Thümmel, Ulbrich, Luca, Bülthoff, and Ernst, 2011), or other types of locomotion interfaces such as the GaitMaster (Iwata, Yano, and Nakaizumi, 2001) or the more recent Virtuix Omni[12]. Interestingly all of these interfaces add to the interaction a more or less realistic inertial/haptic feedback related to gait. Simpler devices do not provide such feedback and their use is limited to gait detection. Numerous examples can be found in walking-in-place metaphors making use of sensors of different type as in Templeman, Denbrook, and Sibert (1999) and Feasel, Whitton, and Wendt (2008). Efforts have focused recently on the use of low-cost interaction devices to detect gait, like the Microsoft Kinect (Zheng, McCaleb, Strachan, and Williams, 2012) or the Wii Balance Board (Williams, Bailey, Narasimham, Li, and Bodenheimer, 2011). Being the Nintendo Wii Balance Board equipped with four independent pressure sensors, it allows for more than simple walking-in-place (Haan, Griffith, and Post, 2008; Hilsendeger, Brandauer, Tolksdorf, and Fröhlich, 2009), enabling for instance navigation based on leaning on the board. In fact, foot-based interfaces have the potential to allow for rich, natural and easy

---

[12] `http://www.virtuix.com/`

to learn/remember input actions, including feet position and orientation (both absolute and relative), relative force (for instance shifting weight from one foot to another), strokes (similar to finger gestures) or taps. In order to be able to exploit this potential, specific devices - such as sensor mats - have to be built so as to quickly scan the spatial distribution of applied force. Unfortunately, commercially available spatial sensor mats are mostly used as medical devices and are relatively expensive. In addition, they are sold as integrated hardware/software systems, limiting the ability to add custom foot gesture recognition to systems based on these products. In the appendix A an innovative pressure-sensitive carpet-like device developed is presented.

4

PROJECTOR-BASED IVES

Nowadays, VR applications are increasingly demanding in terms of computational resources. They may require performances exceeding the computational power that a single workstation, even if exploiting multiple processors and multiple graphics cards, is able to deliver. Additionally, immersive visualization systems requires multiple rendering passes to handle stereoscopy.

We developed a software architecture, called *XVR Network Renderer*, that takes advantage of a network of calculators to perform "cluster rendering". Each cluster's node takes care of a subset of the rendering task, thus allowing large output resolution and multiple channels to be handled without requiring high-end or dedicated hardware. We ensured that our solution works using commodity hardware: the various calculators involved are not required to be identical, are connected by means of ordinary LAN devices, and may be safely equipped with low-end graphics cards.

The Network Renderer can be seen as a virtual OpenGL context with very high capabilities, completely transparent to the original application. The virtualization of the graphical context is performed by a software layer that intercepts all the OpenGL API calls issued by the original application, called *master* application, and sends them to a set of programs, called *slaves*, running on the networked machines. In our approach, the master workstation is in charge of distributing the rendering load among the slave workstations directly connected to output devices. In this way the rendered images do not need to be sent back to the master node, which can be a reasonable limitation when the resolution and the number of the managed displays grow up.

The Network Renderer have been designed with complex immersive visualization systems in mind. In particular, we have targeted PowerWall-like [1] and CAVE-like (Cruz-Neira, Sandin, DeFanti, Kenyon, and Hart, 1992) setup. This means that our system is able to handle many of the usual real-life problems that arise when using multiple projectors together, such as adjusting brightness and colours and handling overlapping regions.

---

[1] University of Minnesota. *PowerWall.* `http://www.lcse.umn.edu/research/powerwall/powerwall.html`. [Online; accessed 21-February-2016]

Furthermore it is possible to distort the projected images in order to exploit non-perpendicular projection or curved surfaces. A perspective correction can be performed allowing also to add head tracking capabilities to every application. This viewpoint's change is performed independently by the slave programs, without affecting the computational resources of the master machine. The Network Renderer configuration is centralized on the master machine.

MOTIVATION AND CONTRIBUTIONS We wanted to build a large CAVE system to be used as platform to conduct our studies, however available solutions were not flexible and open enough to give us a fine-grain control on many aspects. Some solutions force to develop application using specific frameworks, while others require significant additional effort to a traditional application into an immersive one. For these reasons we decided to develop a new architecture, highly configurable which give us the full control of the system. The result is a cluster-rendering architecture which is able to run—ideally—any OpenGL application without requiring any modification to it on a variety of visualization systems—among which are CAVEs–. The system has been used in different EU projects (BEAMING[2], VERE[3]).

4.1 RELATED WORK

The rendering phase is notoriously one of the most demanding operations from a computational point of view of a graphical application, especially when photo-realistic quality and high frame rates are required. Virtual Reality is a multi-modal interaction with dynamic and responsive virtual environments. Providing immersion to the user is usually achieved by covering his field of view typically surrounding him with several displays, or putting displays near his/her eyes. Both the increasing number of managed displays and the real-time constraints further increase the computational load requirements. A considerable number of solutions have been developed to manage applications with high rendering-load (Crockett, 1997; Molnar, Cox, Ellsworth, and Fuchs, 1994). A typical strategy to approach the problem is *divide et impera*, that is splitting the rendering task into subsets

---

[2] `http://beaming-eu.org/`

[3] `http://www.vereproject.eu/`

and processing each of them concurrently. This approach is usually called "parallel rendering".

A possible classification of parallel rendering systems is based on the employed hardware components. Systems having multiple graphics pipelines inside a single calculator are called multipipe rendering systems (Molnar, Eyles, and Poulton, 1992). Conversely, systems employing a cluster of networked calculators, each with its own pipeline working concurrently with the others, are called cluster rendering systems (Corrêa, Klosowski, and Silva, 2003; Humphreys, Eldridge, Buck, Stoll, Everett, and Hanrahan, 2001; Humphreys, Houston, Ng, Frank, Ahern, Kirchner, and Klosowski, 2002); in this case, network communication has to be used to perform task assignments and combination of the various outputs, in order to obtain the final image.

Clusters have long been used for parallelizing traditionally non-interactive graphics tasks, but in the last years, there has been growing interest in using clusters for interactive rendering tasks. Thanks to the impressive performance improvements of commodity graphics hardware in recent years and to the appearance of networks with gigabit bandwidth, cluster architectures have become a valid and cost-effective alternative to former proprietary multi-processor systems.

WIREGL In 2001 Stanford University developed WireGL (Humphreys, Eldridge, Buck, Stoll, Everett, and Hanrahan, 2001), a scalable platform for cluster rendering of graphics applications. WireGL provides the OpenGL API to each node in a cluster, virtualizing multiple graphics accelerators into a sort-first (Molnar, Cox, Ellsworth, and Fuchs, 1994) parallel renderer with a parallel interface. In the sort-first architecture, primitives are early distributed in the rendering pipeline—during geometry processing—to processors that can do the remaining rendering calculations. This generally is done by dividing the screen into disjoint regions and making processors responsible for all rendering calculations that affect their respective screen regions.

WireGL follows the well established client-server approach: one or more clients send OpenGL commands to one or more servers, called *pipeservers*. Pipeservers follow the sort-first approach, and collectively manage the rendering of the whole image. Each pipeserver exploits the capabilities of his own graphics hardware and is linked to all clients through a high speed network. The image is split into several tiles, and each server manages one or more of them. The final image is obtained re-assembling the output from each pipeserver. Without special hardware to support image reassembly, the final rendered image must be read out of each local framebuffer and redistributed over a network. A simpler and more efficient way to perform the reassembling operation is to make each pipeserver deal with a single display for each of the partitions it manages. The displays can then be physically joined together in order to obtain the final image. This kind of approach may lead in principle to a non-balanced distribution of the rendering load between the graphics servers. To partially solve this problem, a number of algorithms have been developed. They can be executed by a dedicated module, that could be implemented both in software and in hardware, as for the Lighting-2 system (Stoll, Eldridge, Patterson, Webb, Berman, Levy, Caywood, Taveira, Hunt, and Hanrahan, 2001).

CHROMIUM Chromium (Humphreys, Houston, Ng, Frank, Ahern, Kirchner, and Klosowski, 2002) is a further development of WireGL. Chromium inherits from WireGL the codification used to store the OpenGL commands, the interception mechanism and the client-server architecture. It allows to perform more transformations on API streams, and to arrange cluster nodes in a more generic topology than WireGL's many-to-many-to-few arrangement. The Chromium user decides which nodes of the cluster are involved in a given distributed rendering session, and what kind of communication they are going to use. These parameters are specified through a centralized configuration system, represented as an acyclic graph. Each node of the graph represent a computer of the cluster, whereas the edges symbolize the network traffic. Each node is made of two parts: *transformation* and *serialization*.

Transformation is performed by modules called *Stream Processing Units* (SPUs); they specify how to modify an OpenGL call sequence, in a completely configurable way, usually carried out according to particular stream processing algorithms. This operation generates one or more different sequences of OpenGL commands. SPUs are implemented by a runtime library, in the same way as the WireGL driver.

Serialization is the elaboration of one or multiple command sequences, in order to generate a single output stream. The whole system is initialized by a special component called mothership. It accomplishes the task of configuring and managing the Chromium processes, and it is capable of dynamically reconfiguring the system's components. It also manages the resource distribution and verifies that every SPU chain and every network connection are created as requested by the application. Mothership configuration is not only a matter of setting number of parameters but it requires a dynamic scripting language, making the configuration a demanding task.

MIDDLEVR AND BLENDERVR MiddleVR (Kuntz, 2015) and BlenderVR (Katz, Felinto, Touraine, Poirier-Quinot, and Bourdot, 2015) use a different approach: instead of distributing the graphical commands, the application is distributed and executed simultaneously on different nodes connected to different output devices. The applications are kept in sync to allow a continuous and consistent immersive visualization. The applications' scene-graphs are kept in sync between the multiple instances of the same application by distributing the state changes of each object from the master to the slaves applications. Master/slave synchronization is carried out at each frame. Hardware-level synchronization mechanisms between the workstations allows to synchronize the displays' outputs. These frameworks help the developers to keep the the scene-graph consistent with a minimal effort. However according to the application's logic this effort could became very demanding.

4.2 SYSTEM DESCRIPTION

The main purpose of *XVR Network Renderer* (Marino, Vercelli, Tecchia, and Gasparello, 2007) is to perform OpenGL cluster rendering exploiting the sort-first approach (Molnar, Cox, Ellsworth, and Fuchs, 1994) using a LAN. It employs a single-master multiple-slaves architecture. The system consists of a single module called *Network Driver* (see section 4.2.1) running on the *master* workstation, and one or more *slave programs* (see section 4.2.2) running on the same and/or on other workstations. All the machines are connected to the same local network. The module running on the master node intercepts all the OpenGL API calls, executes them locally and sends them to the slave programs through the network. The slave programs perform the rendering tasks according to a centralized configuration (see section 4.2.5); they may also perform additional operation on their outputs (see section 4.2.6). Master and slaves are synchronized on a per-frame basis (see section 4.2.4). Intercepting graphical commands does not require any modification to the OpenGL master application. Our approach is based on sending graphical commands instead of rendered images over the network. This typically lead to lower bandwidth usage without quality loss: the approach of sending high resolution pre-rendered images would require an extremely high-performance network or a lossy images compression.

OpenGL APIs are constantly evolving, so, although all functions can be potentially intercepted by the driver, not all functions are currently managed. During the development of the system we have focused on supporting applications developed using XVR. XVR is a flexible and efficient framework, that allows to easily develop Virtual Reality applications while maintaining a fine-grained control on the basic aspects of visualization and interaction (Tecchia, Carrozzino, Bacinelli, Rossi, Vercelli, Marino, Gasparello, and Bergamasco, 2010).

The XVR Network Renderer has been carried out in order to provide a cluster rendering architecture suitable to manage visualization systems for Virtual Environments. This kind of facilities can greatly vary in terms of number, kind and physical arrangement of the output devices, and consequently of required computational power. The architecture can scale from a single display setup, to a PowerWall exploiting multiple displays, up to big CAVE-like systems exploiting multiple walls enlightened by multiple projectors. The XVR network renderer has been also successfully used to implement global illumination in CAVEs architectures (Mortensen, Yu, Khanna, Tecchia, Spanlang, Marino, and Slater, 2008).

The network renderer has been employed in a number of EU projects (Normand, Spanlang, Tecchia, Carrozzino, Swapp, and Slater, 2012; Pérez Marcos, Solazzi, Steptoe, Oyekoya, Frisoli, Weyrich, Steed, Tecchia, Slater, and Sánchez-Vives, 2012; Steed, Tecchia, Bergamasco, Slater, Steptoe, Oyekoya, Pece, Weyrich, Kautz, and Friedman, 2012).

4.2.1 *Network Driver*

The Network Driver is the module running on the master node, performing different tasks. First of all, it intercepts all the OpenGL API calls performed by the master application. Similarly to WireGL and Chromium, the interception mechanism relies on a fake OpenGL dynamic library deployed on the master machine instead of the true one.

The collected commands are passed to a module called *packetizer*, which is in charge of encoding them and storing the commands into a buffer. The encoding is performed in a custom, optimized way, by assigning a unique

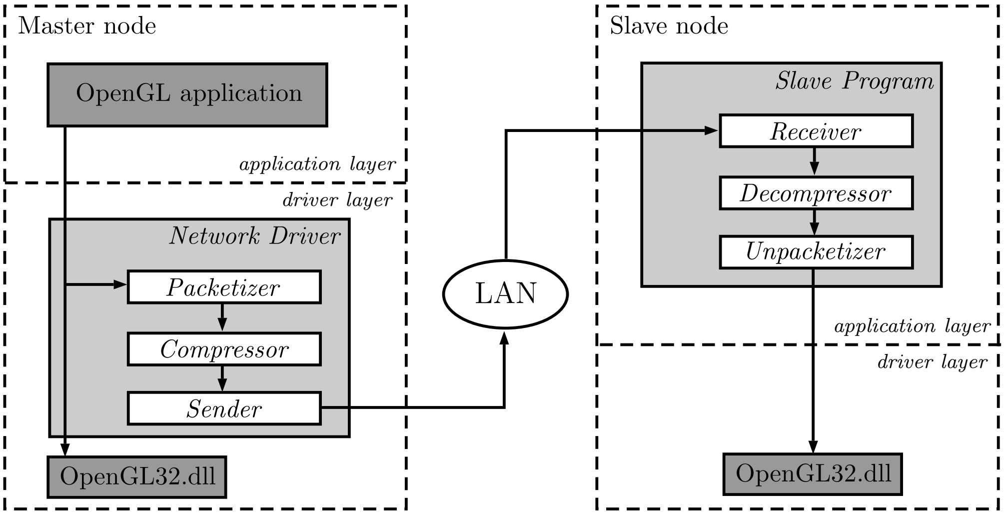


Figure 4.1: Overall scheme of the system's architecture.

identifier to each function and attaching its parameters. Every time a frame ends or the buffer exceeds a threshold , the data are sent through the network to a cluster of slave hosts (see section 4.2.3), in order to get them actually executed. OpenGL is a state machine (Shreiner and Group, 2009). In order to keep consistent the OpenGL state, the graphical commands are executed on the master node too. To be noted that the resolution of the master OpenGL context is totally unrelated from the slaves ones; this allows the master node to render the contents at low resolution to limiting the performance hit due to this "useless" visualization.

In principle, the Driver may also filter or modify the intercepted commands according to some user-defined criteria or in order to reduce network load. Due to the fact that the task of maintaining the OpenGL state consistent is performed by the master node, some of the commands—like all the *glGet* API functions—can be safely executed only on the master node. This allows to save network bandwidth and computing performances by avoiding to send such commands to slave nodes. To further reduce the network load, the Driver exploits data compression; details are provided in section 4.3. Low network utilization and system scalability are ensured by using broadcast or multicast addresses: the network load remains the same, except for the synchronization overhead, when the number of the slave nodes increases. The overall working scheme of the cluster rendering mechanism employed by XVR Network Renderer is shown in Figure 4.1.

4.2.2 *Slave Programs*

Every host of the cluster manages the information exchange with the Network Driver through a slave program. Each slave is identified by a unique name within the cluster. The slave program listens on a connectionless socket and waits first for the cluster configuration sent by the master, then for the OpenGL commands stream. Each slave initialize its own graphical context by extracting its configuration from the cluster's one (see section 4.2.5). After the initialization phase, the slave starts to decode the received command stream and to execute them locally. The slave program may be configured to perform additional calculations such as camera transformations or attenuation and distortion effects (see section 4.2.6).

Each slave's OpenGL context may have a different resolution and exploit different camera transformations. The most useful implication of this feature is that each slave may be configured to replicate just a subset of the master's context, as well as to handle the whole or a part of a bigger virtual context, provided that the original application does not perform any pre-culling phase. Combining the resulting rendering together, a very high resolution output can be obtained without being constrained by the fill rate of a single graphics card. This configuration is intended to work with visualization setup where each slave is directly connected to one or more output devices, in order to avoid sending the rendered images back through the network. This approach is particularly effective when dealing with high-resolution output: sending high-resolution images over the network especially when interactive frame rates are required would not be a negligible problem. Many immersive visualization systems exploits multiple output devices, e.g. CAVEs use multiple projectors in order to surround the user. Projectors are physically arranged to compose a unique seamless visualization. Some of the problems which typically raises in this cases, like overlapping zones or non-perpendicular projections, are handled by our software.

4.2.3 *Network Protocol*

Network traffic generated by the Network Driver has soft real-time requirements and is not tolerant of data loss. Proper functioning of the XVR Network renderer requires a fast LAN, limited network delay and a reliable transport layer with guarantee of in-order arrival. Furthermore the system performances are tied to the network speed. Though the system can exploit any kind of LAN, it is intended to be used over an isolated network. This is a reasonable requirement since the equipment needed to set up a network from scratch are extremely cheap, and there is no practical reason to allow the clustered computers to communicate with other hosts besides the master. Given this requirement, in-order arrival and lack of duplicates are no longer a concern since they are guaranteed by the LAN data-link protocol itself, including Ethernet. Packet losses are still possible, but on an isolated and correctly working network, this rate is close to zero. By the way, an error-recovery mechanism turns out to be important especially as far as per-frame synchronization is concerned (4.2.4). We ruled out the use of TCP, which would have delivered redundant guarantees while introducing a noticeable overhead in network traffic and communication delay. Our system exploits the use of the UDP as its network transport protocol. The reliability of the communication is demanded to the higher-level Fragmented Datagram Protocol (FDP). In order to lower the network load, all the data are sent to a broadcast or a multicast address.

Two application-layer protocols have been developed to manage the cluster's functioning, each of them dealing with a different aspect of the communication. The Network OpenGL Protocol (NOGLP) is the high-level protocol that handles information exchange, per-frame synchronization and data compression. First of all, a NOGLP packet containing the cluster configuration is sent to the slave nodes allowing for cluster initialization. Each slave application extracts the needed informations and answers to the master node. The master ensures that all the clients are ready before starting to send the graphical stream. Following packets contain the commands stream. Sending single-command packets would flood the network, while collecting too many commands to be sent would introduce a high latency. For this reason it is possible to set a threshold: when the collected commands stored in the buffer exceed this threshold the buffer is sent, and the collection continues. Either way, the buffer is also sent every time a frame ends. This approach is a reasonable trade-off allowing to exploit the network bandwidth while limiting the latency. At the end of each frame, the cluster is synchronized in order to present all the images at the same time (see section 4.2.4).

The Fragmented Datagram Protocol is the low-level protocol that handles the fragmentation of those NOGLP packets exceeding UDP's maximum transmission unit (MTU). Through the introduction of acknowledgement messages, FPD prevents possible data loss due to slave-side buffer overflow or network packet loss. The Driver sends a number of packets defined by the window size, and waits for ACKs from all slaves; in case of packet or ACK losses it re-sends the missing informations.

4.2.4 *Per-Frame Synchronization*

Typically, the various output images from the different slave programs have to be merged into a single consistent image, either on a flat screen or on a more complex surface. In order to avoid inconsistencies in the resulting image, a per-frame synchronization has to be performed. Not only the slave programs need to be synchronized with each other, but the master node needs to be synchronized with the rest of the cluster as well. This is to maintain a global consistency between input devices, which are connected to the master node, and output devices, connected to the slave nodes. Synchronization is performed by the NOGLP protocol in a simple fashion. Each slave broadcasts a UDP datagram to every node of the cluster, including the master node. The message is sent just before the end of the frame: the command to display the rendered image (*SwapBuffers*) is issued only after receiving the synchronization message from each and every other slave programs. This scheme avoids to experience cluster desynchronizations, at least at frame level. Please note that we do not try to perform fine vertical synchronization, as this is out of scope of our software. For this, we either rely on hardware support where available, or ignore it altogether. In our tests we never experienced any visible artefact in this respect when using passive stereo projection. However when using active stereo projection the use of hardware synchronization mechanism is strongly suggested in order to synchronize the rendering on a per-eye basis.

4.2.5 *Cluster Configuration*

The cluster can be configured through a set of configuration files, specifying the slave nodes which compose the cluster, the *views* belonging to each slave, and the *tiles* which are managed by each view. Each slave program initializes his own OpenGL context according to the size specified by the configuration. For each slave it is possible to define one or more views; each view conceptually represents an OpenGL viewport with its own perspective matrices, with no perspective-continuity requirements. The possibility for a slave to render multiple views allows a single slave to manage a complex visualization system. Although discouraged for performances reason, it would be possible to connect all the displays of a CAVE-like system to a single slave node, assuming it is physically possible. Furthermore, it is possible to split the output of each view into multiple tiles. Tiles allow to split view's output among multiple displays, without the performance hit that would be caused by rendering the application multiple times with different perspective matrices. The flexible configuration allows both to manage a complex visualization system using a single node, or to balance the same rendering load among multiple nodes.

Configuring the system consists in specifying the size of the virtual walls, their locations and orientations according to a reference system and the load distribution by choosing the slaves/views/tiles arrangement. Using this data, the system is able to calculate the transformation matrices needed to compose a unique coherent visualization. It is furthermore possible to configure several additional slave-side feature like stereo-rendering, tracker-driver camera transformations, overlapping compensation and more.

In our system we decided to employ a centralized configuration scheme in order to avoid to access and configure each node individually. Almost all parameters are configurable on the master node; the only parameter that needs to be set on a slave is its unique identifier, which, anyway, is set once and for all. The Network Driver parses the configuration parameters locally and sends them to the slaves through the first NOGLP message. Each slave, then, extracts from this configuration packet only the information pertaining itself.

4.2.6 *Additional Features*

The collected OpenGL commands belonging to each frame can be processed and modified in order to provide additional features, without the needs to modify the original application.

STEREOSCOPY XVR Network Renderer provides software support to several stereoscopic visualization schemes. This feature is particularly important since most commodity graphics cards do not provide hardware support for stereoscopy. Our software-managed stereoscopy supports anaglyph, side-by-side channels and active stereoscopy if hardware supported. It is also possible to obtain the reverse, that is to display only a single channel when the application uses two of them. The master application is not required to be aware of the availability of stereoscopic modes. If the application is originally monoscopic anyway, it is possible to render it in stereoscopic mode nonetheless. This conversion is achieved by buffering the commands composing a frame during the execution of the left eye and after displacing the point of view executing them again for the right eye. Only necessary commands are executed twice, the others are optimized away. This way, we relieve the application developer from taking care of stereoscopic code. The inter-pupillary distance is configurable too.

PERSPECTIVE CORRECTION The ability of each slave to independently modify the perspective matrices of the original OpenGL scene is useful for two purposes. First to take into account the data sent from a head-tracker connected to the system. Secondly, it is possible to adapt to the geometry of a complex visualization system, particularly when slaves are far apart or not coplanar. This way, it is possible to set up a head-tracked CAVE-like system, where each slave handles a single wall, a single channel of a single wall or even a tile of a single wall, as described below. Tracking data can be forwarded by the master node to the cluster. It is furthermore possible to send head-tracking data directly to the cluster in order to minimize the latency. As a mere consequence, the system allows to simply render the application from a different point of view without modifying the original application.

PROJECTION CORRECTION As stated before, the system has been developed also as a platform to build immersive virtual reality systems, exploiting multiple screens. The system therefore provides tools to obtain a smooth continuity between different projected images. It is possible to compensate in software optical flaws or projection overlapping by configuring a 2D distortion mesh for each tile. A visual tool to create the distortion meshes have been developed; it projects test patterns (grids) that can be manually deformed to obtain a seamless image (see figure 4.2).

Future works include the development of an automatic calibration procedure. It is furthermore possible to specify custom attenuation areas and attenuation profiles. Finally the system allows the projection on curved surfaces, which are managed as 3D meshes on which a virtual viewport is projected according to the user's point of view.

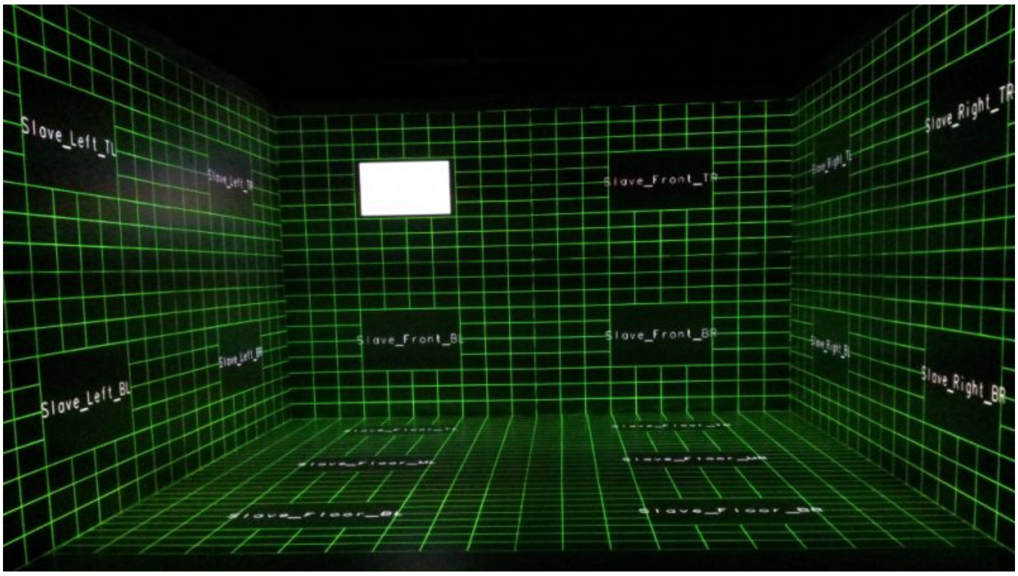


Figure 4.2: Calibrating a 18-projectors CAVE system. Test grids allow for fine distortion compensation.

capturing opengl videos A useful feature of our system is the ability to capture the OpenGL commands streams and execute it later on, as if it would be a video record of a particular run of the master application. The recording is obtained by saving on disk the same data that the master would send to the slaves. Similarly, the playback is implemented by sending the recorded data on the network as if it were generated on the fly. As a result, the recorded file is very compact, and its size is independent of the output resolution. It can also be played with different slave configurations, applying, for instance, additional graphical effects on the output, or modifying the point of view according to data obtained from a head tracker, or running on a different screen geometry.

4.3 parallelism

XVR Network Renderer intends to be a cluster rendering architecture suitable for Immersive Virtual Environments.

The system have been designed paying attention to performances. Both the system's main components—the Network Driver and the slave programs—exploit paradigms of parallel programming. Our architecture adopts a N-stage pipeline model to speed-up performances. Each stage refers to a software module of the system. The number of the pipeline stages is configurable and can vary between 1 and 3 both for the master and for the slave programs.

The three master's modules—the *packetizer*, the *compressor* and the *sender*—correspond to their dual stages on the slave program—the *unpacketizer*, the *decompressor* and the *receiver*—. The packetizer is in charge of intercepting and buffering the graphical commands, the compressor shrinks the size of the data sent over the network, while the sender actually sends them. On the slave side, once the data have been received by the receiver module, they are unshrinked by the decompressor and finally unpacketizer is in charge of decoding the stream and executing the commands in the graphical context.

Multiple compression stages are performed on the stream in order to reduce the network load, and speed-up the performances in case of bandwidth bottlenecks. The overall data compression relies on different compressor modules which applies different shrinking techniques according to the type of the data. If the frame contains geometry description commands, a geometric compression module will handle them (Marino, Gasparello, Vercelli, Tecchia, and Bergamasco, 2010). Similarly, compressible images are reduced in size by a JPEG compressor, whose compression level can be configured to achieve desired ratio.

The beginning of an OpenGL application usually starts with a loading phase, where all the assets are loaded in memory. After that the following frames often consist in the exploration of the model itself. Each frame is mainly a collection of drawing calls and calls to change the camera placement. To exploit frame-to-frame coherence the system uses a diff algorithm (Hunt and MacIlroy, 1976; Marino, Gasparello, Vercelli, Tecchia, and Bergamasco, 2010) in order to avoid sending a large percentage of each frame data (all the geometry information). The above compression stages happens in-place and are performed sequentially by the packetizer, which is single-threaded up to now. If the output of the previous operations does not fit the MTU size, then it is compressed with a general purpose compressor (LZO or Zlib). Indeed, in case of small packets the compression is automatically turned off because the impact of the compression in terms of required time would overcome the benefit of sending a smaller packet. This compression stage is performed by a compressor module in parallel to the packetizer. On the slave node the original data is reconstructed by performing all the stages presented above in reverse order with respect to the master's side. An overall scheme of the parallel architecture is shown in figure 4.3

Even if most of the compression modules internally exploits multi-threading, they necessarily introduce additional latency during the production stage of the frame. Under certain circumstances, the delay is completely overcome by the time saved in sending less data. When the underlying LAN technology allows high transmission rates, compressing data would require more time than transmitting it outright. On the contrary, using slower networks leads to sending times which overcome the compression times. Each compression stage can be enabled or disabled by configuring the Driver in order to obtain the maximum performances from each application, according to its traits.

4.3.1 *Test-bed IVE Setup*

In order to fully exploit the system and to be able to deeply test it, we have intentionally used a complex CAVE^{TM} setup for conducting our performances tests. The CAVE is composed of four projected walls arranged in the shape of a room, with a $4x4m^2$ floor and walls $2.4m$ high. Each wall is back-projected using 4 projectors, while the floor is front-projected by 6 of them, see figure 4.4. In this setup the cluster consists of 5 nodes, running slave programs, in addition to a master node on which the Driver is installed and running the XVR application. The 4 projectors belonging to each wall are connected to a single node, running a single slave managing a single view divided into 4 tiles (see section 4.2.5). The 6 projectors belonging to the floor are instead connected to two different nodes, each one running one slave program configured as a single-view with 3 tiles. Each projector has a 1280x720 resolution.

The system exploits active stereo rendering performed at 60 Hz, and the per-eye synchronization of the cluster relies on hardware capabilities. All nodes are connected to the same isolated gigabit LAN, capable of a maximum transmission unit of 64 kB.

4.3.2 *Test-bed Applications*

In order to conduct our tests we have chosen a set of four applications (see figure 4.5), which differ in terms of complexity, number and size of the assets and average frame size; these characteristics are summarized in Tab. 4.1. The *Rollercoaster* demo is characterized by a mid-low rendering load and by a small average frame size. The *CAD* demo's average frame size is

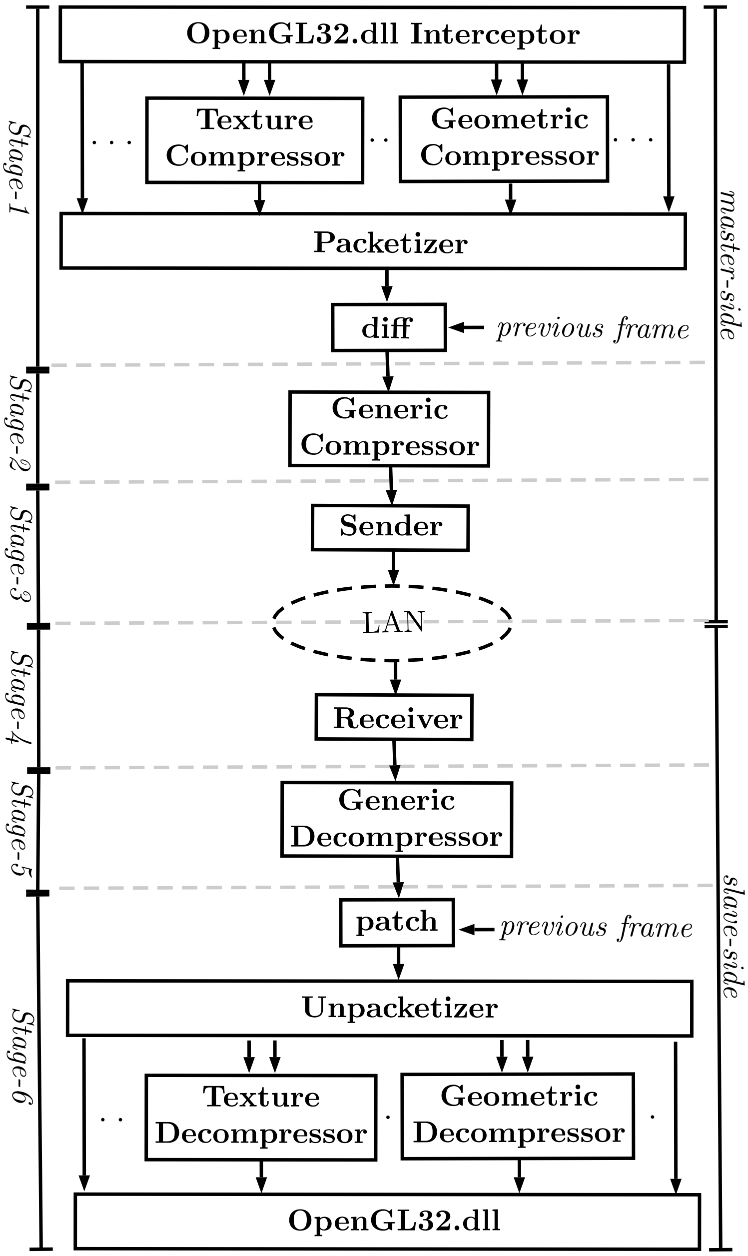


Figure 4.3: Overall scheme of the system's architecture. The system exploit a 6-stages pipepline parallel strategy to speed-up performances.

small as well, but the rendering load is far higher. We have therefore chosen two applications characterized by large frame sizes due to the continuous streaming of video contents: *Kinect* and *Virtual Museum*. The two differ in terms of computational load, the former performs a demanding GLSL shader which generates on-the-fly the geometry by processing the input stream. Screenshots of the application are shown in figure 4.5.

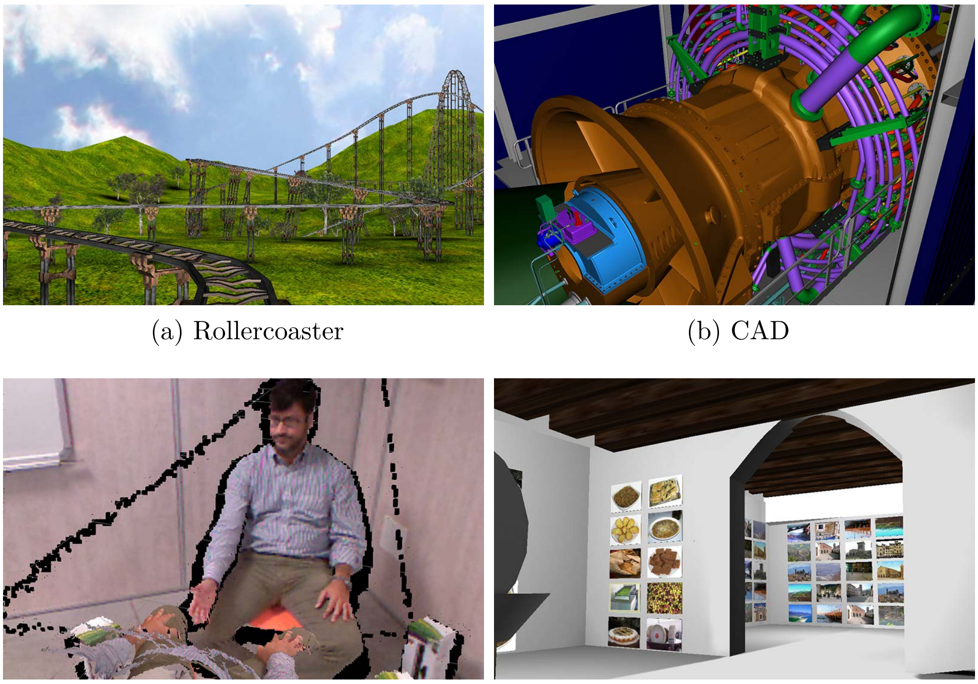
Figure 4.4: The 6-nodes and 18-projectors CAVE setup used to run performances tests.

Table 4.1: Statistics showing the different characteristics of the test-bed applications.

	Roller	CAD	Kinect	Museum
Models complexity (*triangles*)	330k	4.8M	0	234k
Streaming video (*pixels per second*)	0	0	23M	30.7M
Frame size	13 kB	223 kB	1.49 MB	3.67 MB
Compressed size (LZO)	13 kB	57 kB	920 kB	2.67 MB
Compressed size (Jpeg + LZO)	13 kB	57 kB	250 kB	688 kB
OpenGL cmds	2105	38136	324	2145
Display lists	514	18992	4	471
Bound Textures	2	1	10	18
Transformations	10	3	14	41

4.3.3 *Testing Methodology*

The applications used to perform each test have been chosen according to the feature we want to test. Performances have been measured in terms of frames per second that the system have been able to run at. The frame rate have been obtained by averaging one minute of each applications' runs. Graphs reports average, minimum and maximum values. All the measures


(a) Rollercoaster (b) CAD

(c) Kinect (d) Virtual Museum

Figure 4.5: Screenshots of test-bed applications.

have started after that each application finished loading all the assets, in order to avoid that the start-up phase would greatly influence the results.

The cluster configuration used to conduct all tests exploits the above described CAVE setup. The master render the application at VGA resolution in order to minimize his impact on cluster performance. The distortion correction have always been turned on. Even if the vertical sync needs to be turned always on in order to avoid tearing problems, we have decided to disable the V-Sync for the test, in order to be able to register the maximum performances reachable by the system.

All the results referred to the *Local* configuration, have been collected by running the application locally on a slave workstation. The context resolution is the same as if configured as a node of the cluster (WQHD). The local results can be considered as the maximum performances reachable by the system. The master node is more powerful than a slave node and is asked to run the application at lower resolution; in this way it cannot behave as a system's bottleneck.

4.3.4 *Impact of the architecture*

In order to assess the impact of the cluster rendering on performances, we have firstly executed each test-bed application locally on a slave node; secondly we have executed it exploiting the network rendering. Figure 4.6 shows how the cluster rendering negatively influences the applications performances, due to the communication and synchronization overheads. We need to take into account that the cluster rendering has been performed in a stereoscopic way, negatively affecting the performances. Considering all, we have been able to turn a “simple desktop” application into a “fully immersive VR” application without even the needs of modifying it, by degrading and the performances by less than half compared to the local run.

We would also like to assert the impact of the system parallelism exploiting a pipeline paradigm on performances. We have therefore conducted tests on the system configured to run first in sequential mode and then exploiting the pipeline paradigm. Figure 4.7 shows that all test-bed applications have benefited of the higher parallelism. The higher parallelism has led to better results, thanks to the reduction of the waiting time spent for the cluster synchronization.

When the frame size is low, and consequentially the network load is low as well, the pipeline load results to be extremely unbalanced. In this case the impact of the higher parallelism is negligible. On the opposite, when the pipeline load is well balanced—like in the *Kinect* and *Virtual Museum* demos—the performance gain reaches it’s maximum. However, we expect a slightly higher latency when the pipeline parallelism is exploited, but we plan to carry out more tests about this aspect as future work.

4.3.5 *Start-up Phase*

Usually, the first phase of an application’s run consists in loading in memory all the assets used later. We refer to this phase as *start-up phase”*. The start-up phase is obviously slowed down when the cluster rendering is enabled; this is mainly due to the transmission of the assets to the cluster. Figure 4.8 shows how data compression and different LAN technologies affect the start-up loading times, by comparing the initialization times of the applications exploiting or not the cluster rendering. The *Local* loading

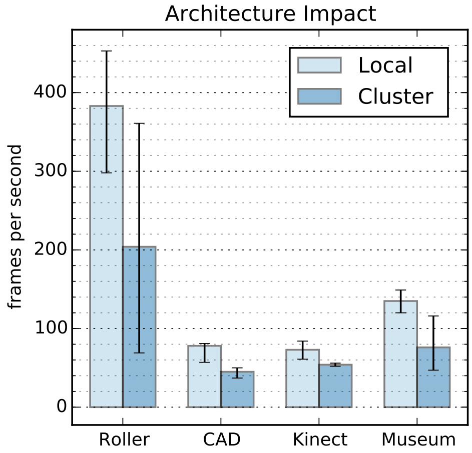


Figure 4.6: Impact of LAN speed and compression on performances.

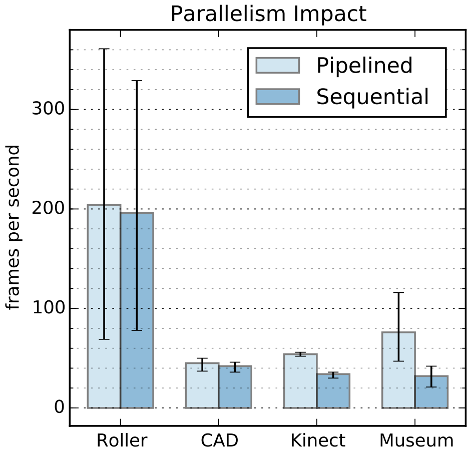


Figure 4.7: Impact of parallel pipeline paradigm on performances.

time can be considered a baseline to estimate how much time is spent to distribute the assets among the cluster.

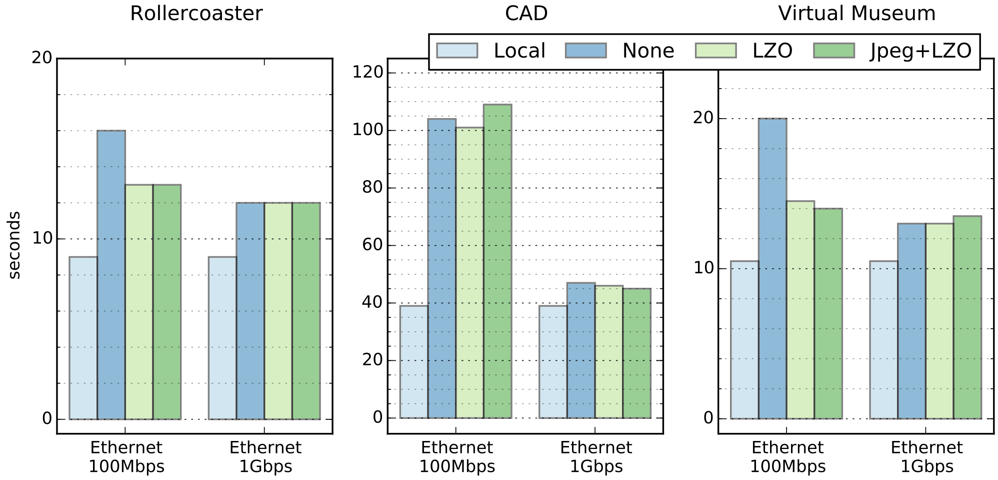


Figure 4.8: Applications' loading times exploiting or not the cluster rendering, and varying the LAN technologies and compression methods.

When the underlying LAN technology allows high transmission rates the loading time is not heavily influenced, furthermore compressing data have resulted to be irrelevant or even counter-productive. Differently, when the LAN speed turns out to be the main system's bottleneck, the more the data are compressible the more the loading times are reduced.

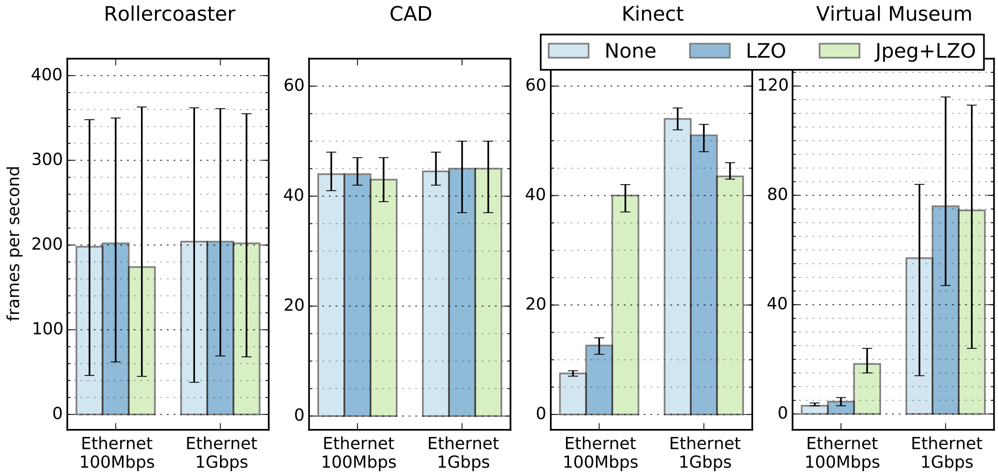


Figure 4.9: Impact of compression and LAN technology on cluster's performances running different demos. No compression–only LZO–Jpeg and LZO enabled.

4.3.6 *Application Complexity and Data Compression*

In order test the impact on performances of the application's complexity and of the different compression methods supported by the system, we have measured the frame rates of 4 different VE applications running on two different LAN technologies.

As reported in figure 4.9, the impact of the compression strictly depends on the frame size and on the available network bandwidth. When the bandwidth required by the application is far lower than the available one, as in the Rollercoaster and CAD cases, the impact of the data compression is negligible. On the contrary, when the application extensively uses the network, as in the Kinect demo and even more in the Virtual Museum's case, the available bandwidth and a greater compression turned out to be extremely important.

Using a network link with a speed of 100 Mbps, both the Kinect and the Virtual Museum demos have highly benefited of greater data compression. When the available bandwidth is far lower than the required one, the higher is the compression ratio the better the system performs. Indeed, we have registered up to a $6x$ speed-up in Virtual Museum which corresponds to a registered $1:6$ frame compression ratio. The higher compression ratio is due to the high compressibility of the streamed contents which have taken advantage of the Jpeg shrinking.

When the network speed is raised up to 1 Gbps, the compression benefits are totally overcame in the Kinect application, while the Virtual Museum demo turned out to be more bandwidth demanding benefiting of the data compression even in this case.

4.3.7 *Scalability*

Many immersive virtual reality system, as stated before, exploit several output devices, like the CAVE we have used to conduct our tests. A first scalability test aimed at comparing the impact of the cluster's configuration on performances. The comparison have been carried out by maintaining the same visual output of a slave node running differently configured. The first configuration exploits multiple slave programs, while the second one exploits a single program configured with multiple views; the latest configuration consists of a single slave program with one view and multiple tiles. The test have been conducted only to assess that the flexible slave-view-tiles configuration (see section 4.2.5) have been designed by paying attention to the performances. As reported in figure 4.10, the best performances have been obtained with a single-slave/single-view/multiple-tiles configuration thanks to the single rendering pass (considering the stereo rendering as unit). Worst performances have instead been obtained by using multiple views, followed by running multiple slaves. In both cases the rendering phase is performed multiple times. While the slave programs are able to take advantage of the multi-core architecture of the used workstations, in the case of the multiple views the rendering passes are performed sequentially. By the way, running multiple slaves means higher memory utilization, that's why we have not been able to run 64 slave programs on a single node.

Looking at the scalability of the system according to the cluster's size (see figure 4.11), significant differences in performance were not found increasing the number of slave nodes from 1 to 5.

4.4 DISCUSSION

The system turned out to be an efficient and flexible solution, which allows to control multiple output devices and workstations even with different specification, and allowing to efficiently distribute the workload among the cluster. The final system performances strictly depend on the characteristics of each specific applications in terms of type of workload. Different levels of

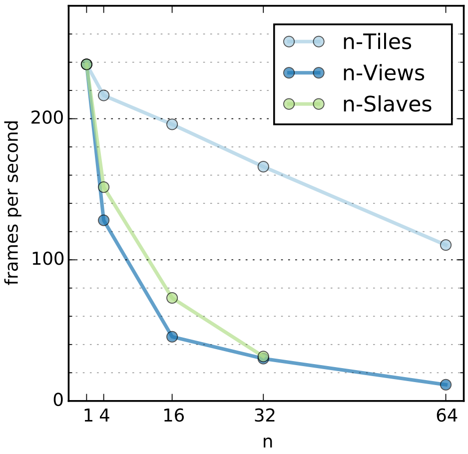


Figure 4.10: Impact of the configuration on system performances running *Rollercoaster* demo.

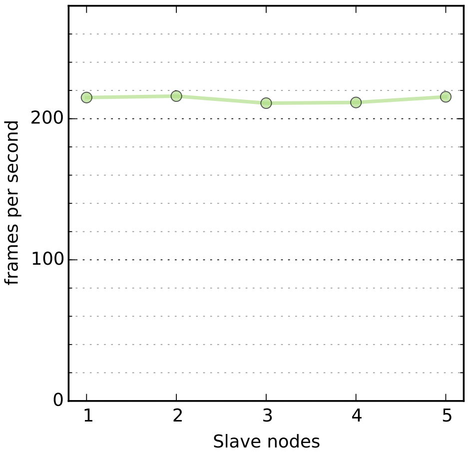


Figure 4.11: Scalability of the system increasing the number of the slave nodes.

compressions and parallelism allow to squeeze of the best from each single application. This software allows a single master workstation to control complex visualization systems such as PowerWalls or CAVE-like systems.

5

HEAD MOUNTED DISPLAY-BASED IVES

Most popular fully immersive VR systems are, nowadays, based on head-mounted displays. CAVE is another popular IVE system, however the most relevant difference between the two are related to the interaction and the self presence. Today, VE are becoming more and more popular also as training simulators. VR applications are always more interactive and VE are designed to be very responsive to user's actions. When a user is immersed in a CAVE, he can see his own body and if properly tracked interact with the VE in a direct natural way. In a HMD-bases IVE instead, the user can see only a representation of the self, but can still perform the same kind of interaction. Seeing directly the self would of course result more realistic, however different problem arises in this case. A user immersed in a CAVE system can see his own body, but the lighting conditions are not optimal: the projectors beams are responsible for the environment illumination and standard lighting conditions cannot be achieved. Furthermore, if the user's body is tracked by means of visible sensors (like wearing a motion-capture suit) the sensors itself are visible to the user; on the contrary when the user can see only a virtual representation of the self, it is possible to control both the lighting conditions as well as hiding the instrumentation attached to the user. But the most annoying problem, is that a natural interaction with own hands would lead to partial occlusion of the projected environment, causing discontinuities in perception and limiting the sense of presence. The situation is even worse if the user wants to grab and manipulate a virtual object: the real body will always occlude the virtual object. On the contrary, using HMD-based systems inter-occlusion between the user and the environment can he properly handled.

MOTIVATION AND CONTRIBUTIONS During our research, we developed two HMD-based IVE systems. A purely virtual reality system and a mixed reality system.

A state-of-the-art VR system was developed in order to serve as a platform to conduct our investigations on virtual training. We aimed at assessing the level of knowledge transfer achievable using VR—of both theoretical and practical notions—when instructing operators in performing maintenance procedures compared to traditional methodologies. The system itself does not represent a novelty, however the contribution consists in giving a clear view of what are the technologies available and which are the key components needed to build a state-of-the-art VR system.

The development of the MR system was driven by our willingness to demonstrate the potentialities offered by a system which allows the user to see his own body and naturally interact with the VE. Our first hypothesis is that the introduction of the photo-realistic capture of user's hands in a coherently rendered virtual scenario induces in the user a strong feeling of embodiment without the need of a virtual avatar as a proxy. Our second hypothesis is that the user's ability to grasp and manipulate virtual objects using his own hands will not only provide an intuitive user interaction experience, it will also improve the user's self-perception as well as the perception of the environment. The architecture of this system represented a novelty in the field when it was presented (of course today similar solutions exists). We used this system to perform two main investigations. First we wanted to assess the effectiveness of using MR systems—allowing a natural interaction and showing a realistic capture of own body—to instruct operators on how to assembly or disassembly industrial machinery. Then we investigated the impact of NUIs and immersive displays on user engagement and social presence in games both from a technological perspective and from a game mechanic design point of view. The studies and the outcomes are presented in dedicated chapters, while in the following we're going to introduce the systems from a technological point of view.

5.1 THE VR FRAMEWORK

With the aim of pushing the boundaries of training simulators we developed a fully immersive Virtual Reality system exploiting the latest devices available and we used the new Unreal Engine 4 framework [1] as software development platform.

The realism we want to achieve is both from a graphical point of view as well from a physical interaction standpoint. The virtual objects must correctly behave according to the physics rules. At this stage of the development no haptic feedback is provided.

---

[1] `https://en.wikipedia.org/wiki/Unreal_Engine#Unreal_Engine_4`; `www.unrealengine.com`

Figure 5.1: A user experiencing the immersive VR training simulator. The user physically grabbing two controllers can see his hand replicated inside the VE.

The system was subject to continuous evolutions, so three configurations were developed.

HMD FRAMEWORK V1 The hardware configuration is based on an Oculus Rift DK2 HMD tracked by means of an Optitrack tracking system. The tracking system exploits 4 high-resolution cameras ($2040x2048pixels$) providing tracking data at the speed of 240 Hz. Three markers are placed on the HMD, mounted trying to minimize partial occlusions of markers from the cameras lines of sight (see figure 5.2). A custom head-tracking module was developed to obtain adequate tracking performances in any condition describes in section 5.1.1.

The hand tracking is based on two Nintendo Wii controllers tracked by means of the Optitrack system and connected via bluetooth to allow the use of the controllers' buttons as input devices (see figure 5.3).

The workstation is equipped with an Intel Core i7-3930K Hexa-Core Processor, 32 GB of ram and a Nvidia Titan X graphic card equipped with 12 GB of GDDR5X memory. A portable version of the system exploits 4 tripods with the Optitrack cameras mounted on top of them, placed at the corners of the tracked space and a workstation as well as the HMD and the joysticks.

Figure 5.2: The hand controller: a Wii controller with an Optitrack's constellation mounted on top.

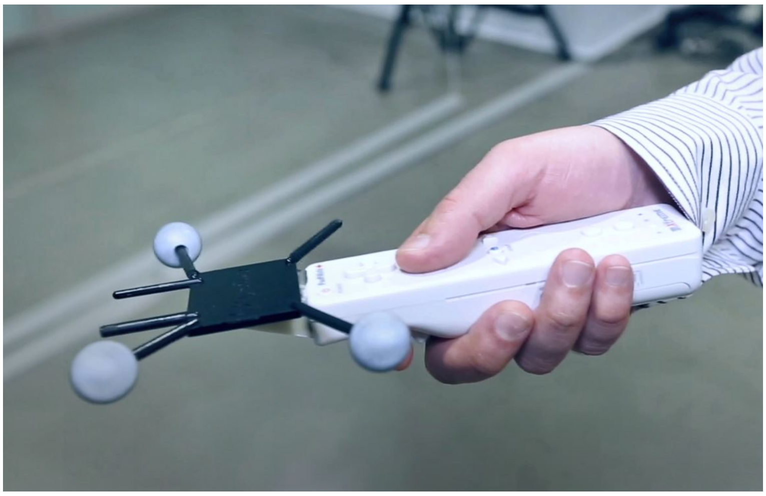


Figure 5.3: The hand controller: a Wii controller with an Optitrack's constellation mounted on top.

HMD FRAMEWORK V2 A more recent version of the system exploit the newer HTC Vive HMD and the provided hand controllers (see figure 5.4). The typical tracked area of the Vive systems is a 4 by 4 meters square. The virtual scenarios was designed to match this size. However, it is possible to track bigger spaces according to own needs by using other tracking systems - like the Optitrack one. Figure 5.4 shows a user seen from the outside while using the system and a display presenting what the user is seeing. Finally a Perception Neuron MOCAP suit is used to provide full motion capture of the user allowing for the perception of the whole own body rather than of the only hands (see figure 5.5). The suit allows finger tracking, however the interaction still does not fully exploit the potential of a physically realistic hand interaction.

Figure 5.4: A user interacting with the VE seen from the outside.

Figure 5.5: A user wearing the MOCAP suit. On the left the user is performing the suit's calibration procedure. On the right the user is interacting with the VE.

HMD FRAMEWORK V3 The more recent fully portable version of the system exploits the use of an HTC Vive HMD as well as a MSI VR ONE[2]

---

[2] `https://vr.msi.com/Backpacks/vrone`

back-pack workstation to allow a fully untethered experience (see figure 5.6) for the user. The MSI VR ONE is a back-pack workstation optimized for the use with VR, equipped with a Intel Core i7-6820HK cpu, 8 GB of RAM and a Nvidia GeForce GTX1070 with 8 GB of GDDR5X memory. The backpack is powered by two hot-swappable batteries able to deliver up to 1.5 hours of continuous untethered experience.

Figure 5.6: The untethered VR system exploiting a back-pack workstation.

5.1.1 *Head Tracking module*

A low-latency robust tracking is among the most important components that allows a pleasing VR experience rather than causing a very annoying effect of motion sickness. For this reason a module to perform an effective 6-DOF head tracking exploiting a custom algorithm was developed. It performs a fusion between the positional data coming from the optical tracker and the rotational data provided by the IMU embedded in the HMD. The optical tracking is achieved by placing 3 retro-reflective markers on the HMD (see figure 5.2). Due to the unpredictable behaviour of the user and the highly interactivity of the user's hands, one or more markers can easily be occluded. The more common solution is to place several redundant markers to compensate partial occlusions of the optical constellation. Our different approach is similar to He, Şen, Kim, Sadda, and Kazanzides (2014) and Enayati, De Momi, and Ferrigno (2015) even if developed before. The final position is computed by the fusion of the markers' positions obtained form the optical system and the markers' positions estimated using the inertial data. The rotational tracking relies on the IMU data. In this way the positional tracking is guaranteed even if only one marker is visible. The result is a 6-DOF tracking robust to partial optical occlusions. The positional estimation is updated at 100 Hz while the rotational one at 1 kHz. Lastly, the module allows the automatic alignment between the two reference systems of the optical and the inertial tracking systems. During the development phase a CAVE system was used to visually assess the functioning and the performances of the head tracking module. The impact point on the CAVE walls of a laser pointer rigidly mounted on the HMD is shown in Figure 5.7 compared to a virtual laser pointer.

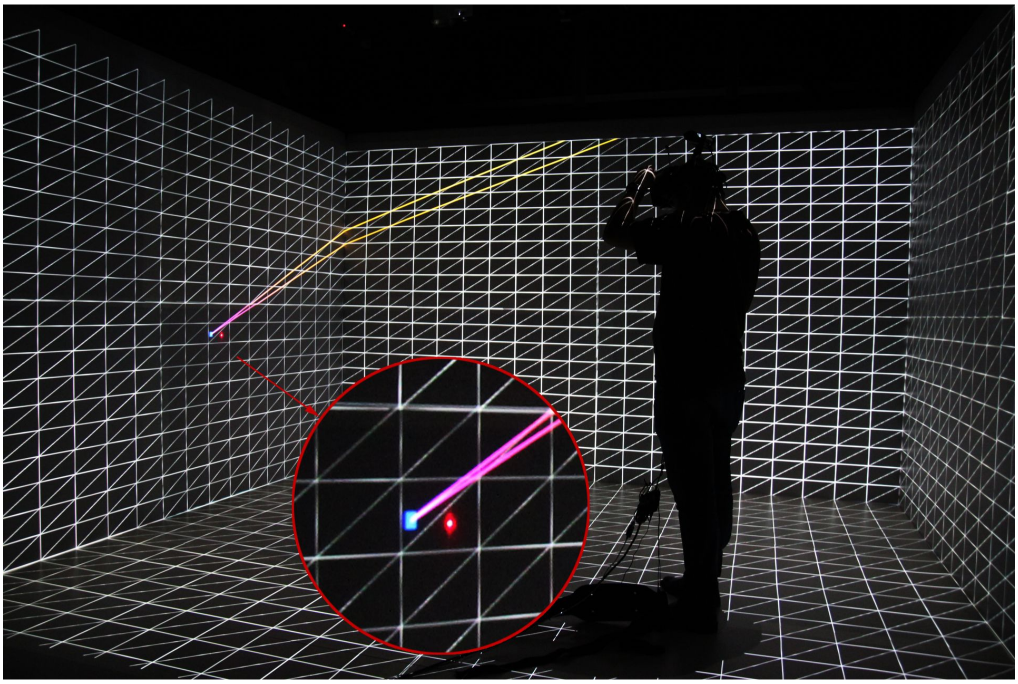


Figure 5.7: A CAVE is used to visually debug and assess the performances of the head tracking module.

5.1.2 *Implementation*

Based on the cited hardware configuration, a training simulator (and a framework) was developed. The simulator was developed using the Unreal

Engine 4 framework integrated with custom modules that handle different tracking systems and input/output devices.

A first plug-in for the engine to allow an effective integration of external tracking systems (like the Optitrack) with the system was developed. A second plugin to allow the use of the bluetooth Nintendo Wii controllers was developed as well.

A framework to ease the development of industrial scenarios was developed as well. The framework consists of a collection of reusable components with pre-defined behaviours allowing for an easier and higher-level development of new scenarios. These components reproduce many standard actual objects (like valves, physics handles, tools, personal protection equipments etc.). Furthermore the framework provides a set of base class and interfaces to easily implement logic and behaviour of generic components (like components that can be grabbed and manipulated, static components, objects that can be damaged and replaced, components that can be attached to the user or to other objects and so on). Mechanisms to define rules and logic of the tasks that users are required to perform was defined. The infringement of the defined rules could lead to predefined outcomes and feedbacks to the user (like explosions, electric short-cuts, alarms etc.). A mechanism to allow users to move across different scenarios based on a elevator metaphor was implemented.

The framework allows to multiple participants to share the VE, simultaneously interact with the environment and performing a common task. The multiple systems need to be connected using a low latency networks to allow a well synchronized shared experience. Users can see each others representation consisting of a head and the two hands, or of a full body skeleton if users are wearing the MOCAP suits (see figure 5.8).

The system does not represent a novelty in this field, however we wanted to build a state-of-the-art VR system that can be used to conduct our research on training simulators.

5.2 THE MR FRAMEWORK

Our interest in the use of fully immersive training systems is the main motivation that brought us to develop the Mixed Reality system presented in this section. One of the most important differences of living the training experience in a totally virtual context (as in VR), or keeping the vision on the

Figure 5.8: Two users sharing the same virtual space. The user on the left is wearing a MOCAP suit. The monitor shows what the user on the right is seeing.

real context (as in MR), is related to the body self-visual perception. This has of course an impact in training, especially in tasks where manipulation operations, or other types of direct interaction with the body, takes place. Avatar representations can be used as a proxy for the user interaction, but at the current state of technology avatars accuracy is still limited, so that they rarely correspond exactly to the dimensions or the current posture of the user and might, although slightly, mislead the self-perception limiting the effectiveness of the virtual training.

Our fully immersive MR system exploits the 3D capture of user's hands, that are reconstructed in real-time and graphically embedded in a synthetic Virtual Environment. The system, embedding some real elements (hands) into a predominantly virtual environment, falls in the Augmented Virtuality branch of the classification proposed by Milgram and Kishino (1994). The user wears an HMD and is free to walk around the scene using his own hands to interact with virtual objects in the scene. It is important to notice that our approach is fundamentally different from previous work on 2D or 2.5D video-based egocentric avatars like the ones described in Bruder, Steinicke, Rothaus, and Hinrichs (2009) or in Fiore and Interrante (2012): the use of a single RGBD camera mounted on the user head is more similar to the work of Suma, Krum, and Bolas (2011) (but they use the camera to see other people, not the self) and allows to have a proper real-time 3D reconstruction of what the user sees; from the data we can compute—on the fly—a geometrically triangulated mesh of the user hands. This has profound consequences on the user visual perception of the self inside the environment: objects-hands inter-occlusion are properly handled, dynamic and geometrically correct virtual shadows can be casted over the virtual objects and any stereoscopic rendering discrepancy between the virtual environment and the captured 3D is implicitly avoided.

While our long-term goal is the realization of a fully untethered Virtual Reality system where the user is immersed in VR by means of a wearable computer and no cables, in this first version of the system some desktop-grade equipment was still in use. As illustrated in Figure 5.9 our prototype is composed of the following items: an optical tracking system, which is used for positional head tracking, an Oculus Rift DK1 HMD connected to a workstation for visual feedback and a 3D camera mounted on top of the HMD support, and integral to it, which is used both for the real-time 3D capturing of the user hands correctly co-located in the virtual environment (and, of course, all the other parts of the body framed by the camera) and for the tracking of the two fingers. Finger tracking takes place by means of two coloured thimbles (blue and green) placed on the index and the middle finger of the dominant hand.

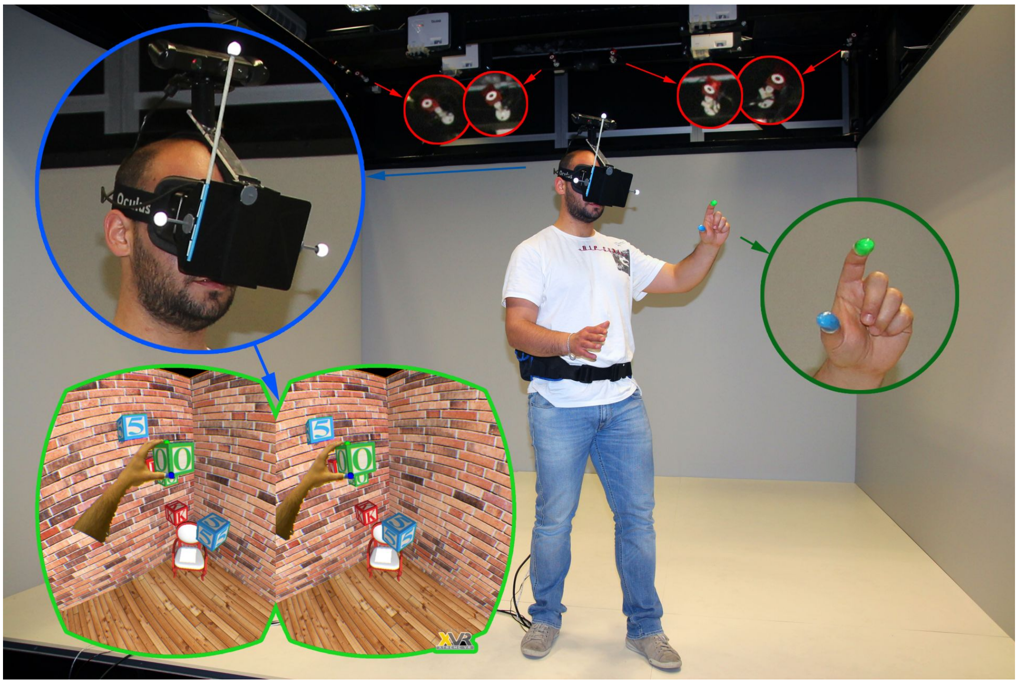


Figure 5.9: The MR system physical layout: OptiTrack cameras in red, finger thimbles in dark green, HMD with markers in blue and an example of what user sees in light green.

A simple collision detection algorithm is applied to this data in order to enable grabbing and dragging interactive virtual objects. The implemented interaction is therefore almost completely natural, with the only added metaphor simulating a simplified grabbing. In fact, no actual physically-based contact is retrieved: when the two fingers touch each other within or in close proximity of an interactive object it can be grasped and moved. A visual feedback for the finger tracking is provided: two small spheres are shown on the finger tips; when the user grasps an object the two spheres become one single red sphere. The user can interact with just a single object at a time.

5.2.1 *Hardware configuration*

The first version of the system adopted the wide field of view (approx. 110° diagonal) Oculus DK1 HMD provided with a $1280x800$@60 Hz display, while subsequently the updated version (DK2) was adopted due to it's higher refresh rate of 75 Hz and resolution of $1920x1080$. The setup used for the optical tracking consists of 8 OptiTrack Flex:V100 cameras, each one equipped with a CMOS sensor capable of providing VGA images at 100 Hz. This system has been used to track the user head position, while the orientation of the user head is estimated by means of the IMU built-in in the Oculus DK1; the inertial sensors used in the HMD is reported as a custom 9-axis tracker (gyroscope, accelerometer and magnetometer) with a 1000 Hz update rate and $2ms$ latency based on the Adjacent Reality Tracker. It can provide 3DOF rotational tracking with yaw drift correction. The RGBD camera is a Primesense Carmine 1.09 (the short range flavour). The camera has a FOV of $57.5°x45°$ and returns a depth map of (approx.) $640x480$ depth samples at the rate of 30 Hz as well as a RGB map of the same resolution and frame rate. Being the short-range version, it can see objects as close as 35 centimetres. The software modules of our system run currently on two workstations: the first one is allocated to the optical tracking software, and it is equipped with a Core i7 3770 CPU (4 core with HT @$3.4Ghz$), 24 GB of Ram and a FirePro V7900 GPU. The second workstation is used for the real-time rendering, the Primesense data handling and the general management of the VR application, and it is equipped with a Core i7 960 (4 core with HT @$3.2Ghz$), 24 GB of Ram and a Nvidia 680GTX GPU with 1.5 GB of memory. Both systems are running Windows 7 64bit. The graphical workstation is connected to the HMD by means of a $10m$ video cable, a 10m USB active extension cable and a power supply cable for the HMD.

5.2.2 *Implementation*

To handle most of the basic VR requirements (loading the 3D model of the environment, performing stereoscopic rendering, gather sensors data) we use the flexible and efficient XVR framework (Tecchia, Carrozzino, Bacinelli, Rossi, Vercelli, Marino, Gasparello, and Bergamasco, 2010), that allows us to have a fine-grained control on the basic aspects of visualisation and interaction. For the more specific task of real-time reconstruction and visualisation of the data captured by the Primesense camera we have then developed a custom rendering plugin based on the hardware-accelerated approach described in Tecchia, Alem, and Huang, 2012. Two external modules—developed in C++—use the RGB data coming form the depth camera and 6DOF head-tracking combining the positional data coming from the optical tracker with the rotational data coming from the Oculus IMU (see section 5.1.1).

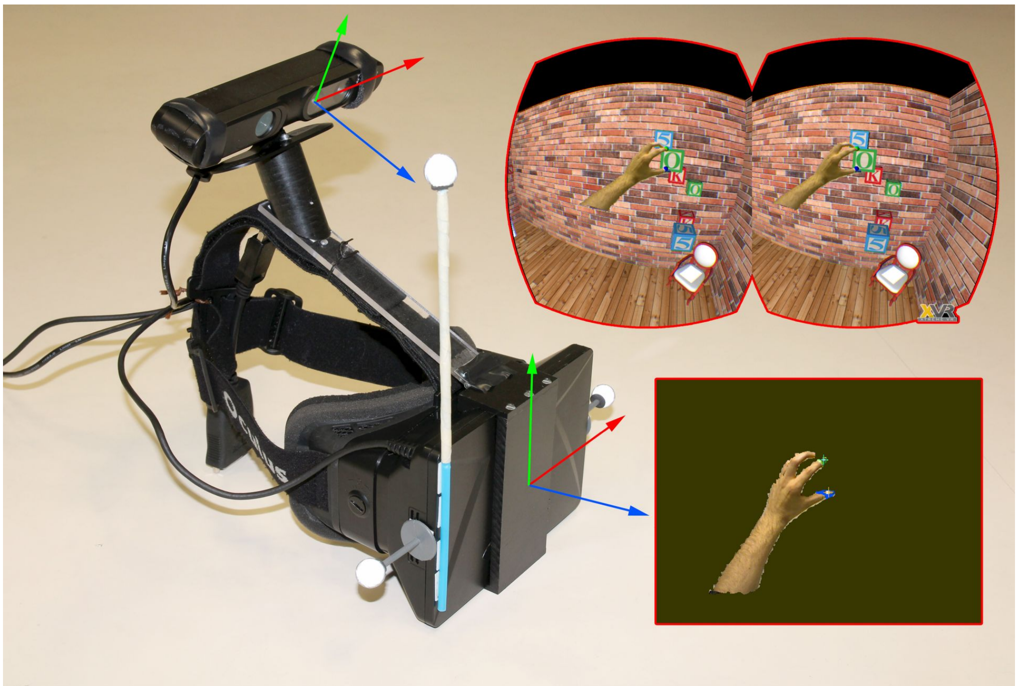


Figure 5.10: Detail view of the HMD system used. In the picture the coordinate reference frames of the Oculus Rift and of the depth camera are shown.

Figure 5.10 shows how we mounted the 3D sensor on the same support of the HMD to capture in real-time what the user is doing with his hands. Notice that the camera is mounted quite a bit behind the HMD, and it's basically placed at the very top of the user head. This placement offers three fundamental advantages: it allows for a better capturing of the user hands movements, as objects placed too close to the Primesense would not be picked from the camera (so mounting it directly on the HMD would basically results in camera blindness when the hands are close to the face), it alleviates the fact that the camera FOV is limited (so placing it at the back allows for a slightly larger workspace to be captured), and finally it reduces the user perceived camera physical mass as the resulting weight distribution minimizes the angular momentum during the common head rotations. In order to render what the camera captures we use the OpenNI SDK to access both the RGB and Depth image array. Both streams are acquired at 30 Hz at the resolution of $640x480$ pixels. Real-time meshing of the data is performed on the GPU using the depth information, and the resulting 3D mesh is co-located in the 3D space where the user is. This is actually easy to achieve, as the 3D sensor is rigidly connected to the HMD, so once the 6-DOF tracking of the HMD is performed we can compute the frame of reference of the 3D sensor and therefore the exact location and orientation for the generated 3D mesh of the hands. The overall process works surprising well and the user has a quite convincing sensation to be looking at own real hands. Having a 3D mesh also fits well the need to generate a stereoscopic view of the environment. Rendering performances on our development machine is definitely sufficient for real-time visualisation: with v-sync disabled our test system is able to exceed $180fps$, including hands and virtual environment stereoscopic rendering and image warping used to compensate for the Oculus lens deformation. To simplify the task of finger tracking, the user of our system wears two coloured thimbles, a blue one on the thumb and a green one on the index. Fingers tracking is then performed using RGBD data obtained by the Primesense sensor. Simple colour filtering is used in order to identify which pixels of the RGB image matches the thimbles colours. The algorithm also uses the depth map data in order to efficiently pre-cull away those pixels that are too far to be part of the user hands. A “grasp” is detected by the application every time the distance between the two finger positions is less than few centimetres. Grasped objects are subjected to fingers translation, so they move together with the user's hands until they are released. In order to avoid continuous releasing/grasping due to imprecisions in fingers detection, a temporal/spatial filter is applied to the tracked positions of thimbles. When the distance between the thimbles exceeds a given threshold, a “release” is detected. The grasping is maintained even when the fingers are not visible by the depth camera, in order to prevent premature releasing of grasped objects. This choice has been made because users naturally tend not to look at hands while moving.

5.2.3 *Usability Pilot study*

Figure 5.11: Showing the overlay of the virtual scenario on the physical space where the user is moving.

For an initial system validation we opted for a simple combination of navigation and manipulation tasks: the user has to walk around in a structured virtual environment avoiding a variety of obstacles as walls and furniture, use the hands to grab some floating cubes randomly distributed in space and then navigate back in one of the rooms of the virtual environment to compose a “virtual puzzle” by placing the cubes on the (virtual) surface of a table. Figure 5.11 helps describing the test scenario: six objects acting as landmarks have been placed in a virtual room: a table, a refrigerator, a sofa, a painting, a chair and a TV Set (see figure 5.12). Eighteen floating coloured toy boxes have been spread across the room. Figure 5.12 shows the virtual room and the position of the six landmarks. The toy blocks have been spread across the room in order to force navigating the whole environment. and have been arranged in three groups close to the landmarks in order to make them more noticeable. The room and furniture arrangement has been designed to enforce obstacle avoidance in order to test navigation ability and spatial awareness. The task consisted in recreating the same cubes layout depicted in the painting placed on a wall (landmark 5).

Figure 5.12: Top view of the virtual room. Landmarks have been numbered.

Method

A total of 14 volunteers aged between 24 and 57 years ($avg = 32.71, SD = 9.12$) recruited among colleagues and students attending the laboratory took part to the study, half males and half females. Before the experiment they have filled the informed consent to participate in the experiment and an entry questionnaire to collect demographic and background informations. Previous experiences with 3D gameplay ($avg = 1.86, SD = 1.1$) and HMD ($avg = 2.28, SD = 1.32$) have been assessed on a 5-point Likert Scale from

1 to 5. Before starting the experiment each subject has been informed about the experiment procedure and has filled the entry questionnaire and the informed consent. Each participant, before to start to perform the task, was left free to familiarize with the MR system for how much time they like. The users placed in a virtual room was able to freely interact with four cubes to get used to the implemented natural interaction. The average duration of this stage was 158.3 ($SD = 52.7$) seconds. The participants were then physically positioned in front of the virtual table and then the application has started and the task begun. No time limit has been set during the task. The experiment session has ended as soon as the subject recreate the same sequence of cubes depicted in the painting.

Upon completion of the assembly task, participants have been asked to fill out a questionnaire aimed at collecting subjective measures about awareness, embodiment and ease of interaction. Measurements was assessed on 5-point Likert (1 to 5). Furthermore they was asked to produce a sketch map of the VE on which they have had to locate the landmarks. A quantitative assessment of the mental representation of the virtual space based on the number of remembered landmarks using a $0 - 6$ score was performed similarly to Huang and Alem (2013). Task's completion time and user's movements was been recorded. Finally an informal debriefing session with the experimeters was conducted to further collect impressions and anecdotes.

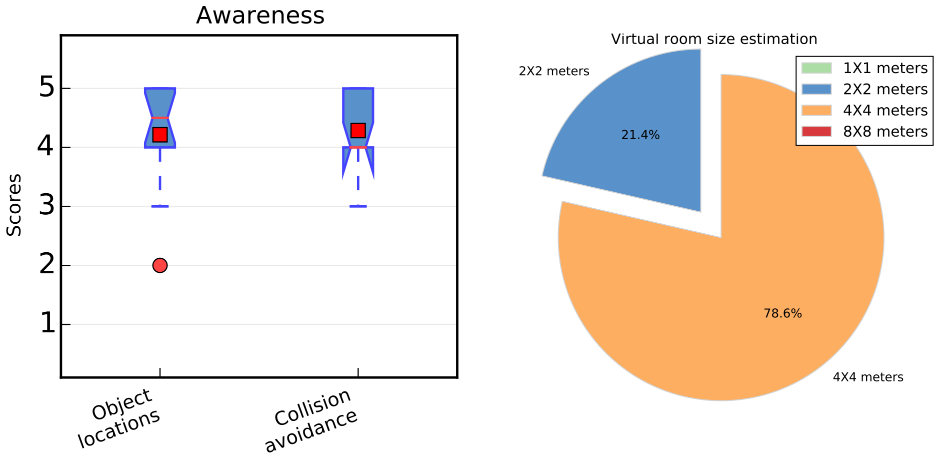
Results

The average time needed by the participants to accomplish the assembly session has been 455.35 ($SD = 113.04$) seconds. The results was analysed to assess the liking of the MR system when performing a simple interaction task. The Mean and Standard Error of the Mean (SEM) are reported for the questionnaire's answers.

AWARENESS Figure 5.13b reports the level of spatial awareness (4.21 ± 0.26) and self awareness (4.28 ± 0.19) reached during the experiment.

Figure 5.13a reports the position of the fourteen participants. Each dot represents the position of a subject in the virtual space. The chart is overlaid on the top view of the virtual room. As shown in the picture, all the users have been able to navigate in the space avoiding collisions with virtual objects and walls. Just three of them — participants 5, 8 and 12 —

(a) Points represent user position during the experiment



(b) Awareness

(c) VE size estimation

Figure 5.13: Results of the pilot study.

have intentionally decided to pass through the walls in order to finish the task more quickly.

In the questionnaire has been asked to the participants to estimate the virtual room size. As reported in figure 5.13c, the 78.57% of the participants have been able to correctly estimate the size of the VE. The subjects have been able to remember almost all the landmarks encountered $(4,71 \pm 0,91)$.

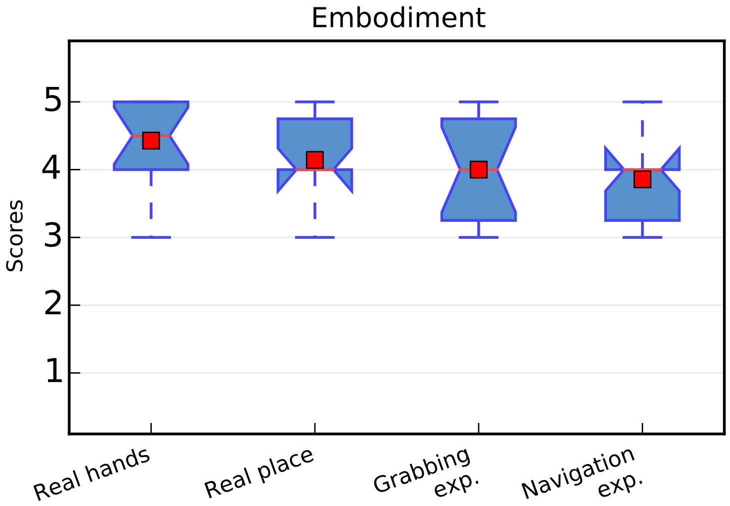


Figure 5.14: Participants embodiment.

EMBODIMENT As shown in figure 5.14, the subjects have had a strong feeling of embodiment. They have perceived the virtual proxy as a real representation of themselves (4.43 ± 0.17) and they have been convinced to be in a real physical place (4.14 ± 0.18). Furthermore the participants have strongly perceived both the interaction with the virtual objects (4.0 ± 0.21) and the navigation in the VE (3.86 ± 0.18) as a real physical tasks.

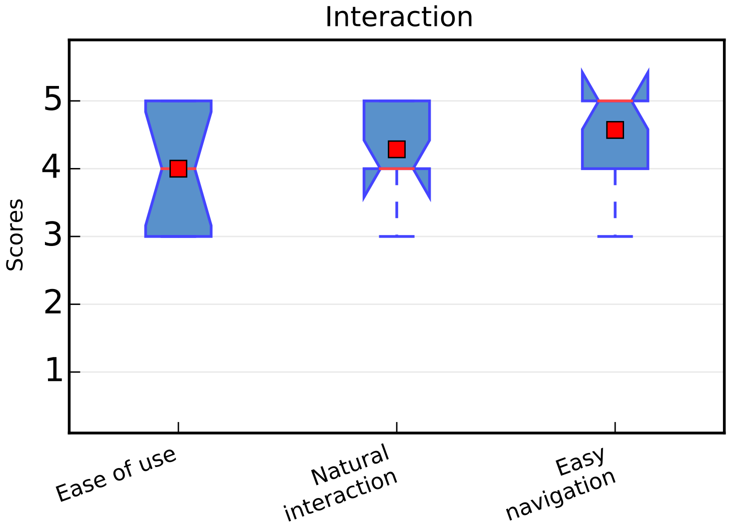


Figure 5.15: Participants interaction.

INTERACTION As reported in Fig 5.15, participants have found the interaction with the virtual objects easy (4.0 ± 0.23) and natural (4.28 ± 0.19). They have also found easy to navigate the VE (4.57 ± 0.17).

5.2.4 *Discussion*

Almost all the players enjoyed the proposed immersive system. Participants have highly rated usability and immediacy of the NUI.

Participants showed a high level of awareness always been conscious of the surrounding environment and demonstrating to know how to reach the target objects. They have purposely avoided obstacles. Experimenters noticed that all the user progressively became more and more confident with the system as the time pass. After a first short period when they were moving carefully because conscious of the real environment they cannot see, they started to move and interact faster. They quickly lost the connection with the real world becoming completely engrossed in the task.

At the end of the game almost all the participants have been able to correctly estimate the virtual room surface and they have clearly remembered how the virtual room was composed and where the landmarks were positioned. A high level of embodiment has been registered during the first experiment thanks to the high immersion achieved by the system. Experimenters have indeed observed—and the results confirmed—that subjects tended to avoid collisions with virtual objects as if they were real as they do in the real life. They have been scared of hitting the virtual objects or walls as they are scared of hitting a real one. However, three participants—mentally aware of the real world—have intentionally crossed the virtual walls to accomplish the task more quickly. Nonetheless the first time they crossed a wall they have been extremely careful afraid of hitting a real one. The Natural User Interface reported a good overall evaluation. The participants have found easy to use the system. Some participants have found limiting the interface as it allows only object translation but not rotation. Almost all the subjects have tried at least once to rotate the cubes the first time they have grabbed one. The finger tracking system resulted to be the the most problematic part of the interface: the finger tracking is performed only when the thimble are visible, so—differently from real life—it is not possible to move a virtual object while not looking at it. Solving this flaws whole greatly improve the NUI usability, but nonetheless the actual implementation has been highly rated.

The result of the pilot study, although preliminary, seems to suggest that this form of self-representation in the virtual environment has great potential to constitute a valid alternative to a more traditional avatar-based user representation. Also, manipulating virtual objects with our own physical hands and navigating in the virtual space using our own physical body has the potential not only to improve the quality of our interaction with the objects in the environment, but also to improve our spatial understanding and self perception in the virtual environments. We believe these benefits to provide a solid basis for supporting further exploration of learning and training processes.

6

REAL-WORLD OCCLUSION FOR OPTICAL SEE-THROUGH HMDS

To obtain an augmented view of the real environment, users wear see-through HMDs to see 3-D computer-generated objects superimposed on their real-world view. This see-through capability can be accomplished using either an optical HMD or a video see-through HMD. In optical see-through HMDs virtual images are combined with the real-world view by means of half-transparent mirrors or by light-additive transparent displays. In video see-through HMDs, instead, the real-world view is captured with two tiny video cameras mounted on the head gear and the synthetic images are digitally combined with the video representation of the real world (Fuchs, 1990). OST systems offer an essentially unhindered view of the real environment, also providing an instantaneous real-world view that assures a perfect synchronization of visual and proprioception information. VST systems forfeit the unhindered view in return for improved ability to see real and synthetic imagery simultaneously (Rolland and Fuchs, 2000).

Most notably, in order to provide a truly convincing experience many challenges need to be effectively addressed. A constant research activity can be seen around topics such as: display technologies (Zhou, Duh, and Billinghurst, 2008), accurate tracking of the user's movements (You and Neumann, 2001; Zhou, Duh, and Billinghurst, 2008), registration of the virtual contents with the real world (Bajura and Neumann, 1995; You and Neumann, 2001) and minimization of the latencies between the user's motions and the presentation of the augmented contents to the user's eyes. Recently, key advances in see-through AR displays have been accomplished in terms of wide field of view (Cheng, Wang, Hua, and Sasian, 2011; Rolland, 2000), real-world masking capability (Gao, Lin, and Hua, 2012; Kiyokawa, Kurata, and Ohno, 2001; Maimone and Fuchs, 2013; Santos, Gierlinger, Machui, and Stork, 2008) and focal depth cues (Hu and Hua, 2012); nonetheless, it is still extremely difficult to embed all this features in a single display. At present, available optical see-through AR headsets—including Microsoft's HoloLens [1]—make use of light-additive displays which are not able to selectively occlude the real environment. This lack have a

---

[1] `https://www.microsoft.com/microsoft-hololens/en-us`

significant shortcoming: the augmented contents are affected by the real-world lighting. For instance, blacks appear to be totally transparent and synthetic objects appear as translucent poorly contrasted "ghosts". The design of AR applications is therefore heavily influenced by this weakness, designers are forced to use of bright objects in their VEs or adopt visual tricks. An example of an adopted trick is visible in figure 6.1a-b: Unity performs a sort of object outlining to improve objects' visibility (it's not due to the lighting of the virtual scene).

AR would also allow to explore perceptual techniques that have no real-world counterparts. In AR often the user may want to see a virtual object inside a solid real one—for instance a giving a sort of "Superman's X-ray vision" to the user. It could be a doctor which looks inside a patient's body The general problem is how to render an object which is inside a real object, not just a tumour inside a patient, but electrical wiring or plumbing inside walls of a house; a lock mechanism inside a car door. Also in this case, the real object's surface needs to be masked, otherwise depth conflicts could arise. This kind of vision is not natural, and it is not yet understood how the human visual system reacts to information displayed with purposely conflicting depth cues, where the depth conflict itself communicates useful information.

For these reasons we strongly believe that the adoption of some form of real-world occlusion is an essential requisite to realize the full AR potential. It is possible that Microsoft or other companies will address the issue, however we propose a viable method to add such capability to the current generation of AR displays. In order to be able conduct our studies we developed an AR system based on the Microsoft's HoloLens exploiting a fine lighting-control obtained replacing the standard room lighting with projectors. *Occlusion shadows* (Bimber and Frohlich, 2002) of the virtual objects can be "projected" onto the real world (Maimone, Yang, Dierk, State, Dou, and Fuchs, 2013) allowing to fine control the level of blending between the virtual and real worlds. The resulting framework based on Unity[2] is made freely available to anyone interested in conducting their own studies.

We have therefore conducted two experiments in order to evaluate the importance real-world masking in AR. The first experiment aims at evaluating the ability of the user in identifying the shapes of the virtual objects when the projective occlusion mask is turned on or off. This should

---

[2] `https://unity3d.com/`

give us a first indication on the benefits/detriments of our approach. The second experiment, instead, aims at assessing the user's ability to localize virtual objects placed behind real surfaces when such surfaces are masked or not. The design of the experiment takes inspiration from a medical procedure—a needle biopsy—to assess the effectiveness of using an AR system that enables x-ray vision (Rosenthal, State, Lee, Hirota, Ackerman, Keller, Pisano, Jiroutek, Muller, and Fuchs, 2001) when performing specific tasks.

MOTIVATION AND CONTRIBUTIONS Most available optical see-through AR headsets are not able to selectively masking the real environment, as results it is not possible to show solid virtual objects in some lighting conditions. We strongly believe that the adoption of some form of real-world occlusion is an essential requisite to realize the full AR potential. It is possible that Microsoft or other companies will address the issue, however we propose a viable method to add such capability to the current generation of AR displays. In order to be able conduct our studies we developed an AR system based on the Microsoft's HoloLens exploiting a fine lighting-control obtained replacing the standard room lighting with projectors. "Occlusion shadows" (Bimber and Frohlich, 2002) of the virtual objects can be "projected" onto the real world (Maimone, Yang, Dierk, State, Dou, and Fuchs, 2013) allowing to fine control the level of blending between the virtual and real worlds. The resulting framework based on Unity[3] is made freely available to anyone interested in conducting their own studies.

We have therefore conducted two experiments in order to evaluate the importance real-world masking in AR. The first experiment aims at evaluating the ability of the user in identifying the shapes of the virtual objects when the real-world occlusion is turned on or off. This should give us a first indication on the benefits/detriments of our approach. The second, instead, takes inspiration from a medical procedure –a needle biopsy– to assess the effectiveness of using an AR system that enables "Sumeprman's X-ray vision" (Rosenthal, State, Lee, Hirota, Ackerman, Keller, Pisano, Jiroutek, Muller, and Fuchs, 2001) when performing specific tasks. Our study compares user's ability to estimate the position of a virtual object placed behind an opaque real surface when the real-world masking is enabled or not.

---

[3] `https://unity3d.com/`

6.1 FINE LIGHTING-CONTROL

Our approach exploits a fine lighting-control as proposed by other works (Bimber and Frohlich, 2002; Maimone, Yang, Dierk, State, Dou, and Fuchs, 2013), to add real-world occlusion capabilities to the Microsoft HoloLens headset by means of stereoscopic projectors and shutter glasses to allow an enhanced stereoscopic vision. The idea is to mask the real-world surfaces that lay behind the augmented contents in order to improve AR imagery by "projecting blacks" on top of them. Of course, projectors share the original HoloLens display limitation to not being able to "project blacks", they can in fact only add light to the environment but not subtract it. To address this limitation we need to achieve a full control of the lighting by replacing the standard lights with projectors so that they are the only light sources in the environment. In this way the whole environment can be illuminated except for those areas that lay behind the virtual objects showed on the AR display from the user's point of view; those areas are called "occlusion shadows". From the user's point of view, occlusion shadows are projected exactly behind the virtual objects so that they are not directly visible but contribute to providing well contrasted and solid virtual objects. In this way it is possible to mask the real-world surfaces, and more in general it's possible to modulate the lighting in a way that allows to completely hide or turn some real-world surfaces to appear transparent, allowing for instance the achievement of special effects like the "x-ray vision".

In order to compute the occlusion mask the geometries of both the real and the virtual environments must be known. The real environment can be modelled by using a CAD software or acquired by means 3D scanners or RGB-D cameras. If the real environment is static an initial acquisition of its geometry is sufficient, conversely for dynamic scenes the geometry needs to be continuously updated. Theoretically the real-time 3D reconstruction of the surroundings performed by the HoloLens would perfectly satisfy this needing; practically, due to limitations in computational power and precision, the definition of acquired geometry is too rough to allow an accurate computation of the occlusion shadows in our case. We hope this problem will be solved in the next device releases. For this reasons we modelled the local environments using CAD software.

Once the geometries of the real and the virtual worlds are known, the stereoscopic occlusion shadows for each projector can be computed in this way:

1. Render the real environment from the projector's perspective and save the resulting depth buffer.
2. For each of the two eyes, render the virtual environment from the user's eye perspective and save the resulting depth buffers.
3. For each pixel of the projector's depth image, project the corresponding depth value onto each eye's depth buffer.
4. For each correspondence, if the depth value is closer than the projector's one the pixel is rendered as black (virtual object closer to user than real environment), otherwise as white.

Figure 6.1: Actual see-through footage of a test scenario captured by a camera located behind the AR headset. (A) Black virtual object on the right next to a real one without occlusion mask. (B) Occlusion mask enabled. (C) Occlusion mask and virtual shadows casting on the real environment enabled.

Figure 6.1a-b clearly shows the first advantage of using occlusion masks: better image contrast and virtual objects appearing more solid. Using a similar approach is also possible to calculate and cast shadows of virtual objects onto the real environment; in our case the positions of light sources correspond to the projectors positions (see figure 6.1c).

6.2 SYSTEM SETUP

The environment lighting had to be turned off and one or more stereo projectors needs to be arranged so that they can illuminate the entire area of interest. Shutter glasses are placed in front of the HoloLens display (see figure 6.3) in order to separate the occlusion masks calculated for each eye and projected by the stereo projectors.

Positioning the shutter glasses between the eyes and the HoloLens display is not possible due to the HoloLens display technology. The display runs at $240Hz$ (custom colour-interleaved rendering) while the commonly used stereo projectors runs at $120Hz$: placing the shutter glasses in between results in colours interferences.

Stereo projectors are connected to one or more workstations in charge of computing and displaying the occlusion shadows according to the user's point of view, to the projectors arrangement and to the environment's geometry. The real-time head pose estimation is performed by the HoloLens inside-outside tracking module and is streamed wireless to the workstations. The synchronization between the projected occlusion mask and the augmented contents displayed on the HoloLens relies on the synchronization of the tracking data between the HoloLens and the workstation. For this reason low latency streaming of the tracking data is important to avoid de-synchronization between the occlusion mask and the virtual world displayed on the headset. A simple setup exploiting a single projector is shown in figure 6.2.

Adding shutter glasses to a device which is not necessarily slim form plus requiring to replacement lights with stereoscopic projectors could result too much for many. Our aim in this paper is to develop an experimental platform to test the importance of occlusion, not to propose a practical solution. A practical solution remains illusive for wide-angle, compact AR headset. We envision a custom construction of stereo shutters in front of a wide-angle AR headset that would be quite compact and more comfortable for future experiments and user studies.

CALIBRATION In order to accurately compute and present occlusion mask, an accurate calibration between the headset, the projectors and the real-world must be performed. The framework provides a semi-automatic calibration procedure to easily align the reference systems: by placing 2D markers (any image with enough visual features) on the projectors and on the real environment it is possible to use the RGB camera embedded in the HoloLens to estimate the markers poses with respect to the Hololens. The HoloLens is capable of recognizing an environment previously acquired so the persistently stored calibration remains valid until the environment changes even after a device restart. The calibration of the projectors' intrinsics – required to compute the occlusion shadows – can be performed using any

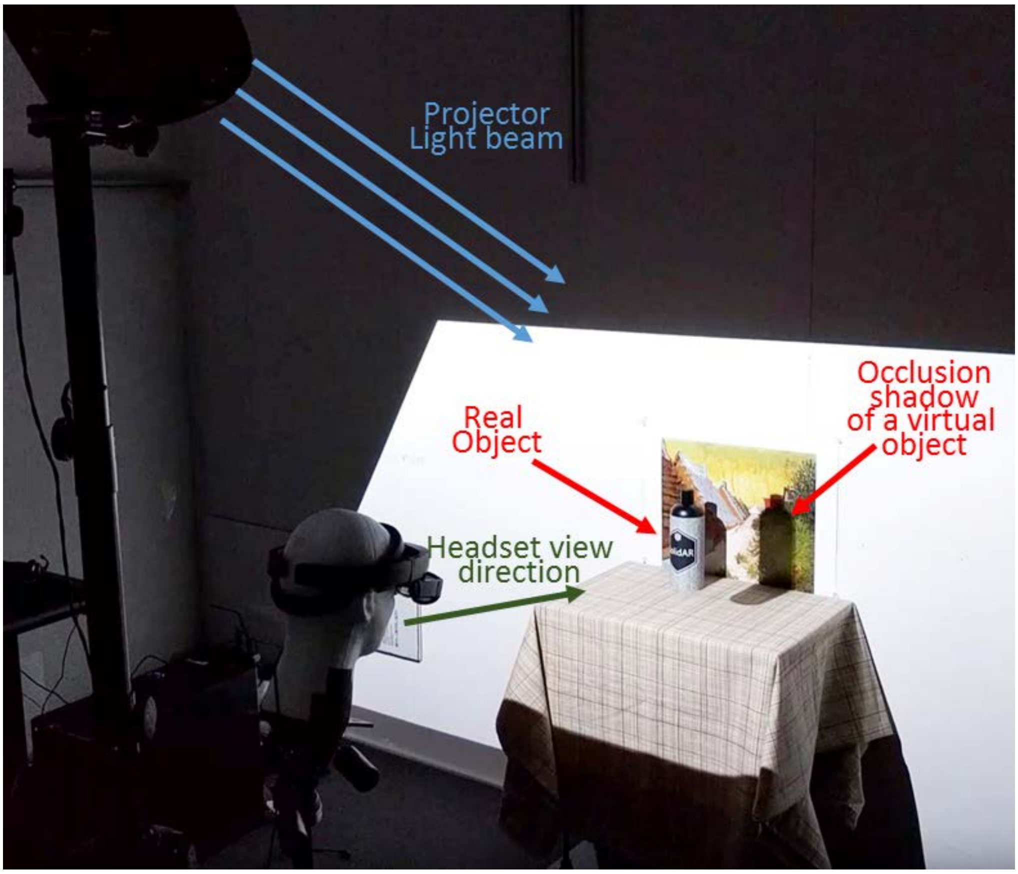


Figure 6.2: Overview of the system. A simple real word scenario is illuminated by a stereo projector. The occlusion shadow of a virtual bottle is visible. The AR headset is placed on a Styrofoam head.

of the several standard computer vision techniques (Kimura, Mochimaru, and Kanade, 2007; Martynov, Kamarainen, and Lensu, 2011).

6.3 EVALUATING THE OCCLUSION IMPORTANCE

The main objective of this paper is to evaluate the importance of integrating a real-world occlusion mechanism in any AR headset. Two experiments have been conducted to assess the benefits of the projected occlusion mask while performing some tasks.

6.3.1 *Experiment 1*

The first experiment aimed at investigating which real/virtual environments conditions could benefit of real-world occlusion.

Figure 6.3: Microsoft HoloLens headset with active shutter glasses. Styrofoam head contains a camera used to capture actual footage of the see-through display.

6.3.1.1 *Experimental Design and Procedure*

The idea behind this experiment is to assess the ability of the user to identify the shape of a virtual object. The experimental setup consists of four real bottles with different shapes placed on a table; a virtual bottle—whose shape matches one of the four real bottles—is shown on the AR display as if placed in the middle of the real bottles and the user is asked to identify the correct match. The real bottles are numbered to ease the answer's collection. The task is very simple, so that it doesn't require any mental effort to the user, but is purely based on his visual perception. The experiment arrangement is shown in figure 6.4.

In order to investigate how real-world occlusion in AR displays could impact on the user's perception of the object's shape different factors have been tested. The first factor consists in enabling or not real-world occlusion in the AR headset. This functionality is simulated using the proposed system. This factor have of course been chosen cause we're interested in investigating the impact of this functionality. We expect that projected occlusion mask will greatly helps in identifying dark objects, while is useless to enhance white bottles' perception. The second factor consists in modifying the

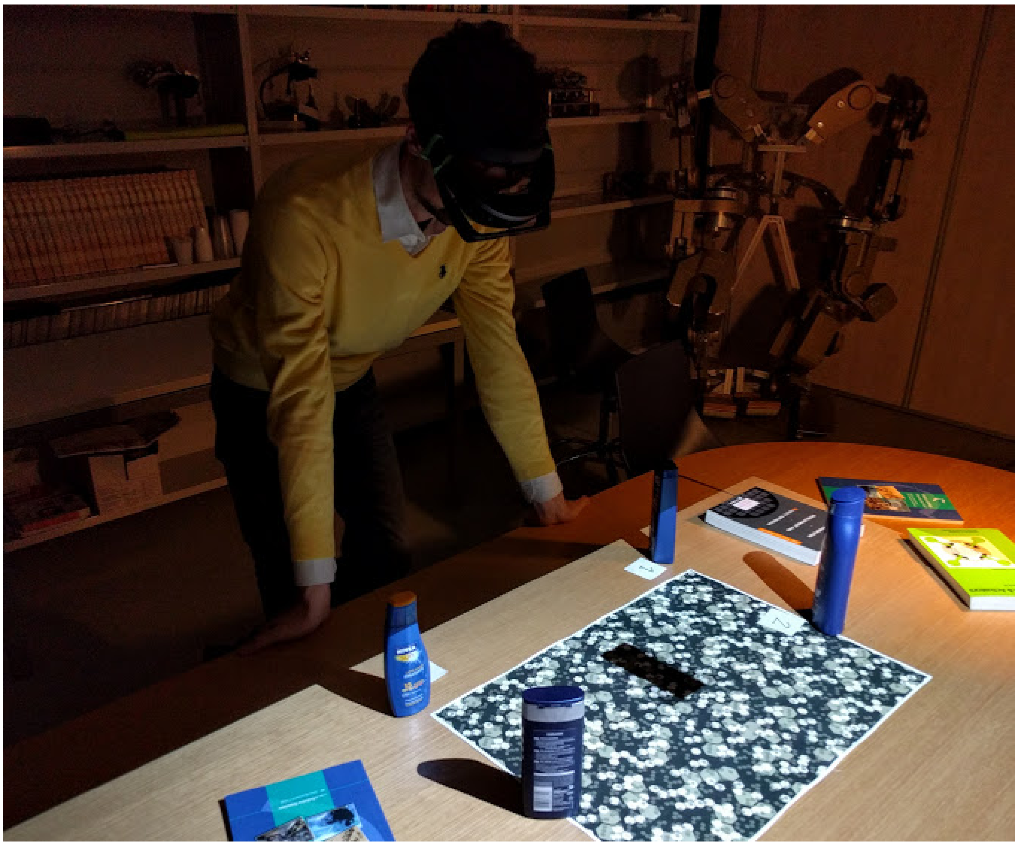


Figure 6.4: The first experiment setup.

virtual bottle's shape. The shape is chosen among the 4 shapes of the real bottles. We don't expect any influence of the shape on the identification ability. The third factor taken into consideration is the material of the virtual bottle. As we already discussed optical see-through displays struggles in showing dark colours—as they appear transparent—, so we decided to dynamically apply different materials to the virtual bottle to evaluate their influence on the shape's identification. The worst case—according to our expectations—consists in applying a totally black material to the bottle, the best case uses a very bright white material, and the last case should fall in the middle using a multicolour texture. The same multicolour texture has been applied to the table surface to make more difficult to distinguish an object with a similar texture placed on it. We expect that the users will experience difficulties in identifying shapes of the dark bottles, while bright ones should be clearly perceived. The last factor consists in modifying the size of the virtual object to appears slightly smaller, bigger or equal to the real counterpart. We choose to introduce this factor because making the object bigger the user should see it better while making it smaller should have the opposite effect. The increment/decrement size has been set to

±5% in volume in order to avoid to make the object appear too big to fit in the HoloLens field of view or too small avoid that the display resolution could impact the shape identification.

The 4 factors tested:

1. occlusion (2 levels): on, off;
2. shape (4 levels): 4 different bottle's shape;
3. material (3 levels): white, black, complex;
4. size (3 levels): smaller, equal, bigger.

Figure 6.5: First experiment results. Comparing object's shape recognition varying texture. $*p \leq 0.05$.

The different shapes, materials and sizes are visible in figure 6.5.

Using the proposed setup the user is able to choose from which point of view look at the object. To estimate if the object's shape can be immediately identified—considering the ease of the task—the virtual bottle is shown only for 3 seconds. No time limit was given to the participant to provide the answer which was manually recorded by the experimenters.

A within subject design has been adopted. Ten participants aged between 24 and 45 ($avg = 32.6, SD = 5.31$) took part to the experiment. Each participant tested all the 72 combinations of the 4 factor, and for each trial the number related to the perceived shape was collected. In order to mitigate potential transfer effects the conditions order was randomized for each participant.

6.3.1.2 *Results*

The test took an average time of $10 : 24$ minutes ($SD = 0 : 35$) to be completed by each users. This means that 8.67 second is the average time spent by the user to give his answer and by the experimenters to collect it.

In order to determine the statistical difference between the investigated methods, a two-way ANOVA for repeated measures test was conducted. In all ANOVA tests, the full model was conducted first. The data sphericity was tested using Mauchly test, and when violated the Greenhouse-Geisser correction was applied. When significant interaction was detected, focused ANOVA was conducted by fixing the levels of one of the interacting factors. When no interaction was detected, reduced ANOVA model with only one main factor was performed. The significance level for all analyses was set to $p \leq 0.05$.

Statistical analysis reported significance of the main factors “occlusion” ($F_{(1,9)} = 9.669$, $p = 0.013$) and “material” ($F_{(2,18)} = 13.065$, $p = 3.126e-04$) but not of the remaining factors “shape” ($F_{(3,27)} = 2.010$, $p = 0.136$) and “size” ($F_{(2,18)} = 1.268$, $p = 0.305$). Also the interaction between the occlusion and the material factors resulted to be significant ($F_{(3,27)} = 12.094$, $p = 4.686e-04$).

Pairwise comparisons—corrected with Bonferroni—of the material factor showed significant differences between the black and the complex textures ($p = 0.014$) and between the black and the white ones ($p = 0.014$). As expected the recognition rate of the black bottles’ shapes resulted to be much lower than when using the other textures. However no significant differences was found between the other two materials.

Comparison between the two occlusion conditions was found significant ($p = 0.013$), and the activation of the projected occlusion mask led to a greater recognition rate.

Post-hoc multiple comparisons tests of the occlusion*material interaction was performed. When the occlusion is turned off, significant differences in the shapes’ recognition rate between the black material and both the others was found (white: $t_{(11)} = 4.15, p = 0.002$; complex: $t_{(11)} = 3.614, p = 0.005$). The user can worse recognize the shapes of black objects when the occlusion mask is turned off. Conversely, when the occlusion is enabled, no significant differences was reported between the 3 levels of the material factor: the user was equally able to distinguish the shapes when the real-world occlusion is enabled. Differently from what we expected, no differences was reported in the comparisons between the two brighter textures. This outcome could be due to a design flaw in choosing a too easy task: recognizing the shape of an object which can be seen—even if not perfectly clear—is an easy task itself (recognition rates in these cases are in fact above 96%). Choosing shapes which require more effort to be distinguished between each other or asking the user to read something on the bottle could have led to different results. A summary of the occlusion*material interaction results is shown in figure 6.6.

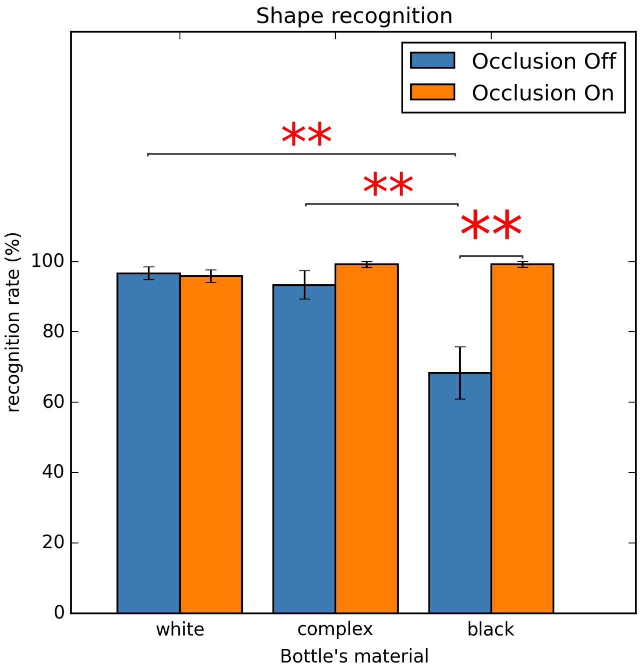


Figure 6.6: First experiment results. Comparing object's shape recognition varying texture. $*p \leq 0.05$.

As expected, projecting an occlusion mask dramatically improved performances in the worst case: the recognition rate of dark virtual objects reached the same level as when showing bright objects. Even if the occlusion mask didn't improve performances in the case of bright objects, it didn't lead to worse results either, suggesting that occlusion mask is useful and not detrimental. This simple experiment is our first evidence that including occlusion mask in optical see-through AR systems would allow their use in wider range of real-world scenarios.

6.4 PURSUIT "X-RAY VISION"

The ability to virtually see what is hidden behind real objects offers tremendous potential to AR users allowing a sort of "Superman's X-ray vision". To our recollection, we first heard this AR capability likened to Superman's X-ray vision by VR pioneer Warren Robinett. Being able to see objects behind solid surfaces doesn't happen in the real world, the closest may be that we see an object inside a semi-transparent object. Different X-ray vision visualization metaphors have been proposed in literature, each one suitable for different purposes and providing different amounts of additional informations to the user (Livingston, Dey, Sandor, and Thomas, 2013). Perhaps the most natural metaphor consists in modulating the opacity of the surfaces, making them appear as transparent and allowing to see what's behind. It is therefore important not to show too many depth layers simultaneously which could result in a misunderstanding of depth order (Livingston, Swan II, Gabbard, Höllerer, Hix, Julier, Baillot, and Brown, 2003) and overloading the user with information. A popular metaphor which could result more plausible consists in cutting virtual holes in real surfaces to show what's behind them, allowing the virtual objects to be perceived at the intended depth without generating confusion (Ellis and Menges, 1998). Also, object's shadows gives important information about spatial relation with the environment, and are therefore useful to accurately estimating its position, so being able to project virtual shadows would be helpful.

All these approaches require to "modify" real-world surfaces—by modulating their opacity, cutting virtual holes or projecting virtual shadows on top of them—so it is important for OST AR systems being able to mask real elements.

Pioneers have used AR to show "live" ultrasound echography data visualized within a pregnant human subject (Bajura, Fuchs, and Ohbuchi, 1992). In attempt to avoid conflicting visual cues, a virtual hole is created inside the real solid abdomen of the patient—with the ecographyc images showed inside it—aided by a occlusion mask for the virtual hole's visible surfaces. Following, State, Livingston, Garrett, Hirota, Whitton, Pisano, and Fuchs (1996) developed an AR video see-through guidance system to help physicians in performing needle biopsy procedures providing localized information directly inside the patient's body. "The system merges rendered live ultrasound data and geometric elements with stereo images of the patient acquired through head-mounted video cameras and presents these merged images to the physician in a head-mounted display. The physician sees a volume visualization of the ultrasound data directly under the ultrasound probe, properly registered within the patient and with the biopsy needle".

One of the major problems they have encountered is the cumbersomeness of the headset. The adoption of more recent devices—like the HoloLens—can address this problem, however, the adoption of an optical see-through display would introduce the previously discussed problems. Grey-scale ultrasound imagery are usually not very contrasted and so hard to distinguished in most lighting conditions. Also, the key parts of the virtual scene are located inside the patients body, hence allowing the x-ray vision without occluding the patient's skin could lead to misleading spatial localization of the virtual content due to conflicts between depth cues.

Ellis and Menges (1998) summarized a series of experiment showing that in the near-field a physical surface can affect both the appearance and the localization of a proximate virtual object. They also found that cutting a virtual hole in the occluder reduced the depth judgement bias compared to superposition.

6.4.1 *Experiment 2*

The second experiment was designed taking inspiration from the work of State, Livingston, Garrett, Hirota, Whitton, Pisano, and Fuchs (1996); however they don't provide solution that we can use, since it was video see through and our system, HoloLens, is optical see-through AR, and we think that OST is in general more useful. For this reason we use our AR system exploiting fine lighting-control to conduct the tests.

In this experiment we aim at investigating the importance of the real-world occlusion by comparing the occluded x-ray vision condition to the non-occluded condition.

6.4.1.1 *Experimental Design and Procedure*

No patients have been involved in the study and the ultrasound imagery of the medical experiment we have been removed. The participant grabbing a stick with one hand is asked to estimate the position of virtual targets located behind a real opaque surface by pin-pointing at them. The task is in analogy to the physician that needs to estimate the lesion's location inside the patient's body and perform the needle biopsy. It would be impossible for the participant to reach a target behind a real surface with the tip of a real stick, for this reason the stick is half real and half virtual. The real stick—optically tracked—is virtually extended so that virtual targets can be pin-pointed with the virtual tip. Distances between the positions of the stick's virtual extension tip and the position of the virtual targets were collected. Exocentric measurements allow to that calibration errors reflect onto the measurements being both the virtual tip and the virtual targets affected by the same error.


(a) (b)

Figure 6.7: The "simple" (6.7a) and the "complex" (6.7b) textures used to simulate different real-world scenarios.

We want to investigate how the participant's precision in pin-pointing targets when the x-ray vision is provided. We also wanted to investigate how different lighting conditions and environment's complexity could affect this task. The environment conditions was simulated by applying different textures to the opaque surface. A "simple" environment is simulated by a texture depicting a patch of skin while a more "complex" one by a texture showing a pile of colourful buttons (see figure 6.7b).

We also wanted to investigate if providing support for the x-ray vision helps to mitigate depth cues conflicts leading to better performances. The adopted solution consists in creating a virtual hole inside the real solid object and display the virtual object inside that virtual hole. The visualization can also be aided by a occlusion mask for the virtual hole's visible surfaces, even though the virtual hole is behind the surface of the real object into which the hole is "cut".

It is more than a hole (cut in a surface), it's rather like scooping out a hole in a solid real object—showing the sides and bottom of the hole—and putting in the virtual objects. The sides and the bottom provide further indication to ease the localization of the virtual objects and how deep are the shafts inside the hole.

Three conditions was explored: showing the virtual target behind the real surface while providing no support (**NoHole-NoOccl**); cutting and showing a virtual hole inside the real surface without the occlusion mask aid (**Hole-NoOccl**) and lastly showing the virtual hole aided by the occlusion mask (**Hole-Occl**). We expect a greater localization accuracy when the environment is simple, while a great number of visual cues as in a more complex environment could generate more confusion in localizing the virtual objects behind a real surface. We suppose that providing additional support for the x-ray vision positively impact on user's performances. Showing a virtual hole should improve localization ability thanks to the additional geometric cues provided. We also suppose that enabling the real surface masking would further increase the accuracy because no more depth conflicts are perceived by the user.

Figure 6.8: Second experiment setup: a user performing the task.

The user, wearing the Hololens and grabbing the stick (virtually extended by 25*centimetres*) is asked to pin-point virtual targets placed behind the surface of a real table (see figure 6.8).

Figure 6.9: A screenshot taken from the user's point of view: the needle, the target and the hole are visible.

The virtual hole, when present, extends from the table surface up to $20cm$ behind it, and is 12 by $12cm$ wide (see figure 6.9). One virtual target at a time is presented to the user and randomly positioned within the virtual hole's volume. Twelve participants aged between 26 and 48 ($avg = 31.67, SD = 5.95$) took part to the experiment. Six conditions were taken into consideration in this experiment: the environment complexity (simulated by texture) and the level of support for x-ray vision (virtual hole/real-world occlusion) are the two independent variables. The distance between the stick's tip and the target object is the dependent variable. A within subject design has been adopted. For each participant 9 repeated measures for each of the 6 conditions have been collected. No time limit was given to the participant to pin-point a target. In order to mitigate potential transfer effects the conditions order was randomized for each participant.

6.4.1.2 *Results*

The average time to collect all the 54 measures was 11.79 ± 4.22 minutes (˜13 seconds per trial).

In order to determine the statistical difference between the investigated methods, a two-way ANOVA for repeated measures test was conducted. In all ANOVA tests, the full model was conducted first. The data sphericity was tested using Mauchly test, and when violated the Greenhouse-Geisser correction was applied.When significant interaction was detected, focused ANOVA was conducted by fixing the levels of one of the interacting factors. When no interaction was detected, reduced ANOVA model with only one main factor was performed. The significance level for all analyses was set to $p \leq 0.05$.

Looking at the impact of the environment complexity, no differences have been found between the two textures ($F_{(1,11)} = 0.406$, $p = 0.537$). Looking at the impact of the second factor, the level of support for x-ray vision, no relevant differences have been found again ($F_{(2,22)} = 1.521$, $p = 0.241$). However, a significant interaction effect between the two factors has been found ($F_{(2,22)} = 3.515$, $p = 0.047$), for this reason focused ANOVA tests have been conducted and some results are shown in figure 6.10.

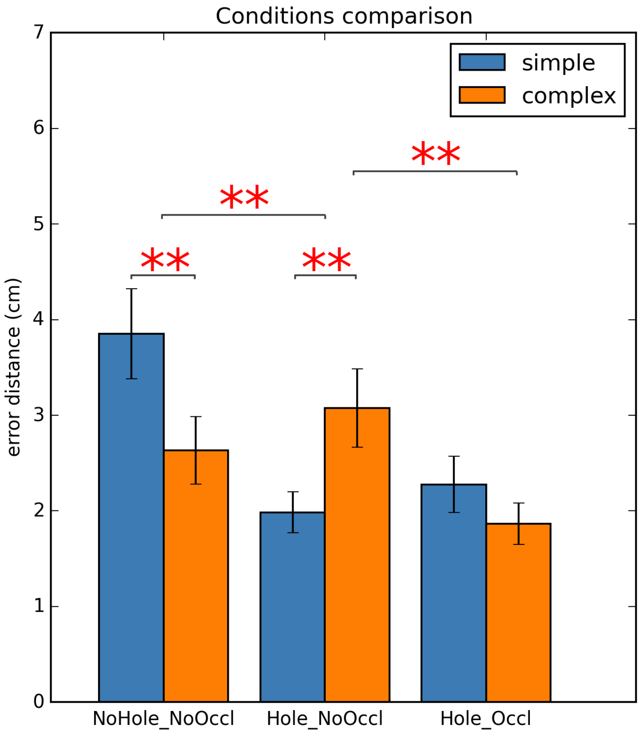


Figure 6.10: Second experiment results: comparison between conditions and textures. $*p \leq 0.05$, $**p \leq 0.01$.

If we aggregate the results collected with both the environmental conditions (simple and complex textures), accuracy improvements are obtained in all conditions where the virtual hole is shown ($p \leq 0.01$). However, adding the real-world occlusion help level doesn't further improve performances.

Looking separately at the two conditions, we found the same results when the environment is kept simple. The additional depth cues provided by the geometry of the virtual hole turned out to be very effective in helping the user to localize the virtual targets behind the real surface. Translating the experiment's results to the needle biopsy scenario, when the surfaces of the real environment are simple, for instance like looking at the clean skin of the patient, providing geometric reference appears to be enough to improve performances. Masking the real world doesn't lead to better performances: the flatness of the environment does not impact on the user's localization ability even if the virtual imagery benefit from better contrast and more colours fidelity.

Differently, when the environment is more complex, post hoc analysis reported less accuracy when only the virtual hole is shown without occluding the real world compared to the condition when no helps are provided ($p \leq 0.05$). The fuzziness of the texture could have led to an improper perception of the additional cues provided by the hole, generating more confusion to the user and reducing the ability of localizing the virtual objects at the intended depth. Conversely, an improved accuracy has been found when the real-world is occluded both compared to the simplest condition ($p \leq 0.05$) and to the condition when the virtual hole is present ($p \leq 0.01$). Occluding real world helps to more clearly perceive the virtual hole and the targets at the intended depth leading to improved localization accuracy. Referring again to the needle biopsy scenario, the projected occlusion mask appears to substantially increase user performance when the operation is performed against a complex textured surface (as it would be the case of a real operation).

Fixing the support factor, results unexpectedly showed that when no support is provided the localization accuracy was higher when the environment was more complex ($p \leq 0.01$). This could be justified by the lower average brightness of the complex texture leading to more contrasted AR imagery and to better results. However, further experiments should be carried out to deeply investigate this outcome. The simple environment scenario benefited more from the virtual hole guidance ($p \leq 0.01$); it was indeed the only one that benefited from the virtual hole support when the real surface is not masked.

Providing different support levels to x-ray vision to help the user in perceiving augmented contents placed behind real surfaces is not effective in all environmental conditions. However, there are some specific conditions that greatly benefit from this support.

6.5 DISCUSSION

The current generation of commercially-available optical see-through displays is not capable of masking real-world surfaces. As a consequence, virtual objects—especially the dark ones—may appear as translucent ghosts floating in front of the display, the imagery contrast is poor, and virtual objects cannot cast their shadows onto the real environment as they would appear totally transparent. Additional metaphors, like the “x-ray vision”, also suffer this limitation.

We presented a system exploiting stereoscopic projectors to add real-world occlusion capabilities to the Microsoft HoloLens, and developed a framework that can freely used by anyone interested in conducting studies in which this capabilities matter. The framework, that we called *SolidAR*, exploits only commodity hardware that can be easily bought by developers and researchers. Even if the system, due to the necessary instrumentation of the environment, is not practical enough to be used in any condition, there are some specific scenarios—like the needle biopsy scenario that we presented—where the benefits obtained by the improved capabilities overcame the high system complexity disadvantage.

One of the aims of this work is to evaluate the advantages of providing such capabilities to optical see-through AR headsets, for this reason we’ve conducted two experiments. The first study aimed at comparing the impact of the projected occlusion mask on the user’s ability to recognize and compare virtual and real objects’ shapes. Results showed that real world occlusion is highly beneficial when showing dark AR contents in a bright environment, while not useful to the task in other conditions, for instance when the objects are bright enough. The second experiment aimed at comparing the precision of the user in localizing virtual objects placed behind a real opaque surface. The comparison have been conducted by varying the complexity of the environment and providing different levels of support to the “x-ray vision”. Results suggest that providing geometric references—by showing a virtual hole cut in the real surface—is sufficient to improve user’s accuracy when the environment is simple, but not sufficient when the environment is more complex. In the latter case occluding the real world is important to allow the user to better resolve the depth cues and improve the localization ability. An unexpected outcome was the better accuracy reported when the scenario is more complex when no support is provided for the x-ray vision. We hypothesize that this effect was caused by the lower average brightness of the environment and thus to the better contrasted and less transparent AR imagery. However further investigations are needed.

Providing real-world occlusion led to better results only in some specific environmental conditions, however, it didn't worsen performances in any of the other tested conditions either. Furthermore, real-world occlusion is able to improve AR imagery in terms of contrast and colours accuracy in almost all conditions, as well as allowing for realising some effects that have no counterparts in real life. For these reasons we believe the providing AR headsets with real-world occlusion capability would allow to use the devices in a wider range of situations.

It would be interesting to perform the same investigations we did using devices with embedded occlusion capabilities when they will be available, This would allow to tests any environmental condition without requiring a fine lighting-control. Although the presented framework is ready to be used, some components could be improved. Support for additional external tracking systems can be added, as well as a better handling of multiple overlapping projectors and multiple workstations can be implemented.

7

IVES FOR SIMULATION AND TRAINING

One of the most important applications of VEs are training simulators for a variety of industries and fields. Examples can be found in the mining industry (Van Wyk and De Villiers, 2009), in the aerospace industry (De Sa and Zachmann, 1999), in the automotive industry (Li, Khoo, and Tor, 2003), in logistics (Bergamasco, Perotti, Avizzano, Angerilli, Carrozzino, and Ruffaldi, 2005) and, in general, in the sector of maintenance (Magee, Zhu, Ratnalingam, Gardner, and Kessel, 2007). Virtual prototyping is an example of the us of VEs in the industrial field. Using VEs it is possible to simulate and render all characteristics relevant to physical mock-ups generating digital mock-ups. The digital mock-up can be used for the verification of assembly and disassembly procedures, assessment of product ergonomics, and visualization of functional simulations. Replace, at least partly, physical mock-ups by software prototypes would allow for a cost reduction as well a faster prototyping loop and therefore to costs reduction (De Sa and Zachmann, 1999).

Immersive VR (IVR), in particular, allows to realize simulators that enable an effective transfer of the skills acquired in the virtual context. To this purpose realistic sensorial feedback and natural interaction must be provided so as to match as close as possible real-life conditions. A number of challenges have been highlighted, ranging from minimizing overall latency, interacting intuitively in the virtual environment, increasing user's perceptual awareness of the virtual world and providing the user with a strong sense of immersion and embodiment. The high levels of presence is an important driver of user engagement and, consequently, impacts on motivation and training efficacy. This is one of the reasons why VR is becoming an important and powerful training tool, as it allows to perform simulated hands-on operations in a controlled and safe environment, reducing costs and risks associated to these activities. Real industrial environments can be dangerous or simply unavailable; training taking place outside the direct working environment can often produce only incomplete experiences and a limited impact. VR-based training can, instead, simulate real-life working conditions but in a safe playground. The idea is to challenge operators with dynamic cases in order to train them to respond quickly in unusual situations, thus enabling them to effectively recognize and recover anomalies and malfunctions (Manca, Brambilla, and Colombo, 2013). VEs provide, in fact, a "sandbox" where certain operations can be performed and learnt safely, under full control, and with the possibility of replicating the experience multiple times, exactly in the same way or with any desired modification. In the industrial field training is important not only to optimize working skills but also to avoid incidents and fatalities. A learning experience can be developed offering a virtual experience in place of or after a classroom-taught lesson on safety. A virtual experience could in fact improve the attention or the awareness on safety, hence the users can be exposed to virtual risks without the dangers of "real-life" experiences. For all workplaces, it is possible to recreate the risk conditions where workers are subjected to. When placed in a virtual environment, users can explore all the solutions and the effects of their actions, including potentially dangerous ones, inside a virtual workplace. VR applications for safety have been already used in some fields. Van Wyk and De Villiers (2009) studied how VR applications could help miners to improve their safety in South African mines, using all the peculiarity of the natural environment. Miners work in confined areas, in steeply inclined excavations, using handling heavy material and equipment and in the proximity of moving machinery. The virtual environment that they reconstruct reproduces these conditions. Previous studies where the VR is applied to the power production fields exists (Cardoso, Prado, Lima, and Lamounier, 2017). Power electric systems require continuous maintenance in order to maintain public safety, emergency management, national security and business continuity. These companies extensively use 2D diagrams as support for servicing activities. For this reason,often operator have difficulties in matching the provided information to the real world machinery in order to carry on the maintenance. VR simulators instead allow to model real objects according to their dimensions, appearance and features. A user trained using VR systems could easily recognize the real scenario allowing for an easier knowledge transfer.

MOTIVATION AND CONTRIBUTIONS Virtual Reality is widely regarded as an extremely promising solution for industrial training as it allows to perform simulated hands-on activities in a controlled and safe environment. Despite the fact that VEs are already used in a variety of contexts for multiple aims, there is a lack of studies which investigate the efficacy of training people using this tools and provide a comparison against traditional teaching approaches. We therefore aim at filling this lack by presenting the results of two users studies. The first one aims at comparing the efficacy in learning assembly/disassembly procedures of a machinery using MR tools with a traditional hands-on teaching methodology. Being difficult to allow a realistic interaction in VEs using own hands—due to the lack of very effective finger-tracking and haptic technologies—the comparison focused on the assembly/disassembly sequences memorization rather than on the manual skill development. The second study aimed at assessing knowledge transfer—of both theoretical and practical notions—and trainee's involvement while instructing operators in performing maintenance procedures—adhering to the related safety rules—using a VR system compared to traditional methodologies. Also in this case the comparison focused on the procedure memorization rather than the manual skill development. Two VR/MR systems—presented in chapter 5—have been developed and used to perform our studies.

7.1 LEARNING ASSEMBLY PROCEDURES

Global industrial manufacturing capacities constitute a large part of the world wealth and economy. A key component of any manufacturing business is training: training a specialized workforce as well as training the customers about the produced machineries requires huge amounts of time, resources and logistic facilities. Training has spill-over benefits for the industry (by providing a pool of skilled workers) and for the society (the improved employment outcomes and flow-on effects such as improved health and lower social welfare costs). Currently, in the field of industrial manufacturing training is a hugely expensive activity traditionally burdened by a number of issues such as the cost of realizing a training environment, the cost of using machineries beyond the working hours, security risks when a trainee uses an equipment and more. These considerations have in time lead to the suggestion that the use of Virtual Reality could introduce significant benefits in the training processes, by removing the need of physical mock-ups in the training process or at least in some of the procedures.

Assembly simulation, is one of the most challenging applications of virtual environments, mostly due to the very high level of interactivity. In a human assembly task, the interaction mostly involves the human hands, so the interaction simulated in a VR application result to be as natural as possible. Furthermore, in manual tasks worker utilizes natural constraints to obtain precise and efficient manipulation of parts and tools. To result natural and so effective, the system can make uses of haptic interfaces ad force-feedback to simulate certain types of constraints. In general an assembly task performed in VR highlights various issues related to the interaction with the system, collisions between virtual parts, and functioning simulations. Previous researches tried to singularly address these problems (Abe, Zheng, Tanaka, and Taki, 1996; De Sa and Zachmann, 1999; Jayaram, Connacher, and Lyons, 1997; Seth, Vance, and Oliver, 2011; Zachmann and Rettig, 2001).

This section presents our ongoing work for training workers in assembling or disassembling complex mechanical machineries.

7.1.1 *Virtual Assembly Trainer*

A VR training system intended to be used by industrial companies who need to train their operators on the tasks of assembly, disassembly or maintain machinery has been developed. The system is based on the mixed reality platform presented in section 5.2. What motivates the use of Virtual Reality is that a real copy of the machine could be cumbersome and expensive, and very likely it might result impossible to work together on the same machine at the same time. Moreover usually an expert assistant is required during the training phase in order to assist the operator. The proposed system provides the needed metaphors to interact and manipulate a 3D model of the machine in absolute autonomy with the purpose of following out a task. This system provides a controlled and safe training environment, in which damages to the real machine are reduced or avoided; hence, inexperienced users can take advantage of virtual training before actually facing the real machine. The system immerse the trainee in a simple VE consisting in a $4mx4m$ room with a 3D model of the machinery placed in the middle of it. The participant is able to freely move inside the environment and naturally interact with the machine with his own hands to perform the assigned assembly/disassembly task. The machinery model is divided into several parts according to the actual machinery composition, so that each piece can be independently grabbed in order to allow a realistic assembly or disassembly procedure. The training instructions are provided to the trainee by a series of instruction tables hung on the walls of the room.

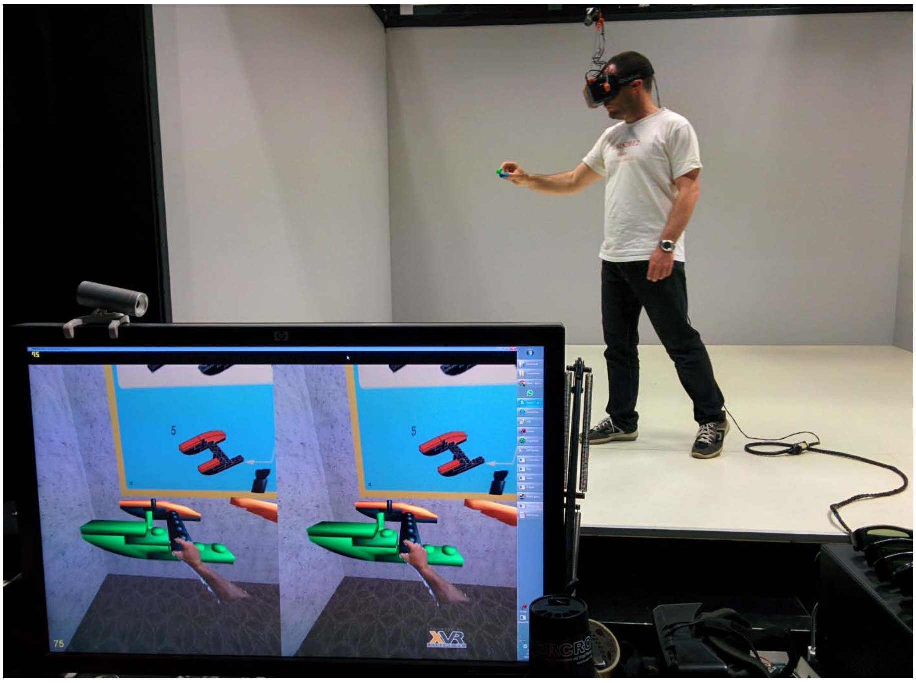


Figure 7.1: A user performing a training session: in green, the group on which he is operating.

The training system consists of two fundamental modes of operation: the authoring mode and the training mode.

Authoring mode

This modality is intended to be performed only once by an expert who owns already deep-knowledge of the machine, of the procedures that can be performed on it, such as maintenance, and of each of the steps that needs to be followed to disassemble it. The expert can disassemble the machine, piece after piece. During this phase, the expert defines steps and sequences. In our mind a *sequence* consists of an ordered list of steps, and each *step* consists of an unordered list of pieces to move. According to this notation, inside a step the pieces can be moved without a specific order, but inside a sequence the steps must be sequentially performed. The expert can also mark a step as a *group* of pieces: this implies that the step will consist only in the translation of the entire group into its target position. The expert thus disassembles the group piece-by-piece, with the possibility to further define nested groups. The concept of group is of primary importance especially in case of complex hierarchical machines: it allows to assemble/disassemble portions of the machine in a location different from the final one, helping the operator to have a more organized and schematic view of the entire

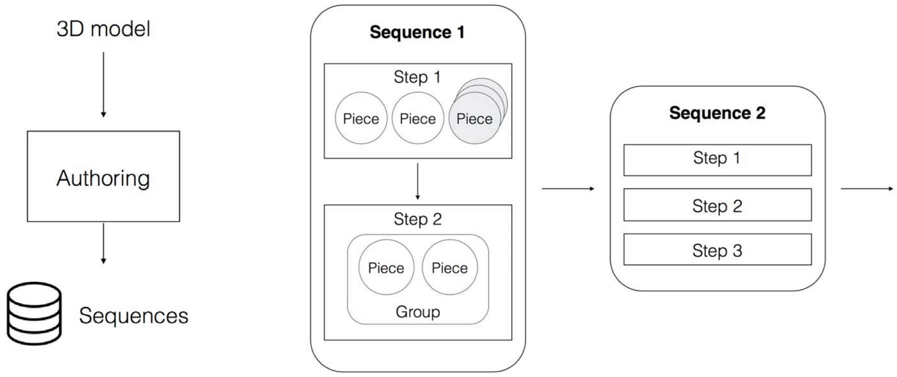


Figure 7.2: The authoring stage produces the sequences structure starting from the 3D model.

machine. If during the disassembly the expert makes a mistake, he can navigate through the steps, undo the changes and start over.

While authoring, the expert places each item in a specific location. This location is marked as starting position for the specific piece during an assembly session: in other words, when the trainee will assemble the machine, at the beginning of each step he will find the pieces exactly in the position the expert left it. If inside a step, two or more pieces are left in the same position (and have the same dimensions) they are marked as "equivalent". In this case, only one of the equivalent pieces is shown with a label indicating the number of equivalent pieces of which consists. This implies that in assembly mode, the operator can place an equivalent piece in the target position of any other equivalent (e.g. if the expert marks some screws as equivalent, the operator will be able to place screws in any well fitting location). During this phase the expert can also take snapshots of the state of the machine from its own point of view. These snapshots are stored and can be used and modified in order to define instructions tables that can be presented to the operators during the training phase. At the end of the authoring session an ordered list containing all the sequences, the steps and the equivalences defined by the expert is saved on file, as shown in figure 7.2.

Training mode

This mode is intended to be performed by operators for training. The trainee can either work on the machine in a "free" mode, or can perform a training session on the machine assembly or disassembly. In the first case the operator can interact with the machine model, without any constraint

Figure 7.3: The training system flow chart. Different operators can use the same sequences to perform different tasks on the same machine.

in terms of sequences and steps, in order to discover how it is made. When instead the operator performs a training session, he's constrained to the specific sequence previously executed by the expert in the authoring mode. The instruction tables relative to the step the operator is performing and the information about the task's progress (remaining pieces, steps completed, sequences completed and time elapsed) are hung on a wall of the virtual room. It is possible to define help layers that can be dynamically presented to the users that encounter difficulties in performing certain operations. A training procedure could train operators to perform the same task several times - with a decreasing level of help - until they acquire the needed familiarity with the machine. The application's flow chart is shown in the figure 7.3.

The simulator can be used to train both on how to assembly or disassembly a machinery. In Disassembly mode the instruction sequences are loaded from file in the same order they have been saved. At the beginning of the task, the machine is completely mounted and in order to move forward on the task, the operator must complete the needed steps by removing the right pieces from their starting position. Since it is a disassembly task, the operator is not required to put the items into a target position; in order to clean up the scene, at the end of each step, the moved pieces are translated into the position the expert left them. The operator is able to move only the pieces that must be actually moved inside that step. When the operator grabs a piece that can be moved it becomes green, while if the piece cannot be moved it becomes red. Each action results also in an acoustic feedback that alerts the users that the action has been actually performed by the system. In Assembly mode the instruction sequences are loaded from the file in a backward order. At the beginning, the first piece is already placed into its target position. The remaining pieces relative to the step/sequence are shown in their start position, namely where the expert left them during the authoring stage. Equivalent pieces are shown together with a label indicating the number of multiple items. The task of the operator is to put all the pieces into their right target position. If the operator leaves the pieces in a closest range of the target, they are automatically snapped to the correct position. Also in this case each action results in an acoustic feedback that alerts the users that the action has been actually performed by the system.

7.1.2 *Pilot Study*

A pilot study to evaluate the effectiveness of the VR training system was conducted using a comparative approach was used: a VR training phase versus a real world's one. We aim at assessing performances in terms of memorization of a machinery assembly sequence both after a virtual or a traditional training phase. A LEGO® Creator Sea Plane 53-pieces model [1] was used in place of an actual machinery, but the fundamentals of the learning experience are the same.

Figure 7.4: The real LEGO® Sea Plane model.

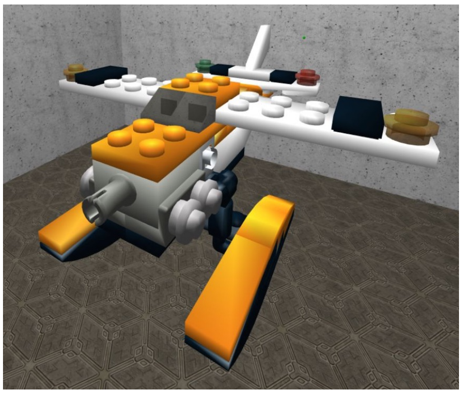


Figure 7.5: The virtual LEGO® 3D model.

The test group consisted of 8 subjects, divided into two sub-groups. Both sub-groups have performed a 30-minutes training, the first one using the real model, and the second one using the virtual simulator showing the digital mock-up of the real model. During the training phase, the subjects

---

[1] *Lego® Creator Sea Plane.* `http://shop.lego.com/en-US/Sea-Plane-31028`. Accessed: 2015-05-08

were provided with the same instruction tables needed to accomplish the requested task and they could use the time at their disposal to perform the task several times, or just to study the model and the relative instruction tables.
After the training phase, each subject has performed the actual task: assembly the actual model of the Sea Plane without time constraints and without the guidance provided by the instruction tables. The completion time as well as the number of pieces correctly placed have been registered and used to assess the performances. The authoring stage described in section 7.1.1 to generate the correct virtual assembly sequence used by the VR simulator has been performed by one experimenter strictly following the original instructions provided with the model.

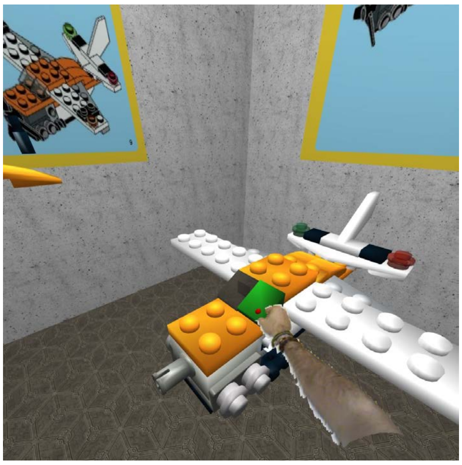


Figure 7.6: Virtual assembly

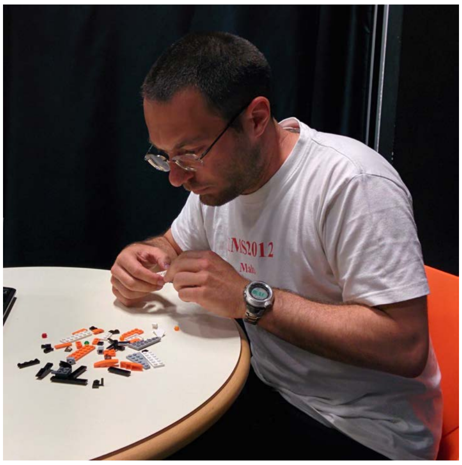
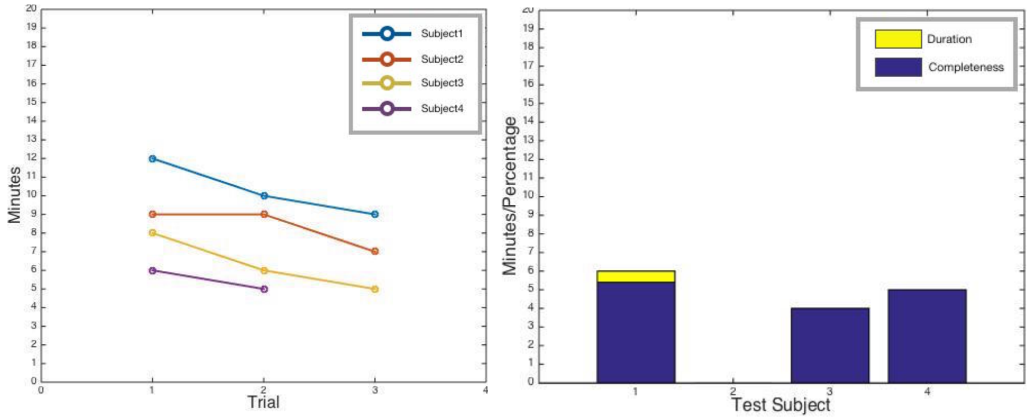


Figure 7.7: Real assembly

7.1.3 *Results*

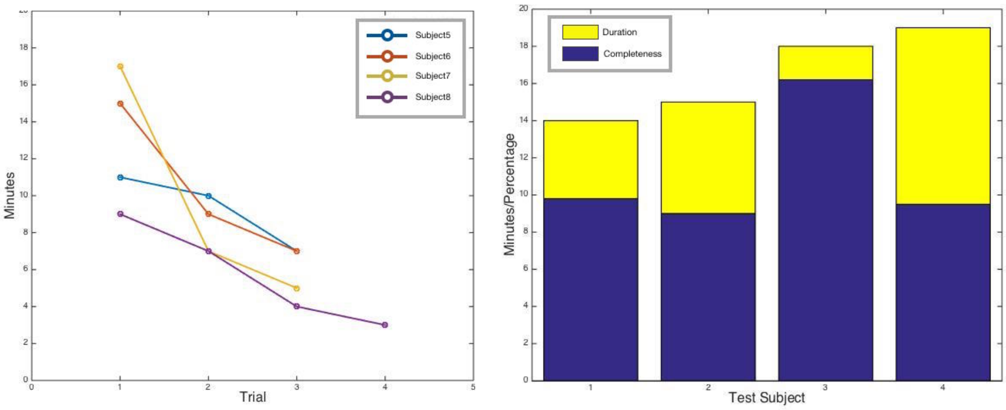
Figures 7.8 and 7.9 show the results of the pilot study. In particular figure 7.8a and 7.9a show the completion times of assembly sequences performed respectively during the real training and the virtual training. Figures 7.8b and 7.9b show the time needed to assemble the real model after the training (in yellow), and the percentage of completion of the model (in blue).

The charts show that the time needed to complete an assembly sequence significantly decreases in the virtual training with respect to the number of trials. The first trial in virtual training requires more time because of the needed familiarization with the environment. Even if the assembly of the real model after the virtual training has required longer time, all the



(a) Duration of training sequences (b) Real model's assembly duration

Figure 7.8: Training on the real model.



(a) Duration of training sequences (b) Real model's assembly duration

Figure 7.9: Training on the virtual model.

subjects have been able to complete the assembly of the model for almost the 50%, with a peak of 90%.

7.2 LEARNING SAFETY PROCEDURES

We've already extensively assessed the importance of educating workers on how to perform their job. However, it is even more important to the workers to learn how to do their job while complying to the safety procedures. There are no more important tasks than safety inspections, especially in power plants and oil refineries. Accidents that can happen in those factories might be more catastrophic and pernicious than anything else. Studies demonstrate that workers are more likely to be exposed to health and safety risks when they underestimate risk due to the confidence on their work environments (Bryson, 2016). All workers must be aware of the safety procedures needed to correctly perform their jobs without risks, not only new employees. As a matter of fact, the risk perception of a worker decreases when the experience increases (Zohar and Erev, 2006); routine in fact can lead to underrate threats. For this reason, even experienced workers needs to be updated and regularly reminded about the safety procedures.

Current industrial safety training methods generally rely on repetitive classroom-taught lessons integrated by directions given in the physical workplace, followed by on the job training (Sheridan, 1992). These lectures sometimes end up being not so engaging or even boring, so exploring alternative or supplementary ways to improve the workers awareness and compliances to the safety procedures is fundamental. Nonetheless finding a more effective way to transfer the knowledge and improving the information retention would be extremely useful. We have therefore developed an industrial Virtual Safety Trainer simulator in collaboration with a power production company, in order to explore new ways to improve the adherence of workers behaviours to the standard safety procedures.

Aim of this research is to develop a VR safety training simulator in order to assess the effectiveness of innovative tools end estimate benefits in terms of knowledge transfer and workers' adherence to the safety procedures.

7.2.1 *Virtual Safety Trainer*

The simulator is based on our technological setup described in section 5.1. The simulator consists of a set of real case scenarios, where workers can be trained in performing exercises, inspections or maintenance procedures. All the procedures have been designed in collaboration with and validated by experts in the field. The virtual environment have been logically modelled to simulate the real counterpart and respond to the users actions as it would do in real life. It is therefore possible for the trainee to experience, in a safe way, the outcomes of wrong actions when performing dangerous procedures. Accidents due to incorrect procedures or lack in observing safety rules are simulated: electrical short-cuts, explosions, burning and electric arcs are reproduced. Providing this kind of feedback in real-life training sessions is not possible or too dangerous for the trainee. Using VR, workers can be therefore safely practice procedures using a learning-by-doing paradigm, and recognize and recover anomalies and malfunctions.

The simulated scenarios include maintenance or securing procedures of low and medium voltage apparatus, inspections of confined spaces, and maintenance or securing of mechanical parts. Figure 7.10 shows an example of scenario. For each procedure a work plan describing the exercise to be performed is provided. Each scenario starts in a dress room where the worker can read the work plan, wear the Personal Protection Equipment (PPE) and choose all the tools needed to perform the task. A lift placed in a corner of each room allows the user to move across different rooms.

The simulator allows one or more trainee and trainers to communicate verbally and also to share the virtual experience. The system can be used both as a training tool, as well as as a validation tool. During a validation case the worker is asked to perform a specific procedure while the supervisor can see what the user is doing inside the VE.

Figure 7.10: Scenario containing low- and medium-voltage apparatus.

7.2.2 *Method*

In order to evaluate the effectiveness of using VR systems for safety training, we first reviewed the literature about evaluation methods of general training practice (Gavish, Gutiérrez, Webel, Rodríguez, Peveri, Bockholt, and Tecchia, 2015). Kirkpatrick developed one of the most used models for measuring the effectiveness of a training procedure(Alliger and Janak, 1989; Kirkpatrick, 1967, 1975, 1979) and provide guidelines for designing effective training. The model defines four levels of evaluation: reaction, learning, behaviour and results.

1. *Reaction level* rates what the participant though and felt about the training session and gives a measure of the user's satisfaction. Reaction level is measured collecting questionnaire, interviews and surveys after the learning session and analysing the verbal reactions of the users.
2. *Learning level* measures knowledge and/or skills increase. This level can be estimated during the training session through a knowledge demonstration or a test. More in details, pre-post knowledge questionnaire and performances tests can be used. Predetermined scoring and coherence between methods of evaluation and training targets can help to minimize risks of inconsistency of results.
3. *Behaviour level* measures the degree to which participants apply what they learned during training on the job and their changes in attitudes. This measurement can be, but is not necessarily, a reflection of whether participants actually learned. For example, the failure of behavioural change can be due to other circumstances such as individual's reluctance to change. The evaluation involves both pre- and post-event measurement of the learner's behaviour.
4. *Results level* seeks to determine the tangible results of the training such as: reduced cost, improved quality and efficiency, increased productivity, employee retention, increased sales and higher morale.

Some authors suggest to add a 5th level in order to measure the return on investment as evaluation parameter (Phillips, 1996). The power of the model is its simplicity and its ability to help people think about training evaluation criteria.

The power producer company—we worked in collaboration with to carry out this tool—currently utilize a traditional training approach. The knowledge transfer is achieved through classroom lecture, manuals and in-loco practice on real machinery. Physical practice with real equipment cannot be easily replaced by VR practice due to factors as a still inadequate haptic feedback. However the theoretical knowledge some practical skills could take advantage of the virtual approach.

In order to evaluate effectiveness of VR training in comparison with a traditional learning methodology an experiment has been conducted.

The subject of the training session is to instruct users how to correctly perform inspections of confined spaces focusing on the adherence to the safety procedures. In order to perform the comparison two groups have been formed: the Control Group (CG) and the VR Group (VRG). The control group has been taught using only traditional techniques. Part of the lectures of the second group have been conducted in a VE.

We have followed the guidelines provided by Kirkpatrick to evaluate the training, in particular the first two steps and part of the third of his model. The experiment have been conducted on workers which have no previous experience on the job, so the last step and the change in attitudes of the third step of the Kirkpatrick's model doesn't apply to our case.

A strong sense of presence has proved to carry the potential to aid training transfer (Li, Daugherty, and Biocca, 2002; Lombard and Ditton, 1997; Sheridan, 1992). We used the Witmer and Singer (1998) Presence Questionnaire (PQ) in order to estimate the sense of presence perceived by users during the VR evaluative session. Control, Sensory, Distraction and Realism are the main factors of the sense of presence.

Procedure

The simulator provides several scenarios where to perform different procedures. Among these, we choose the inspection of a confined space as test scenario for evaluating the impact of the system on the effectiveness of the training session.

The participants to the experiment have been recruited among colleagues and students with no previous experience on the subject of the training. A total of 24 subjects, aged between 24 and 46 (31.36 ± 6.3) took part to the evaluation, 12 for each of the two groups.

The training sessions of both groups have started with a traditional classroom lecture. The trainer instructed the workers about general safety rules that have to be observed when performing any job inside the power plant. During this lesson slides have been shown and commented by the trainer. The lecture explained the importance Personal Protective Equipment (PPE), and the correct procedures and tools needed to operate both in voltage risk environments and in confined spaces. Additional information were provided, as law references, the importance of reading the work plan and of a clean workspace.

After the first common lecture, each group have been taught how to perform an inspection of a confined spaces with different approaches. The control group watched a video of an expert performing the inspection in the correct way: the operations to undertake to accomplish the task and safety rules that need to be respected. The VR group users have been instructed by a trainer about the same procedure while immersed in the VE. The video watched by the CG users consisted in a screen capture of an expert conducting the inspection in the VR simulator, allowing the CG users to familiarize with the VE too. In order to evaluate differences in learning between the two approaches, two different criteria have been adopted: an evaluation based on pre-training (PreQ) and post-training questionnaires (PostQ) and an evaluation of the user's behaviour while performing the procedure in VR. The experiment have been therefore split in two phases, conducted on different days.

During the first phase, users first have answered the PreQ, then have taken part to the two lectures, and finally have answered the PostQ. Comparing the pre- and post-training questionnaires is a classical way—suggested also by Kirkpatrick and widely adopted in different fields—to measure the effectiveness of a training session. During this phase "reactions" (Kirkpatrick, 1979) of the users have been collected.

The second phase of the evaluation took place 5 days after. Each subject have been asked to perform the inspection procedure using the VR simulator without any aid of the trainer. The experimenters have therefore observed and collected all the actions performed by the trainees and the adherence to the safety procedures. Finally, the presence questionnaire have been presented to the users.

A scheme of the procedure adopted for the evaluation is shown in figure 7.11.

The inspection procedure taught during the course took place in a furnace room (see figure 7.12a), while the evaluative inspection took place in a condenser room (see figure 7.12b). Despite of the different scenario, the task to be performed and the safety procedures to respect are the same. As usually happens in real life training session, the workers are trained on procedures which are generally valid for a variety of scenarios.

Metrics

The pre- and the post-training questionnaires are composed of the same 42 true-false questions on the contents of the courses. When evaluating the results, to each question is assigned a score of 1 point if correctly answered and 0 points if not. In this way the total score of the PreQ and PostQ

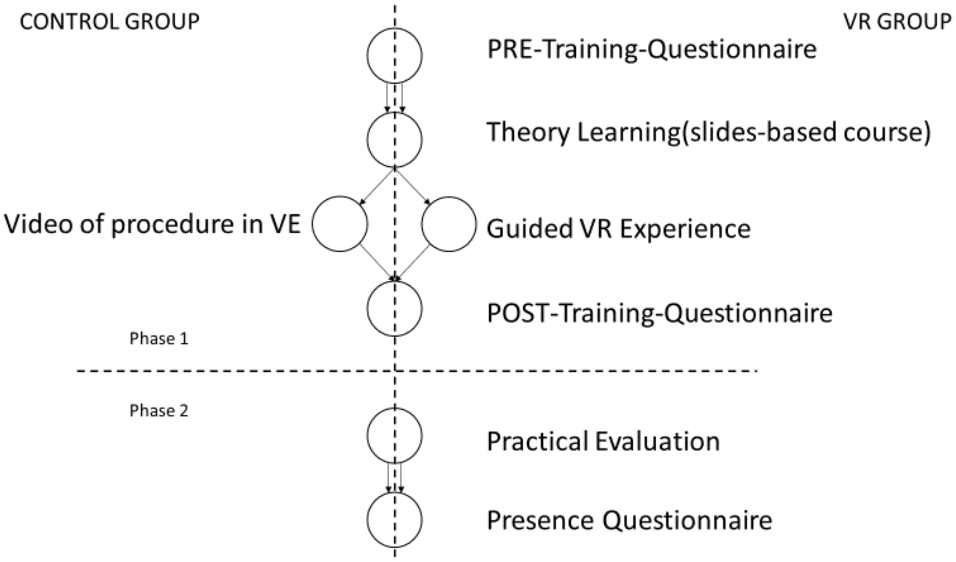


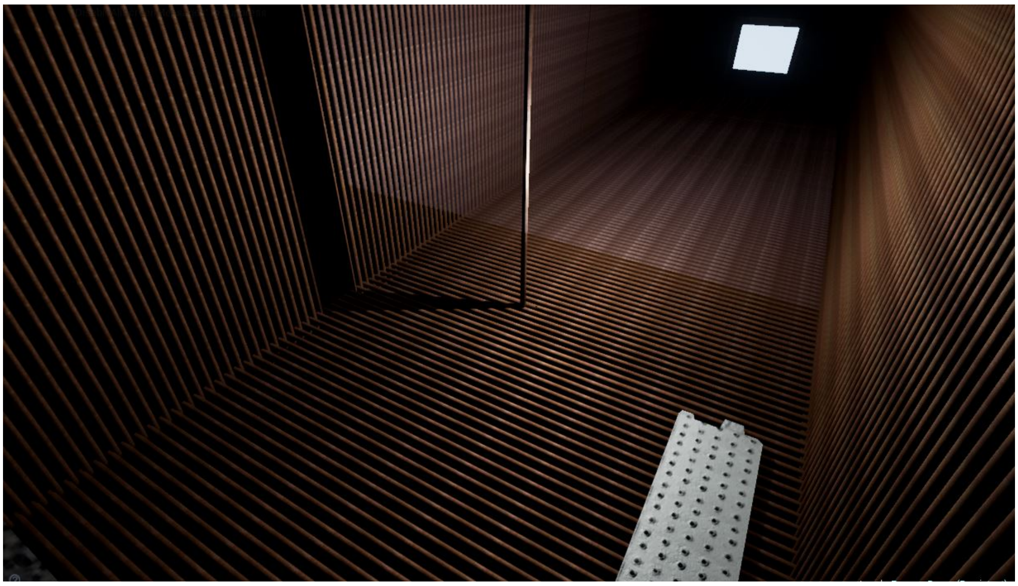
Figure 7.11: Scheme of the adopted experiment procedure for the two groups.

can be directly compared in order to evaluate the level of the knowledge transfer obtained by the two training approaches.

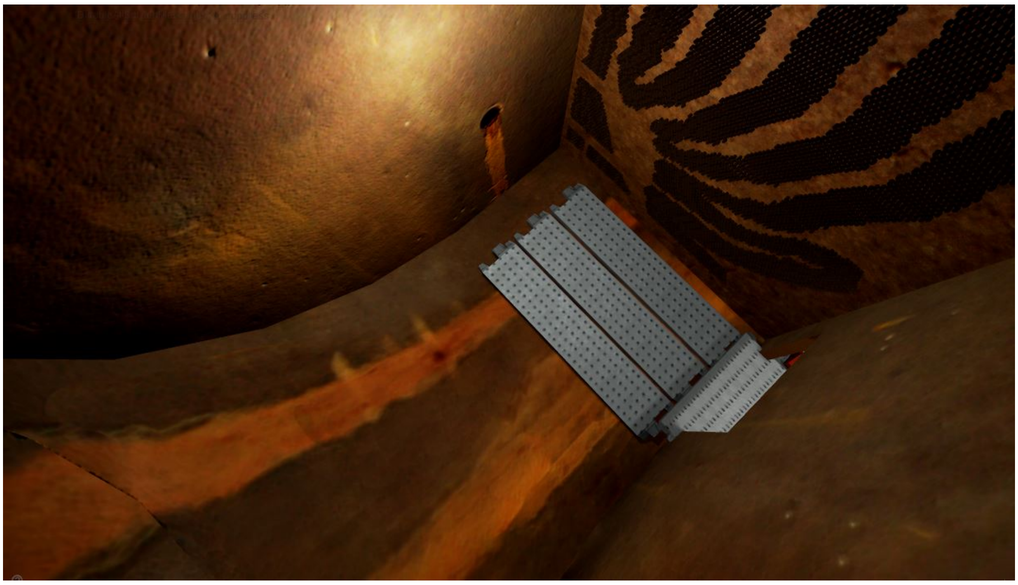
Questions about the pertinence of the questions to the contents taught in the course, the involvement and the perceived effectiveness of the courses have been added to the PostQ. These questions aim at collecting the “reactions” of the users in a 6-point Likert scale.

Similarly, a score is assigned to each action performed during the evaluative inspection procedure. One point is assigned for each correct PPE worn by the user and for each tool correctly used. One point is assigned to each correct step performed by the user to accomplish the inspection task. In this way it is possible to have a second comparison of the two different teaching approaches in terms of practical knowledge transfer. While the questionnaires’ comparison aims at evaluating a more theoretical knowledge transfer, the second approach is closer to a practical evaluation of the training approach.

All the collected scores have been normalized on the maximum score, and are presented as a percentage score.PQ results have been normalized on the maximum score, and are presented as a percentage score.



(a)



(b)

Figure 7.12: The furnace (a) and the condenser (b) confined spaces scenarios.

7.2.3 *Results*

Aim of the work is to compare the effectiveness of VR for safety training in comparison with a traditional classroom-based approach. In order to evaluate the differences in learning—both on theory and when performing an actual procedure in VR—we have asked the subject to answer questionnaires and have collected their behaviours. To statistically compare the differences both in questionnaires and in the collected scores the Wilcoxon signed-rank test have been used.

Part 1: Learning the theory

As shown in figure 7.13, significant differences have been found between the PreQ and the PostQ for both the control and the VR groups. Knowledge of the users improved after attending the course for both groups (CG: $W = 10.00, p = 0.022$; VRG: $W = 0.00, p = 0.003$).

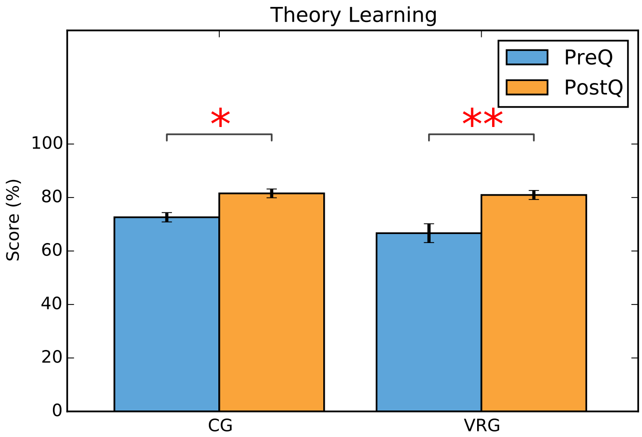


Figure 7.13: Theory learning results for both groups. Bars reports 25th and 75th percentiles. $^*p \leq 0.05$, $^{**}p \leq 0.01$.

The base knowledge of the users before the training was the same for the two groups. As shown in figure 7.14 no relevant differences in the PreQs between the control and the VR groups have been found ($W = 22.5, p = 0.19$). Similarly, according to the PostQ results, no relevant differences in learning have been found between the control and VR group, meaning that both the teaching methods have resulted to be equally effective ($W = 23.5p = 0.68$).

Part 2: Performing procedures

Significant differences have been found between the control and the VR groups during the evaluative inspection procedure. A better compliance to the safety rules and to the steps to undertake to accomplish the task have been obtained by the VRG ($W = 1.5, p = 0.005$) as shown in figure 7.15a. Similar results have been obtained even if we distinguish between the equipment collection task ($W = 1.5p = 0.004$) and the actions performed to a perform the inspection ($W = 10.0p = 0.022$). Individual statistics are presented in fig,7.15b. Those findings agree with the embodied cognition theories which suggest that memory can be aided by performing actions with own body (Barsalou, 2008; Chao, Huang, Fang, and Chen, 2013). The

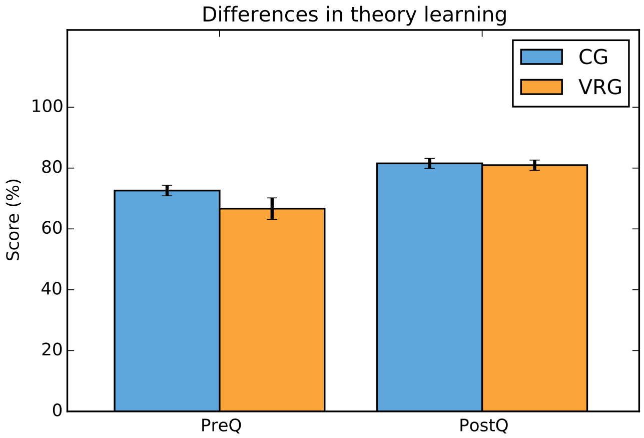


Figure 7.14: Comparison between groups of theory learning according to PreQ and PostQ. Bars reports 25th and 75th percentiles.

training in VR is performed in fact by the user naturally using his/her own body as in real life.

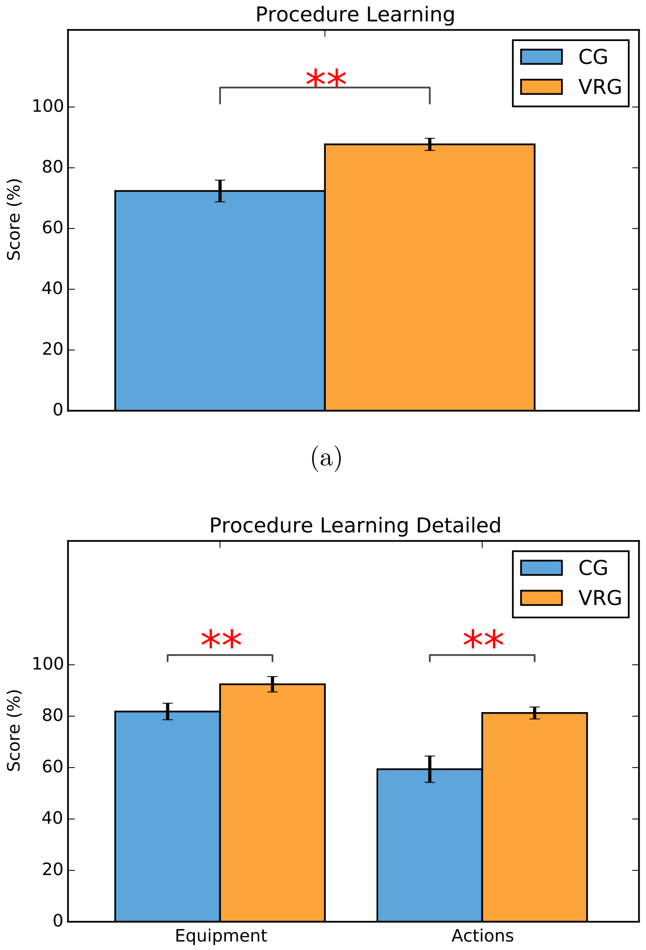
Trainee Involvement

An important factor to be taken into account when evaluating a training course is the overall involvement of the trainees. Both the groups have first taken a classroom course on general safety rules, and then a second course on a specific procedure (different methods for the two groups). Both groups indeed have shown a greater involvement in the second part of the course: video watching for CG ($W = 0.0, p = 0.002$) and VR simulation for the VRG ($W = 0.0p = 0.004$) as shown in figure 7.16 shows the difference.

Furthermore, as shown in figure 7.17, a very significant difference have also been registered between the two courses on the procedure taken by the two groups. The VR experience resulted in a greater involvement of the trainees with respect to watching the video (W = 4.0, p = 0.03), meaning that the VR course have resulted to be much more engaging.

Sense of Presence

Presence Questionnaire results are shown in figure 7.18. The positive factors have obtained high scores (Control: $80.61 \pm 10.05\%$; Sensory: $82.70 \pm 12.17\%$; Realism: $72.04 \pm 12.67\%$), while the negative Distraction factor have obtained a low score ($33.00 \pm 18.54\%$).



(a)

(b)

Figure 7.15: Comparison of procedure learning between groups (a) and (b). Figure (b): distinguishing between collection of the equipment and actions performed. Bars reports 25th and 75th percentiles. $^{**}p \leq 0.01$.

7.3 DISCUSSION

The results of the pilot study conducted using the assembly simulator, although preliminary, are so promising to lead us to believe that this form of training has the potential to constitute a valid alternative to a more traditional training approach. A significant improvement to the proposed system could be brought with the introduction of a more enhanced interaction metaphor. To date, the interaction system allows only the translation of the virtual objects present in the scene. We deem important to allow also rotating and scaling the objects, providing users with a more direct and natural interaction. The addition of a multi-modal feedback to the interaction system would also certainly lead to more convincing results.

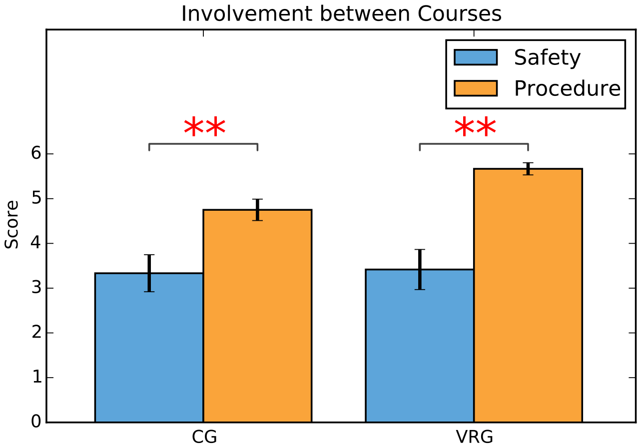


Figure 7.16: Comparison of trainee's involvement between the two parts of each course. $^{**}p \leq 0.01$.

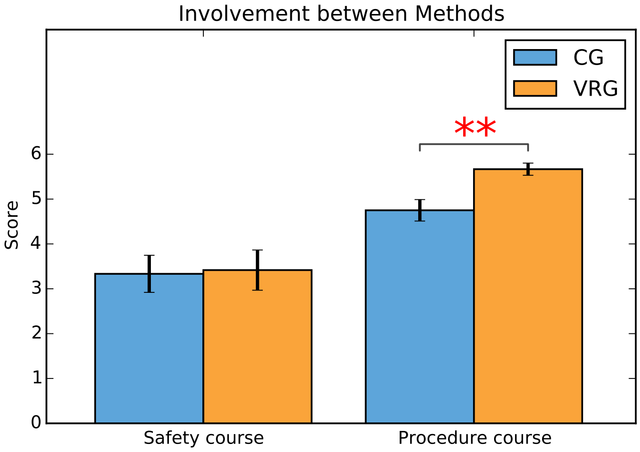


Figure 7.17: Comparison of trainee's involvement between the two groups. $^{**}p \leq 0.01$.

Since we are encouraged by our findings, we are already planning to conduct a series of user studies with more subjects using the new interaction system to operate on several models, in order to further assess the benefits of our approach.

Results of the second study showed that the VR training turned out to be as effective as the classical approach in terms of learning the theory directly measurable with questionnaires. The lack of significant differences between the two training approaches suggests that methods are interchangeable. However higher involvement have been reported by users instructed using VR. A higher involvement means a greater engagement and user satisfaction. Therefore even if the two methods are similar in terms of knowledge transfer, the newer approach is preferable due to the better user involvement. An

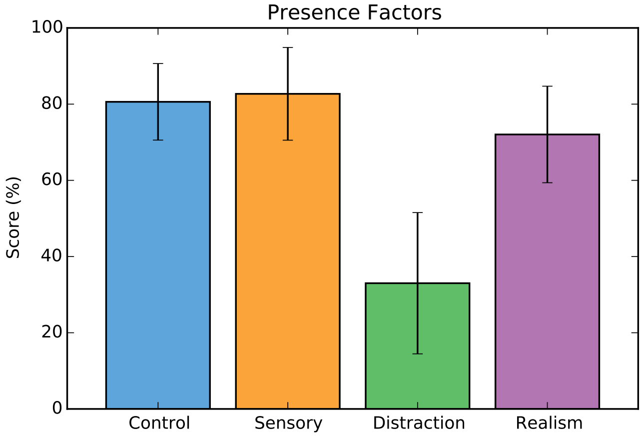


Figure 7.18: Results of the Presence Questionnaire distinguishing between factors.

high level of sense of presence have been registered during the evaluative VR sessions for both groups. A strong sense of presence has proved to carry the potential to aid training transfer. Furthermore, when the evaluation criteria take in consideration the workers behaviour when performing the procedure, the VR training resulted to be significantly much more effective than the traditional approach. The users performed better both in terms of adherence to the safety rules and in terms of correctly performing the assigned tasks. VR training could indeed raise the level of reception and memorization of physical procedures. Two variables should be taken into account when evaluating the practical performances: effect on memory and knowledge transfer. How much the improvement of the VR training are due to a better memorization? How much to a better knowledge transfer? In the experiment conducted it is not possible to distinguish between the contributions of the two factors, because we have only observed the combined results after some days. Thus, could be interesting to evaluate separately the two effects. In order to decouple the contributions we would like to conduct further experiments varying the amount of time between the training and evaluative sessions.

According to the results we think that the introduction of VR as training methodology could lead to better knowledge transfer as well as guarantee a more pleasant and engaging learning experience.

8

COLLABORATIVE MIXED REALITY

Since the beginning, video games took into account the sharing experience of playing. With the advent of Internet the importance of social interaction in games has dramatically improved. Nowadays, thanks to the ubiquitous network availability, videogames provide a never ending social experience, creating sometimes an alternative reality, made of friends and foes, and rich of social interaction.

The importance of the social component as motivation for playing both in digital and non-digital games has been pointed out by many studies (Cole and Griffiths, 2007; Colman, 2013; Yee, 2006a). Lazzaro (2004) describes the main reasons that induce people to play games asserting that social interaction is one of them: "*It's the people that are addictive not the game*".

Social interaction in video-games is mostly defined by the ability to communicate with the other players. The less is the effort required in order to learn how to interact with the others, the better is the experience gamers will have. Being able to effectively interact with the partner is indeed extremely important. Furthermore by providing additional ways to interact in games, players can customize their experience, choosing the most familiar and effective form of communication. The social component is therefore heavily influenced by the technologies enabling the communication. The role of the technology becomes crucial when the game takes place online and players are not physically co-located and consequently the communication relies only on the technological layer.

Nowadays the hardware evolution has lead to powerful solutions able to substantially improve the immersion and interaction of players with and within the Virtual Environment (VE). Commodity immersive visualization devices, like Head Mounted Displays (HMDs), are nowadays commonly used by gamers. Starting with the adoption of touch surfaces by mobile devices, and today present in several different products other than mobiles, passing through the new console controllers—among which Microsoft Kinect and Wii Remote are two noteworthy examples—we are assisting to a new generation of interfaces which are narrowing the gap between the physical world in which we live and act and the virtual world where we play. The interaction between the players and the game is becoming more natural.

The spreading of these technologies is reshaping the way we play and think about gaming. Natural User Interfaces (NUIs)[1] are flourishing and becoming more and more popular. From an interaction between the real world and the virtual game world mediated by abstract metaphors, among which Keyboard & Mouse represents the most popular and used one, we are slightly moving to a more physical interaction. New richer interaction metaphors can be designed indeed in order to improve the game engagement in a social scenario.

There are many social structures that can be used to define the interaction between the players and each one has specific pros and cons (Zagal, Rick, and Hsi, 2006). Choosing the right design for a social game is extremely important. Depending on how the social interaction in the game is conceived, the resulting game experience may heavily differ. Furthermore each game mechanic may exploit differently the technological medium which enables the interaction between the players.

MOTIVATION AND CONTRIBUTIONS The implications of the technological revolution on the social interaction in entertaining applications have not been yet extensively addressed. The present work aims at assessing the impact of the new technological solutions—NUIs and immersive displays—on user engagement and social presence both from a technological perspective and from a game mechanic design point of view. A comparison of the level of engagement and social presence achieved with a classic interfaces like Keyboard & Mouse with a novel natural interface combined with an immersive visualization system is performed. A comparison of two popular games mechanics—competitive vs collaborative—in terms of player engagement and social presence when playing with an immersive NUI is conducted.

8.1 RELATED WORK

Gaming, in its widest meaning, plays a special role in the society and goes much beyond the merely act of playing. Huizinga holds that play is older than culture and goes beyond the confines of purely physical or

---

[1] Natural User Interface identifies human-computer interactions based on typical inter-human communication. These interfaces allow computers to understand the innate human means of interaction (e.g. voice and gestures) and not induce humans to train to the language of computers (e.g., keyboard and mouse).

purely biological activities. It has a significant role in all aspects of life and underlies human society (Huizinga, 1944). Games, contrary to popular belief, are one of the main and best example of social activities as they promote the socio/emotion-relational skills as collaboration, negotiation and respect of agreed rules (Hromek and Roffey, 2009). As claimed by G. Mead is through play that human beings learn about the social world, because the game allows the emergence of the self in relation to the other and it acts as a mediator of the processes in taking of roles (Mead, 1934).

Starting from the early '70s, the games world experienced one of its greater changes. With the first computer game ever released to the public, Nolan Bushnell's *Computer Space* started the era of Video Games. Before the arcade brought video games to the public, games were available only to those who had access to computer labs at universities or corporations. Starting with *PONG* in 1972, video games found success alongside pinball games in arcades (Wolf, 2008).

As gaming is intrinsically a social experience, since the beginning video games took into account the sharing experience of playing. Some of the earliest video games were two-player games, where a single computer or console connected to a display allowed more than one player to play together. With the advent of Internet the social aspects of games have been dramatically improved. Today games leverage on ubiquitous network availability to provide a never ending social experience, providing sometimes an alternative reality to the player, made of friends, foes and rich of social interaction.

The promising experience of arcade games was just the preview of the capabilities of this phenomenon which today is one of the more flourishing business in the world. Just in the United States, according to the statistics published by the Entertainment Software Association (ESA), the video game industry achieved retail sales of $15.2 billion in 2012. The real annual growth rate achieved by the U.S. video game industry exceeded 9.7 percent for the years 2009 through 2012. During the same years, real growth for the U.S. economy as a whole was only 2.4 percent (Siwek, 2014).

Given the great importance of gaming in our society, there is a great interest in finding the best design guidelines defining the correct principles to create a successful and engaging game. It is therefore important to understand the main motivations that lead people to play. Lets start from the most simple assumption on which everyone can agree: if players do not enjoy the game, they will not play the game (Sweetser and Wyeth, 2005). Therefore player enjoyment is the most important goal for game designers.

The greater is the game engagement the better is the user experience: maximizing the game engagement is the objective all the designers.

There are multiple factors contributing to the overall game engagement (Brockmyer, Fox, Curtiss, McBroom, Burkhart, and Pidruzny, 2009; Jennett, Cox, Cairns, Dhoparee, Epps, Tijs, and Walton, 2008; Seif El-Nasr, Aghabeigi, Milam, Erfani, Lameman, Maygoli, and Mah, 2010); flow and social interaction are two of them.

Flow has been defined by Csikszentmihalyi as a state of mind in which a person is completely involved and immersed in an activity (Chen, 2007; Csikszentmihalyi, 2014). We can experience the flow in several every day tasks: when reading a book; when driving a car; when playing a videogame. The concept of flow is central to game evaluation (Brockmyer, Fox, Curtiss, McBroom, Burkhart, and Pidruzny, 2009; Jennett, Cox, Cairns, Dhoparee, Epps, Tijs, and Walton, 2008; Qin, Patrick Rau, and Salvendy, 2009; Sweetser and Wyeth, 2005).

Social interaction represents another key aspect when designing videogames. People often play games to interact with others, regardless of the task (Sweetser and Wyeth, 2005). The importance of the social component as motivation for playing both in digital and non-digital games has been pointed out by many researches (Cole and Griffiths, 2007; Colman, 2013; Yee, 2006a). This can be explained by the fact that in playing with other people an increment in the involvement as well as in the enjoyment of the game itself occurs. Lazzaro (2004) describes in his work the main reasons that brings people to play games. She observed many emotions from gameplay in facial gestures, body language, and verbal comments and she identified four main keys or pathways leading players to emotion in games. Social interaction is one of these four main aspects. Players use games as mechanisms for social experiences: “It’s the people that are addictive not the game”. Yee in his works (Yee, 2006a,c) offers a clear slice of the characteristics and behaviours of Massively Multiplayer Online Role-Playing Games (MMORPGs) players. He analysed the motivation of over 30000 MMORPG users. He identifies three main components that define the motivation to play online games: achievement, immersion and, again, the social component.

Given the relevance of these two factors on gaming experience, it is extremely important to maximize their effects when designing a game. Since the game engagement is largely influenced by these two parameters, one could think that the solution should be nothing other than maximizing social presence and flow in game. But the social interaction can have powerful and contrasting effects on the user experience.

Social interaction in video-games is mostly defined by the ability of players to communicate with the others and, more than other aspects in videogames, it is heavily affected by the technological facilities used. This difference is even more important when players are not physically co-located and the communication among them relies only on the technological layer. Using different hardware solutions new interaction metaphors can be enabled: motor activity-centered games exploit new console controllers (e.g. Wii Remote); depth sensors allow full body interaction.

Nowadays the increasing availability of novel hardware devices (e.g. Depth cameras, HMDs, inertial sensors) provides new and interesting alternatives to game designers. Game designers can take advantage of them in order to create new *communication channels* or improve the existing ones. It is possible to develop novel, powerful and extremely immersive social experiences overcoming the existing communication gap between co-located and remote players. It is today possible to de-materialize the players and teleport them in a shared virtual world where the game takes place.

Gajadhar, De Kort, and Ijsselsteijn (2008) evaluated the effects of co-player presence on player enjoyment according to three common two-player settings (virtual, mediated, and co-located). They used a basic technological setup in which subjects play *PONG* varying the closeness of the players. They found that players enjoy more the co-located setting due to the increased affordance for communication.

Sajjadi, Cebolledo Gutierrez, Trullemans, and De Troyer (2014) investigated whether the choice of interaction mode/controller has an impact on the game experience. They tested a collaborative game using the Oculus Rift and Sifteo Cube[2]. They didn't found any significant difference between the two interfaces on the game experience. They instead observed that almost all participants using the Oculus Rift looked for alternative way of communication trying to use gestures to interact with the partner even if not enabled by the technological setup.

Lindley, Le Couteur, and Berthouze (2008) focus on the impact of the new interfaces involving body movements on player engagement and social behaviour. They found that the amount of social interaction is higher when using input devices which allow body movements, resulting in an higher engagement in the game.

---

[2] An interactive game system built on building blocks and domino tiles.

Kauko and Häkkilä (2010) compared the effect of two different technological setups on social interaction. Subjects played the same multi-player game first on their mobile phones facing each other and then on a typical game-console setting side-by-side. They found an increase in the social interaction in the first setup which enables a socially richer game experience.

Even if both flow and social factors contribute to an increment of game engagement, it is still not clear which kind of interactions occur between them. Sweetser and Wyeth (2005) assert that social interaction, being not an element of flow, can interrupt immersion in games, as real people provide a link to the real world that can knock players out of their fantasy game worlds. Similarly Lindley, Le Couteur, and Berthouze (2008) suppose that by encouraging social interaction, players will in some sense have been drawn out of the game environment and into the real world breaking the flow.

All the previous works highlight the importance of communication between gamers in multi-player video-games. Being able to effectively interact with your partner is extremely important. At the same time the social communication works as a link with the real world because often happens outside of the game. This can affect the flow experience which represents another fundamental aspect for game enjoyment. New technologies offer powerful alternatives to enable new ways of communication.

At the same time when developing a social game, particular attention must be put to the social mechanics adopted. Depending on how the social interaction is conceived, the resulting game experience may heavily differ. In traditional game theory, multi-player games fall into two basic categories: competitive or cooperative. Despite not being acknowledged in game theory for a while, a third category exists: collaborative games (Zagal, Rick, and Hsi, 2006).

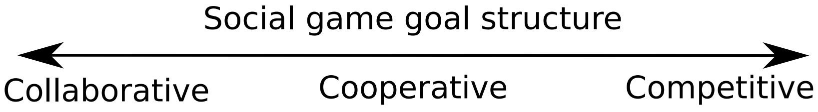


Figure 8.1: Social game goal structure: a graphical representation of the continuum design space between the different social game goal structures: competition, cooperation and collaboration.

The three different goal structures can be seen as a continuum of design solutions. Figure 8.1 schematizes the concept placing on a side pure competitive games in which the goals of players are diametrically opposed: only one player can win by defeating the opponents. On the other side of the continuum there are collaborative games where all the participants work together as a team, sharing pay-offs and outcomes; if the team wins or loses, each player wins or loses. A team is an organization in which the kind of information each person owns can differ, but interests and beliefs are the same. Between the two extremes there are cooperative games where players may have different goals and pay-offs. Cooperating with other players can increase the score or ease a difficult step, but cooperation is not mandatory in order to play and each user is free to choose his/her own strategy to advance in the game.

Collaborative and competitive goal structures are frequently used in video games since they increase engagement and motivation to play. However, empirical evidence on their impact is limited. It is important to investigate whether incorporating multi-player modes can engage and motivate players more, and if so, how different types of multiplayer goal structure impact on player engagement, involvement, motivation and commitment to the game (Peng and Hsieh, 2012).

When designing a game, choosing the right type of social mechanics might depend on several factors and, in turn, influence the game itself. Designers strive to promote interactions among players, as they recognize that these virtual encounters are essential to the success of their virtual games. What makes a difference for many players is the shared experience, the collaborative nature of most activities and, most importantly, the reward of being socialized into a community of gamers and acquiring a reputation within it. These shared experiences, in turn, can greatly increase the appeal and longevity of the game (Ducheneaut and Moore, 2004).

Ducheneaut, Yee, Nickell, and Moore (2006) studied how social interaction in games depends on the game structure. The authors talk about the complexity of designing player's experiences and how the social interaction is planned meticulously into the game to make it more engaging. The data collected in this work reveals also a sort of addiction, a "social pressure" on the players making them more active in the community. Authors explain a different side of game sociality using the oxymoron "Alone together": playing surrounded by others instead of playing with others. Participants in the game seen as audience increase the sense of social presence and the game turns into a spectacle raising the appeal of multi-player games. Yee and colleagues have shown how much the social motivations are a key dimension of reasons for playing MMORPGs (Caplan, Williams, and Yee, 2009; Yee, 2006b,c). Ducheneaut and Moore (2004) studied the social activities in Star Wars Galaxies and highlight the choices made by the game designers to promote social activities.

Online FPS games are clearly designed to maximize the enjoyment of competition. Due to their competitive nature and the violent theme of shooting, FPS games have often been assumed to be socially isolating for young players. In (Frostling-Henningsson, 2009; Xu, Cao, Sellen, Herbrich, and Graepel, 2011) is instead highlighted the strong cooperative component of this games. Even if this genre includes games mainly designed to provide competitive goal structures, players are able to reshape the way the game is conceived to maximize the social relationship and collaboration. Teamwork and cooperation seemed to be crucial for all online gamers interviewed and observed by the authors.

Siu, Zook, and Riedl (2014) investigated and compared how scoring mechanics based on principles of collaboration and competition impact on the accuracy and engagement of players in commonsense knowledge collection tasks. They focus on Games with a Purpose, games in which players generate useful data or solve problems as a by-product of playing, finding that players competing in the game perform faster and enjoy more the game.

Peng and Hsieh (2012) focus on the differences among competition and collaboration as different game goal structures and on the relationship between players in a motor activity-centered computer game. They did not find significant effects of relationship on game performances, motivation and goal commitment. Their findings indicate that cooperation results in greater motivation and effort than competition. Therefore the authors suggest to include cooperative and/or multi-player modes in exergames for physical activity in order to maximize the motivation and performances.

Competition and collaboration are often used also for educational purposes (Burguillo, 2010; Carrozzino, Evangelista, Brondi, Lorenzini, and Bergamasco, 2012; Plass, O'Keefe, Homer, Case, Hayward, Stein, and Perlin, 2013; Romero, Usart, Ott, Earp, and Freitas, 2012). Social serious games are used with the explicit pedagogical intention of fostering collaborative learning processes, which involve working together towards a common goal, sharing and constructing a certain level of common knowledge, understanding and expertise. Researches in this field focus on the definition of design principles for ensuring that all players learn and collaborate efficiently. Plass, O'Keefe, Homer, Case, Hayward, Stein, and Perlin (2013) claim that friendly competition among learners increases motivation and performances while collaboration is inefficient and error-prone compared with the individual mode. Yet, collaborative play also led to greater intentions to play the game again, suggesting that, over time, this negative effect could be resolved.

Physical play settings are a complex mix of social (with friends vs. with strangers), spatial (at home vs. at the Internet café) and media (side-by-side vs. online) characteristics. The impact of socio-spatial settings on player experience is substantial, and mediated by the social presence of the co-player. Gajadhar, De Kort, and Ijsselsteijn (2008) studied how the game enjoyment varies between settings that differ in potential for social interaction. They have evaluated the influence of a co-player presence on player enjoyment, by systematically varying social play settings increasing levels of objective co-player presence or closeness, and the role of social presence, "the sense of being with another" (Biocca, Harms, and Burgoon, 2003), on the final game experience and enjoyment. Results indicate that, compared to playing against a virtual or mediated co-player, playing against a co-located opponent significantly adds to the fun, challenge, and perceived competence in the game. Differences between mediated and co-located co-play lay in the opportunity for richer social interaction between players during co-located play.

Outcomes of these studies suggest that players experience more enjoyment when the communication bandwidth and immersion are increased. The social cues and opportunities for social interaction directly shape player enjoyment in social play.

8.1.1 *Engagement questionnaires review*

Several methods and questionnaires have been proposed over the time in order to evaluate the game engagement of players.

Qin, Patrick Rau, and Salvendy (2009) worked on a questionnaire composed of seven dimensions They concentrate on the evaluation of the effect of the narration inside the game as a gateway to involve and immerse the user in the game. Although focused on the computer game narrative, their questionnaire is also able to measure user immersion in a story-oriented virtual reality games.

Sweetser and Wyeth (2005) developed a model called Gameflow in 2005. It consists of a set of qualitative criteria for measuring eight specific elements of a game: concentration, challenge, skills, control, clear goals, feedback, immersion, and sociality. The model has been validated by evaluating two commercial games, and comparing their results with the ones given by expert reviews.

Seif El-Nasr, Aghabeigi, Milam, Erfani, Lameman, Maygoli, and Mah (2010) with their work introduced an interesting set of metrics for cooperative games analysis. Starting from the observation of players during game sessions, they found noteworthy events that can be used to quantify the social interaction on different planes. Starting from the work of Rocha, Mascarenhas, and Prada (2008), they identified eleven cooperative design patterns. Then they formalized the Cooperative Performance Metrics (CPMs), associated with the observable events within a play session, and they linked each metric to a particular design pattern. The result is a weighted view of how much each event occurs in the different patterns, giving precious information to game designers.

Jennett, Cox, Cairns, Dhoparee, Epps, Tijs, and Walton (2008) distilled the Immersion Experience Questionnaire(IEQ) aimed at consider players subjective immersion experience from three different experiments. They define the immersion as a combination of flow, cognitive absorption and presence involving a lack of awareness of time, a loss of awareness of the real world and a sense of being in the task environment. The immersion is the result of a good gaming experience.

Finally the research of Brockmyer, Fox, Curtiss, McBroom, Burkhart, and Pidruzny (2009) in the design and definition of the Game Engagement Questionnaire (GEQ) has also been evaluated. The questionnaire is built around five psychological states: immersion, presence, flow, psychological absorption, and dissociation. According to the authors, players experience these different conditions while playing, from the first and easily achievable state of immersion to the intense sensation of dissociation representing the maximum engagement achievable in a game.

The same research group also proposed the Social Presence in Gaming Questionnaire (SPGQ)(Kort, IJsselsteijn, and Poels, 2007), which can be used to evaluate the social experience in games. The questionnaire is organized in three main dimensions: co-presence, psychological involvement and behavioural interdependence. Through the SPGQ the social potential and richness of a game can be estimated.

Immersion, sense of presence and flow are common parameters influencing the user experience inside any virtual application. The more the user will be engaged and perceive the virtual world as real the more fruitful will be the virtual experience, the greater will be the enjoyment when playing. Obviously the technologies adopted have a great impact on this parameters and can therefore shape the user experience.

When dealing with multi-player video-games, where more than one user is present and interact with the VE, other variables must be taken into consideration. Being able to effectively interact and communicate with the partner is extremely important. At the same time when the communication happens outside the game world (e. g. when speaking with a partner physically co-located or using a chat when remotely connected), the social component can work as a link with the real world. This can affect the flow experience which represents a fundamental aspect especially of game enjoyment.

User experience can be significantly enhanced by increasing the communication bandwidth to support natural and unmediated body gestures and create a sense of co-location among players. Therefore, the availability of good and powerful ways to communicate has the potential to contribute significantly to a successful design for a social game (Ducheneaut and Moore, 2004) .

Gajadhar, De Kort, and Ijsselsteijn (2008) studied how settings that differ in potential for social interaction affect the game enjoyment. Results indicate that playing against a co-located opponent significantly adds to the fun, challenge, and perceived competence in the game compared to playing against a virtual or mediated co-player. Differences between mediated and co-located gaming lay in the opportunity for richer social interaction between players during co-located play. The social cues and opportunities for social interaction directly shape player enjoyment in social play.

Video-games are always more frequently designed to be played online. It is therefore important to understand how the social presence in virtual not co-located games can be enhanced reaching the same results obtained with co-located play.

At the same time when developing a social game, particular attention must be paid to the social mechanic adopted. Depending on how the social interaction is conceived, the resulting game experience may heavily differ. Collaborative and competitive goal structures are frequently used in multiplayer video-games since they increase engagement and motivation to play. The effects of the two mechanics—competitive and collaborative—have been widely studied using traditional non-immersive systems. However, empirical evidence of the influences of immersive systems using NUIs on multiplayer games is limited. It is nonetheless important to investigate whether these new technological setups can engage and motivate players more than the classic desktop setups, and if so, how different types of multiplayer goal structure exploit the new system.

The subjects of the study have played a jigsaw puzzle game in a shared VE. Three different experiment configurations have been adopted in order to compare both the technological setups and the impact of the immersive/-natural configuration on game mechanics. The experiment has adopted a within-pairs design. A total of 24 subjects (12 couples) have played the game.

In the following, the technological setup and interaction metaphors under evaluation are introduced. Then the game used to perform the study is described. Finally participants, procedure and metrics are presented.

8.2.1 *Technological setups*

Two identical network connected systems have been used. During the experiment the subjects have played a collaborative and competitive version of a jigsaw puzzle game using two different interaction metaphors mapped onto two different system configurations.

The first interface proposed exploits Keyboard & Mouse, one of the most classic gaming interface, as a medium between the player and the VE. It has been used a standard gaming setup using a 24 inches monitor. Players can navigate the environment by using keyboard and mouse. They can grab and position the puzzle tiles and zoom in and out by using the mouse. Verbal communication is enabled by using headphones and microphones. Each player can see both his and the partner's mouse pointers (see figure 8.2). We refer to this setup as "KM" from here on.

The second interface based on our MR system (see section 5.2) exploit natural interaction between the player and the VE. The tracking relies on the tracking camera provided with the Oculus DK2 system (instead on the custom head-tracking module) and is completed with the addition of a wireless headset with microphone to enable not co-located verbal

Figure 8.2: Keyboard&Mouse setup: snapshot of the system highlighting the players pointers during a KM game session.

communication between players. No haptic feedback is provided. We refer to this setup as “OU” from here on.

8.2.2 *The game*

A collaborative jigsaw puzzle game has been developed for the experiment. This popular game genre combines a low complexity with an high level of attention and interaction. Even if videogames of this type are usually two-dimensional, we have designed the game to be played in a three-dimensional environment in order to exploit the immersive capabilities of the HMD.

The game environment has been designed in order to maximize the space needed by a player during a game session and stimulate participants' movement within the limits of the tracked workspace (see figure 8.4).

The scene is composed of a virtual 3x3 metres room in which a table, 2.5x1.0x0.8 m, is placed on a side. On a wall a countdown timer and a poster showing the puzzle solution are hung. The puzzle is made up of 48 tiles randomly disposed on the two sides of the table. Each tile is represented by a parallelepiped (10x10x5 cm) textured with a part of the puzzle image. A

Figure 8.3: Oculus setup: each column shows a user playing in the physical environment and one of the binocular view the user see.

board defining the place-holders where the puzzle has to be arranged on is placed in the middle of the table(see Fig 8.5).

The game is played by two players at the same time. The participants, physically located in two different places, share the same VE during the game session. The game state and the actions performed by a player are directly visible to the partner (e.g. scoring, tiles movements and positioning). When playing OU sessions the two players can see each other. A proxy for each player, made by a textured mesh reconstructed from the RGBD data, and a virtual head, which replicates user's movements, is shown in the VE (see figure 8.3).

During KM sessions the players can see his/her own and partner's mouse pointers (see figure 8.2). Hence players can see each other (or a representation of them in the VE) and the actions each one performs are directly visible to the other (see figure 8.3). The puzzle tiles are shared by the two players and can be dragged and dropped over the table. Both participants can interact with each tile at any time. When a tile is currently grabbed by a user, the other can not interact with it until the tile has been dropped. When a tile is released near a free board place-holder it is attracted and automatically positioned on top of it. If correctly positioned,

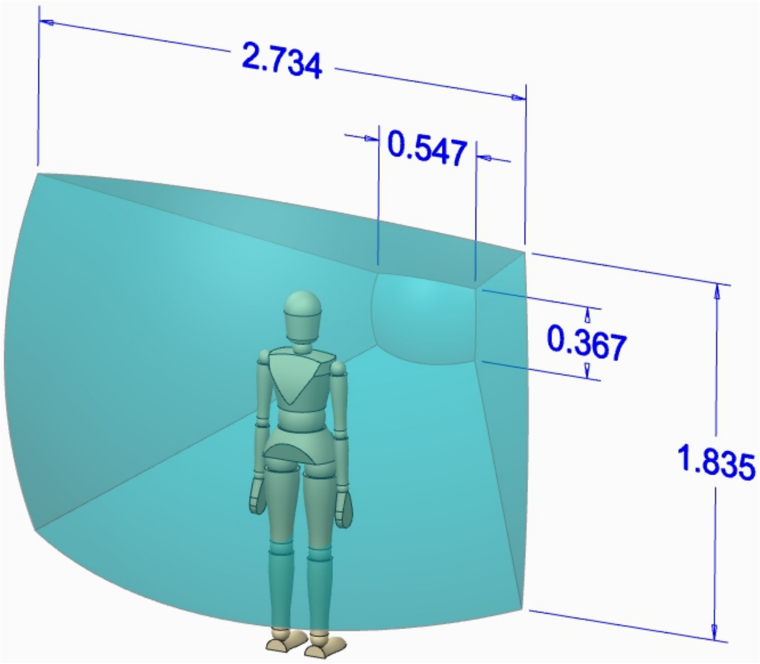


Figure 8.4: Oculus tracking workspace when distant about 2.7 meters from the camera.

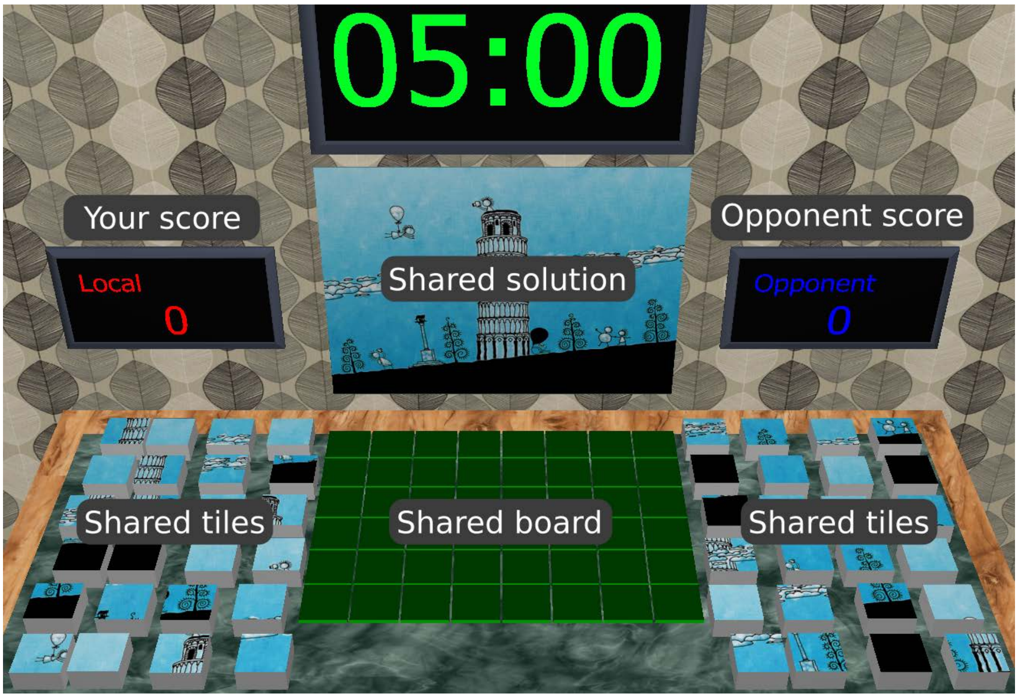
the tile collapse on the place-holder and cannot be moved any more. Sound feedback notify correct or wrong tile positioning.

The two players are characterized by different colours, red and blue. In order to highlight the belonging user, each owned element (score, grabbed tile, etc.) is modulated with the corresponding colour. Players have seven minutes to solve the puzzle by correctly placing all the tiles on the board. The game ends at the puzzle completion or when the time is over.

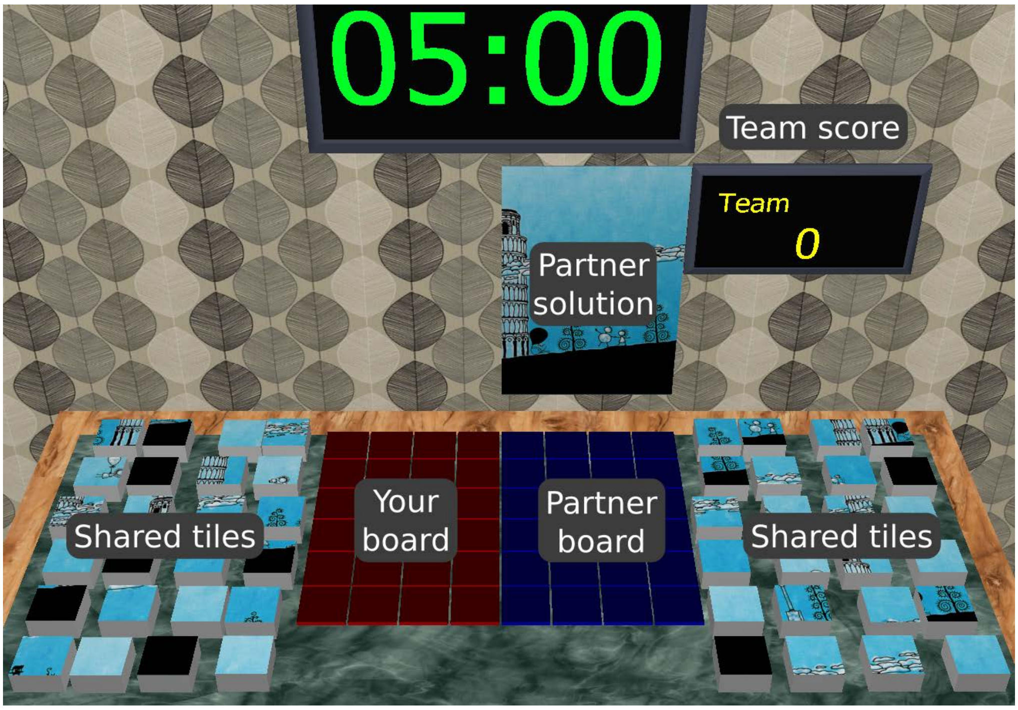
In order to assess the effects on game engagement of the different social paradigms, namely collaborative and competitive, when using the new MR interface two versions of the game have been developed.

Competitive game

Competitive games must promote fair competition by defining clear rules and provide immediate feedback of player achievements (Peng and Hsieh, 2012; Zagal, Rick, and Hsi, 2006). The game has been designed following these principles. The player's goal is to finish the puzzle before the time is over and correctly placing more tiles than the opponent. Both players lose if at the end of the time the puzzle has not been completed. In this setup each player sees the whole puzzle solution and can grab and place all the tiles available. The score of each player is calculated according to the



(a) Competitive puzzle



(b) Collaborative puzzle

Figure 8.5: Game environment: (a) competitive and (b) collaborative game environments

following rules:

- +2 points for each correctly positioned tile
- +1 point every 5 seconds of time left to the winner

- −1 point for each tile left at the end of the time

During the game, the score of each player is updated and shown on two boards placed near to the poster solution (see figure 8.5a).

Collaborative game

The collaborative game is mainly based on the Complementarity and Shared Goals design patterns defined in (Rocha, Mascarenhas, and Prada, 2008). The aim of the collaborative game therefore consists in working together with the partner in order to solve the puzzle before the time is over. In this setup the board is divided in two sections and each player is able to position tiles only on the half board belonging to him/her. If a player tries to position a tile on the half board belonging to the partner, the tile jumps away in a random position. Furthermore, each player can see only the half solution belonging to the partner. Hence, in order to complete the game, players depend on each other and need to help the partner in positioning the tile. Players share a common team score calculated according to the following rules:

- +2 points for each correctly positioned tile
- +1 point for each 5 seconds of time left
- −1 point for each tile left at the end of the time

The team score is displayed on a single board on a side of the poster solution (see figure 8.5b).

8.2.3 *Participants*

The participants to the experiment have been recruited among colleagues and students. A total of 24 subjects, 15 males and 9 females healthy subjects, aged between 23 and 50 (32.04 ± 6.84) have taken part at the experiment. Two of them were not native Italian speaker. During the recruitment they have been asked to read and sign the informed consent. Thus they have filled an entry questionnaire (EnQ) which has been used to collect demographic information like gender, age and level of education.

In the EnQ users have also had to rate on a 5 points Likert Scale, from 1 to 5, their experience with the use of computer (average 3.88 ± 0.85), videogames (average 3.12 ± 1.33), use of immersive virtual displays(average

2.54 ± 1.21), puzzle games (average 3.04 ± 1.2) and online puzzle games (average 1.83 ± 1.05).

8.2.4 *Procedure*

In order to maintain a users flow experience (Csikszentmihalyi, 2014), a game must balance the inherent challenge of the activity and the players ability to address and overcome it. If the challenge is beyond that ability, the activity becomes so overwhelming that it generates anxiety. If the challenge fails to engage the player, he/she quickly loses interest and tends to leave the game (Chen, 2007). In a social game the challenge is heavily influenced by the different abilities of the participants. In the literature (Cairns, Cox, Day, Martin, and Perryman, 2013; Peng and Hsieh, 2012; Schmierbach, Xu, Oeldorf-Hirsch, and Dardis, 2012), when evaluating competitive games, the coupling strategy foresees matching players' skills in order to keep alive the challenge and prevent boredom or frustration. On the other hand there are no indications on how to couple players when evaluating collaborative games in order to maximize the flow experience. Both approaches—matching or not the skills—have pros and cons. Unbalanced abilities can induce frustration in the weaker player and boredom in the stronger one. The same issues can occur also when coupling similarly skilled players: too weak teams can struggle, too strong pair can be not challenged enough. We have therefore decided to use the coupling strategy of matching the abilities.

A pre-experiment aimed at assessing the puzzle-solving abilities has been conducted in order to couple participants according to their dexterity. Before playing they have read the game instructions provided them by the experimenter. The pre-experiment has been played on a non-immersive setup exploiting a 24 inches display, and keyboard and mouse as input devices. The subjects were unaware of the real objective of this session. Hence twelve couples have been formed in order to proceed with the proper experiment.

Before the experiment, each pair has been received together by the experimenters who have informed the subjects about the outline of the session. After the explanation, the players have been divided on the two identical setups prepared for the experiment, located in two different rooms and network connected. The subjects, spatially not co-located, have been able to communicate by using only the enabled communication channels.

Then each user has had time (no more than 5 minutes) to get familiar with the NUI provided. Each subject have played a simplified (12 tiles) single player version of the puzzle game using the interaction metaphor developed. No score has been collected during the trial. Prior to starting each session of the game, each player has been allowed to read again the instructions for the specific session.

The experiment has adopted a within-pairs design. Each couple have undertaken three experimental sessions:

- a collaborative game session played using KM interface (KM_Coll)
- a collaborative game session played using OU interface (OU_Coll)
- a competitive game session played using OU interface (OU_Comp)

During each session the players have had to solve a puzzle. Three different puzzle images have been used during the experiment. The images have been selected out of a list of ten different pictures so that they provide about the same difficulty. The challenge for each image has been evaluated averaging the score results of four same-skilled players—which did not take part to the actual experiment afterwards—playing the jigsaw puzzle game with each one of the ten images. The three pictures obtaining the most similar score have been selected to be used in the experiment.

The different game sessions—KM_Coll, OU_Coll, OU_Comp—as well as the puzzle images have been randomized for each pair in order to exclude potential side effects on the experiment result.

The time limit to finish each puzzle has been 7 minutes.

8.2.5 *Metrics*

Given the questionnaires review (see section 8.1.1), we have decided to adopt GEQ and SPGQ in the current study. Qin et al. (Qin, Patrick Rau, and Salvendy, 2009) focus too much on the narrative aspects of games missing important factors required in our experiment. CPM (Seif El-Nasr, Aghabeigi, Milam, Erfani, Lameman, Maygoli, and Mah, 2010) does not represent a suitable instrument to make a comparison between competitive and collaborative games, as it focuses just on the latter. GameFlow (Sweetser and Wyeth, 2005) represents the baseline from which both IEQ and GEQ started and it is outdated by these two research works. Analysing the IEQ and GEQ questionnaires, we have found that they propose very similar questions even if they namely address different factors, immersion and engagement. Both of them provide an evaluation of player's gaming experience which is equal under many degrees. We have therefore decided to adopt GEQ which is completed by the social questionnaire that is fundamental for our research.

At the end of each game condition, all players have answered a post condition questionnaire (PCQ) composed of a subset of the GEQ items (competence, flow, tension/annoyance, challenge, negative affect and positive affect), a subset of the SPGQ items (empathy and behavioural involvement), awareness and satisfaction questions. After all the conditions have been played, an exit questionnaire (ExQ) has been presented to each player in order to collect their preferences and motivations, friendship relation between the two participants, general impressions and suggestions. Finally an informal debriefing session between the experimenters and both players has been conducted to further register impressions and anecdotes.

Besides data collected through questionnaires and interviews, objective measurements have been recorded through the game in both the preliminary and experimental sessions. Usage and performances data collected comprehend:

- completion time and score to evaluate performances,
- frame-rate and network latency to check if users experience has been affected by malfunctions,
- outcome and tiles positions to identify adopted game strategies,
- player's head movements to analyse the usage of the shared virtual space.

Each session has been video and audio recorded for further investigations. Finally, experimenters have assisted all the sessions taking notes of noteworthy events.

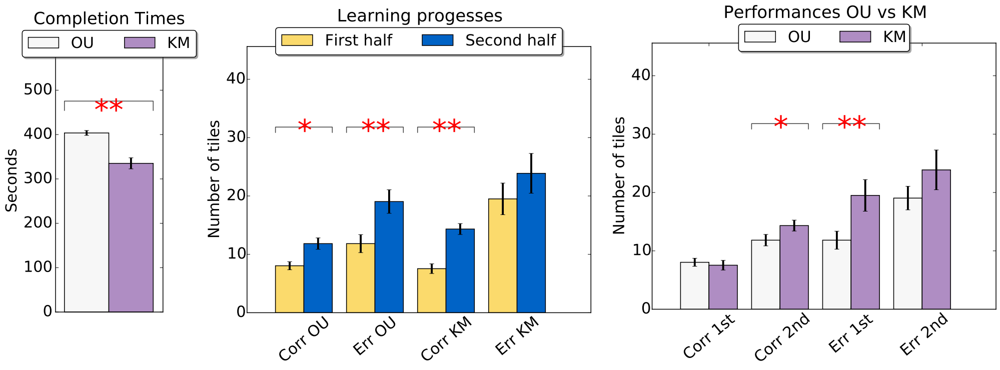
8.3 RESULTS

The questionnaires results have been analysed in order to compare first the differences between the two different technological setups—KM versus OU—and then the influences of the OU configuration on two game mechanics—collaborative versus competitive—both in terms of player engagement, social presence, awareness and performances.

Following the instructions provided by the authors in (Brockmyer, Fox, Curtiss, McBroom, Burkhart, and Pidruzny, 2009), the answers to the GEQ and the SPGQ have been aggregated in order to obtain a value for each one of the eight items: Competence, Flow, Tension/Annoyance, Challenge, Negative affect, Positive affect , Empathy and Behavioural Involvement. A Wilcoxon signed-rank test has been used to statistically compare the results as the distribution of the data has been not Gaussian. The Mean and Standard Error of the Mean (SEM) are reported for the aggregated questionnaire answers and performances measurements. Median, 25th and 75th percentile are reported for the single questionnaire's items, awareness and satisfaction.

8.3.1 *Technological comparison*

In the following, the questionnaires results are reported highlighting the impact of the different technological setups and interaction metaphors on user engagement, social presence, awareness and performances.



(a) (b) (c)

Figure 8.6: Performances results: bars reports 25th and 75th percentiles. $*p \leq 0.05$, $**p \leq 0.01$. Figure 8.6a shows completion times for collaborative and competitive sessions; figure 8.6b shows improvements registered during the two halves of the games; figure 8.6c shows performances registered during the two goal structures.

PERFORMACES Figure 8.6 shows performances registered during the game sessions reporting some significant differences between the two setups.

Only the 41.66% of the couples completed the puzzle before the time ends using the OU setup while the 75% of the couples won the game using the KM

Table 8.1: Time on task comparison among KM and OU sessions.

Factor	**Setup**	**Mean**	**SEM**	**W**	***p***
Correct	OU	403.626	5.484	21.00	$2.26e^{-04}$
	KM	335.089	12.467		

Table 8.2: Comparison of the player performances during the first and second half of the game. The performances are estimated using the number of tiles correctly and wrongly positioned.

Factor	**Period**	**Mean**	**SEM**	**W**	***p***
Corr OU	First half	8.042	0.688	43.50	0.012
	Second half	11.833	0.980		
Err OU	First half	11.833	1.541	53.00	0.006
	Second half	19.042	2.018		
Corr KM	First half	7.542	0.827	18.00	4.14e-04
	Second half	14.333	0.935		
Err KM	First half	19.500	2.709	105.50	0.203
	Second half	23.875	3.406		

setup. As shown in figure 8.6a and reported in Tab. 8.1, participants solved the puzzle significantly faster in the KM setup ($W = 21.00, p = 2.26e^{-4}$)

Figure 8.6b compares the number of tiles correctly and wrongly positioned in the two game sessions respectively in the first and the second half of the game. As reported in Tab. 8.2, a significant difference in the number of correctly positioned tiles has been found both in KM ($W = 18.00, p = 4.14e^{-4}$) and OU ($W = 43.50, p = 0.012$) sessions: participants have been able to position much many tiles during the second half of the game. A significant difference has been registered also in the number of wrongly positioned tiles during OU sessions ($W = 53.00, p = 0.006$) showing that players have tended to make more mistakes during the second half.

Figure 8.6c compares the system configurations in terms of number of tiles correctly and wrongly positioned during the first and second half of the game. As reported in Tab. 8.3, the improvements in the ability in positioning the tiles during the second half of the game is significantly greater when using the KM setup ($W = 45.00, p = 0.025$) with respect to the OU setup. Table 8.3 reports also a significant difference between KM and OU setups in terms of wrongly placed tiles: participants using KM setups have tended to make more errors than using the OU one ($W = 38.00, p = 0.007$)

Table 8.3: Comparison of the player performances during KM and OU sessions. Performances are estimated using the number of tiles correctly and wrongly positioned during the first and second half of the game.

Factor	**Setup**	**Mean**	**SEM**	**W**	***p***
Corr 1st	OU	8.042	0.688	96.50	0.505
	KM	7.542	0.827		
Corr 2nd	OU	11.833	0.980	45.00	0.025
	KM	14.333	0.935		
Err 1st	OU	11.833	1.541	38.00	0.007
	KM	19.500	2.709		
Err 2nd	OU	19.042	2.018	101.00	0.260
	KM	23.875	3.406		

during the first half of the game.

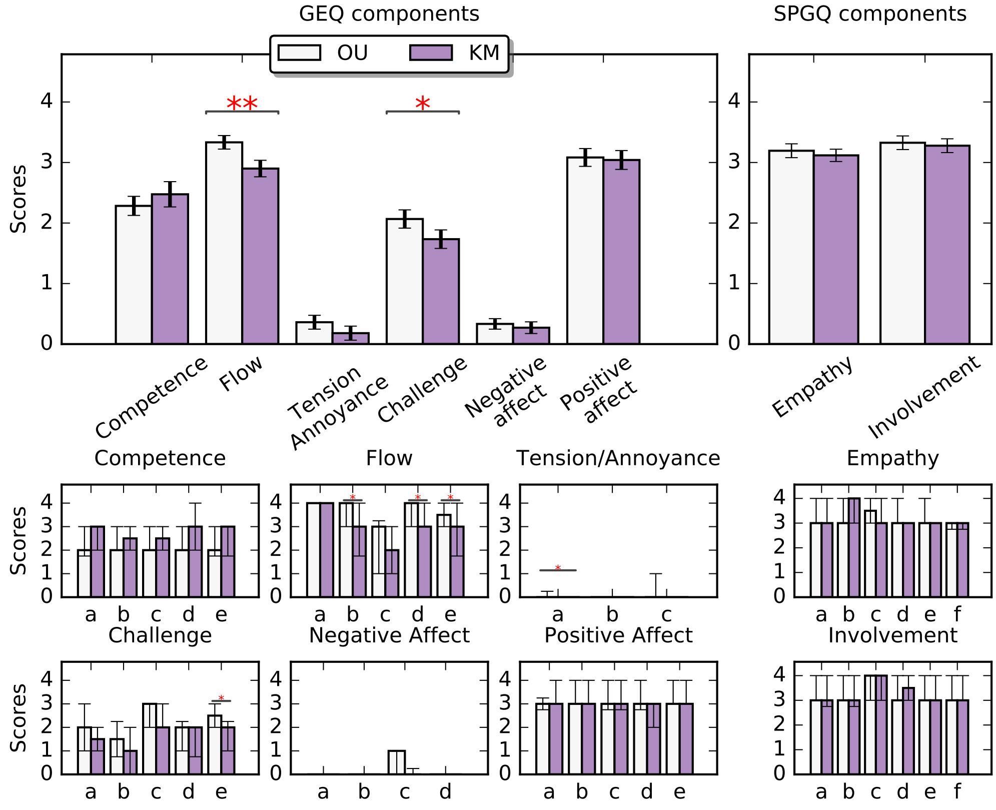


Figure 8.7: GGEQ and SPGQ results: Game Engagement Questionnaire and Social Presence in Gaming Questionnaire results. Bars reports 25th and 75th percentiles. $*p \leq 0.05$, $**p \leq 0.01$.

GAME ENGAGEMENT GEQ questionnaire results indicate an overall positive evaluation of both game setups (see figure 8.7). Participants

Table 8.4: GEQ results for the two game sessions, KM and OU

Factor	**Setup**	**Mean**	**SEM**	**W**	***p***
Competence	OU	2.283	0.158	108.50	0.369
	KM	2.475	0.209		
Flow	OU	3.333	0.112	35.00	0.009
	KM	2.900	0.136		
Tension Annoyance	OU	0.361	0.114	12.50	0.124
	KM	0.181	0.116		
Challenge	OU	2.067	0.151	22.50	0.011
	KM	1.733	0.153		
Negative affect	OU	0.333	0.087	48.50	0.508
	KM	0.271	0.097		
Positive affect	OU	3.083	0.147	105.50	0.728
	KM	3.042	0.156		

have felt competent in both sessions without any relevant difference. An high level of flow have been reached by players with both setups, but the psychological absorption has been significantly greater during OU sessions ($W = 35.00, p = 0.009$) as reported in Tab. 8.4. The differences characterizing the OU experience in terms of flow (see figure 8.7) are highlighted by the questions: (b) "*I forgot everything around me*" ($W = 4.00, p = 0.048$), (d) "*I was deeply concentrated in the game*" ($W = 5.00, p = 0.013$) and (e) "*I lost connection with the outside world*" ($W = 6.00, p = 0.046$).

Challenge has been medium-rated by participants. Players have found the OU session significantly more challenging with respect to KM ($W = 22.5, p = 0.011$). Low values for tension/annoyance and negative affects have been reported for both configurations. The question (a) "*I felt annoyed*" ($W = 0.0, p = 0.023$), have registered a slightly higher value when playing with the Oculus. Players have reported a high positive affects in both sessions.

SOCIAL PRESENCE Participants have high-rated both social components (see figure 8.7), Empathy and Behavioural Involvement. No relevant differences has been found in the results (see Tab. 8.5).

AWARENSESS AND SATISFACTION As reported in figure 8.8, players have had a good awareness of the other's actions, locations and intentions

Table 8.5: SPGQ results for the two game sessions, KM and OU

Factor	**Setup**	**Mean**	**SEM**	**W**	***p***
Empathy	OU	3.194	0.114	103.00	0.664
	KM	3.118	0.102		
Involvement	OU	3.326	0.113	104.50	0.985
	KM	3.278	0.114		

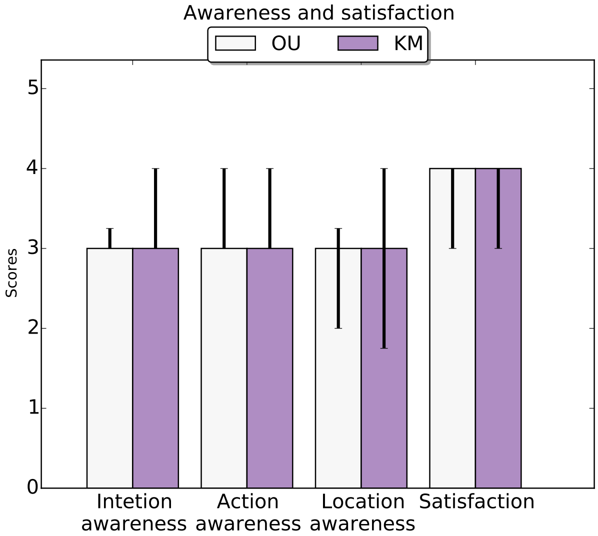


Figure 8.8: Awareness and satisfaction results: bars reports 25th and 75th percentiles.

in both setups.

All the participants have rated both experiences as very satisfying as shown by the question "*Please rate your overall satisfaction*" reporting an high score in both sessions (see Tab. 8.6). Answers to the ExQ have showed a clear preference of the participants for the OU session. To the question *"Which kind of user interface do you prefer?"*, 16 players (∼ 66.7%) answered the natural one.

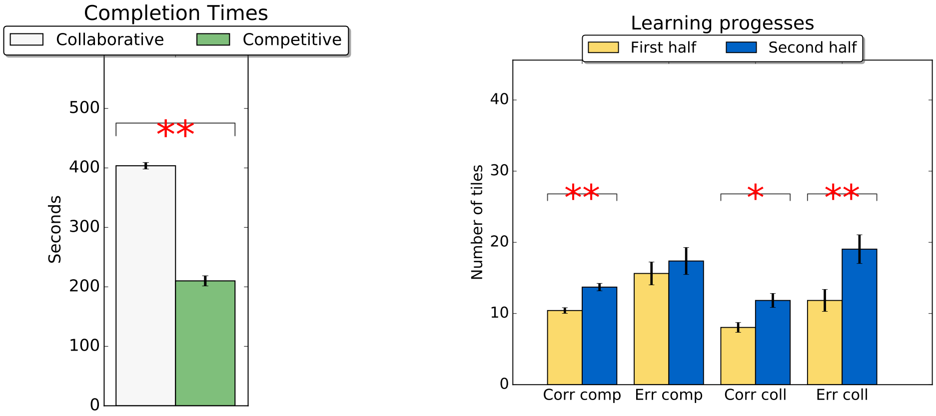
8.3.2 *Game mechanics comparison*

In the following, the questionnaires results comparing collaborative and competitive mechanics in terms of player engagement, social presence, awareness and performances are reported when using the OU configuration.

PERFORMANCES As expected from other studies (Peng and Hsieh, 2012; Plass, O'Keefe, Homer, Case, Hayward, Stein, and Perlin, 2013; Siu, Zook,

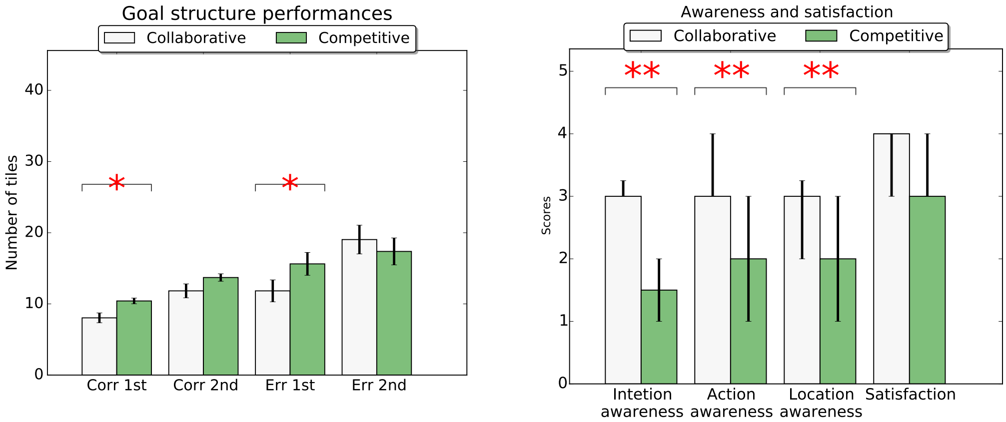
Table 8.6: Awareness and satisfaction results for the two game sessions, KM and OU.

Factor	**Setup**	**p25**	**p50**	**p75**	**W**	***p***
Intetion	OU	3.00	3.00	3.25	32.00	0.926
	KM	3.00	3.00	4.00		
Action	OU	3.00	3.00	4.00	17.00	0.887
	KM	3.00	3.00	4.00		
Location	OU	2.00	3.00	3.25	57.00	0.562
	KM	1.75	3.00	4.00		
Satisfaction	OU	3.00	4.00	4.00	33.50	0.644
	KM	3.00	4.00	4.00		



(a) Completion times for collaborative and competitive sessions

(b) Improvement registered during the two halves of the games



(c) Comparison between goal structures in terms of performances

(d) Participants awareness and satisfaction results

Figure 8.9: Awareness, satisfaction and performances results: bars reports 25th and 75th percentiles. $^{*}p \leq 0.05$, $^{**}p \leq 0.01$

and Riedl, 2014), players have performed much better when competing (see

Table 8.7: Time on task comparison among collaboratve and competitive sessions.

Factor	Setup	Mean	SEM	W	*p*
Correct	Coll.	403.626	5.484	0.00	1.80e-05
	Comp.	210.045	8.561		

Table 8.8: Comparison of the player performances during KM and OU sessions. The performances are estimated using the number of tiles correctly and wrongly positioned during the first and second half of the game.

Factor	Setup	Mean	SEM	W	*p*
Corr 1st	Coll.	8.042	0.688	53.50	0.017
	Comp.	10.417	0.399		
Corr 2nd	Coll.	11.833	0.980	77.00	0.062
	Comp.	13.708	0.525		
Err 1st	Coll.	11.833	1.541	73.00	0.048
	Comp.	15.625	1.605		
Err 2nd	Coll.	19.042	2.018	119.50	0.383
	Comp.	17.375	1.894		

figure 8.9) also when using the immersive/natural setup.

Puzzles have always been completed when competing, while only the 41.7% of the teams have succeeded when collaborating. Figure 8.9a shows a significant difference ($W = 0, p = 1.80e^{-5}$) among the completion times between the two sessions: users required about half of the time to complete the task when competing than when collaborating (see Tab. 8.7).

Figure 8.9b shows the comparison of the number of tiles correctly and wrongly positioned in the two game sessions respectively in the first and the second half of the game. The chart reports a statistically significant difference ($W = 13.0, p = 5.72e^{-4}$) in the number of correctly positioned tiles of competitive sessions. As regard as the collaborative goal structure, a highly significant difference has been registered both in the correctly ($W = 43.5, p = 0.012$) and wrongly ($W = 53.0, p = 0.006$) positioned tiles (see Tab. 8.9).

Figure 8.9c compares the two goal structures in terms of number of tiles correctly and wrongly positioned during the first and second half of the game. The chart reports a statistically significant difference in the number tiles positioned during the first half of the sessions (see Tab. 8.8).

Table 8.9: Comparison of the player performances during the first and second half of the collaborative and competitive games. The performances are estimated in terms of number of tiles correctly and wrongly positioned.

Factor	**Period**	**Mean**	**SEM**	**W**	***p***
Corr comp	First half	10.417	0.399	13.00	5.72e-04
	Second half	13.708	0.525		
Err comp	First half	15.625	1.605	86.50	0.312
	Second half	17.375	1.894		
Corr coll	First half	8.042	0.688	43.50	0.012
	Second half	11.833	0.980		
Err coll	First half	11.833	1.541	53.00	0.006
	Second half	19.042	2.018		

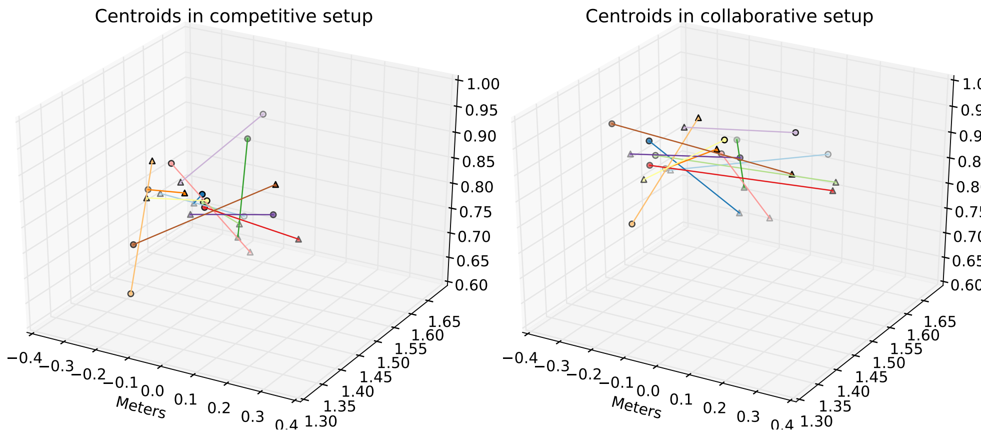
Table 8.10: Awareness and satisfaction results for collaborative and competitive games.

Factor	**Game**	**p25**	**p50**	**p75**	**W**	***p***
Intetion	Coll.	3.00	3.00	3.25	21.00	4.83e-04
	Comp.	1.00	1.50	2.00		
Action	Coll.	3.00	3.00	4.00	8.00	6.20e-04
	Comp.	1.00	2.00	3.00		
Location	Coll.	2.00	3.00	3.25	4.00	0.003
	Comp.	1.00	2.00	3.00		
Satisfaction	Coll.	3.00	4.00	4.00	22.00	0.527
	Comp.	3.00	3.00	4.00		

AWARENESS As shown in figure 8.9d, players have had a good awareness of the other's actions, locations and intentions.

Table 8.10 reports statistically significant differences between competitive and collaborative game sessions for all the awareness components: intention awareness ($W = 21.00, p = 4.83e^{-4}$), action awareness ($W = 8.00, p = 6.20e^{-4}$) and also location awareness ($W = 4.00, p = 0.003$). Players have been less aware of the other when competing.

SPACE USAGE Users have tended to move more($W(23) = 0.0, p = 1.82e^{-5}$) during collaborative sessions ($58.96 \pm 12.25m$) than during competition ($37.19 \pm 7.78m$). On the other hand, players have moved faster ($W(23) = 35.0, p = 0.001$) when competing ($0.19 \pm 0.027m/s$) than when collaborating ($0.16 \pm 0.035m/s$). Figure 8.10 shows the centroids of the collected positions for each user grouped by pairs. The comparison of the


Centroids in competitive setup

Centroids in collaborative setup

(a) Distribution during competition (b) Distribution during collaboration

Figure 8.10: Distribution of users positions' centroids grouped by couples.

Table 8.11: SPGQ results for collaborative and competitive game sessions

Factor	**Game**	**Mean**	**SEM**	**W**	***p***
Empathy	Coll.	3.194	0.114	0.00	1.81e-05
	Comp.	1.910	0.172		
Involvement	Coll.	3.326	0.113	0.00	3.95e-05
	Comp.	1.549	0.167		

centroids' distances shows a significant difference ($W(11) = 11.0, p = 0.028$). Distances have been greater in collaborative sessions ($0.330 \pm 0.14m$) than in competitive sessions ($0.23 \pm 0.08m$).

SOCIAL PRESENCE Both the social components—empathy ($W = 0.0$, $p = 1.81e^{-5}$) and behavioural involvement ($W = 0.0$, $p = 3.95e^{-5}$)—measured with the SPGQ are much more significant in the collaborative scenario than in the competitive one (see Tab. 8.11).

As shown in figure 8.11 all the single components of both items report greater outcomes when collaborating ($p \leq 0.01$).

GAME ENGAGEMENT The results of the GEQ indicate an overall positive evaluation of both games (see figure 8.11).

A high level of flow has been obtained in both games as shown in figure 8.11. While the difference between the aggregated flow items doesn't result to be statistically significant, the question d, "*I was deeply concentrated in the game*", reported a significant difference ($W = 0.0, p = 0.0143$).

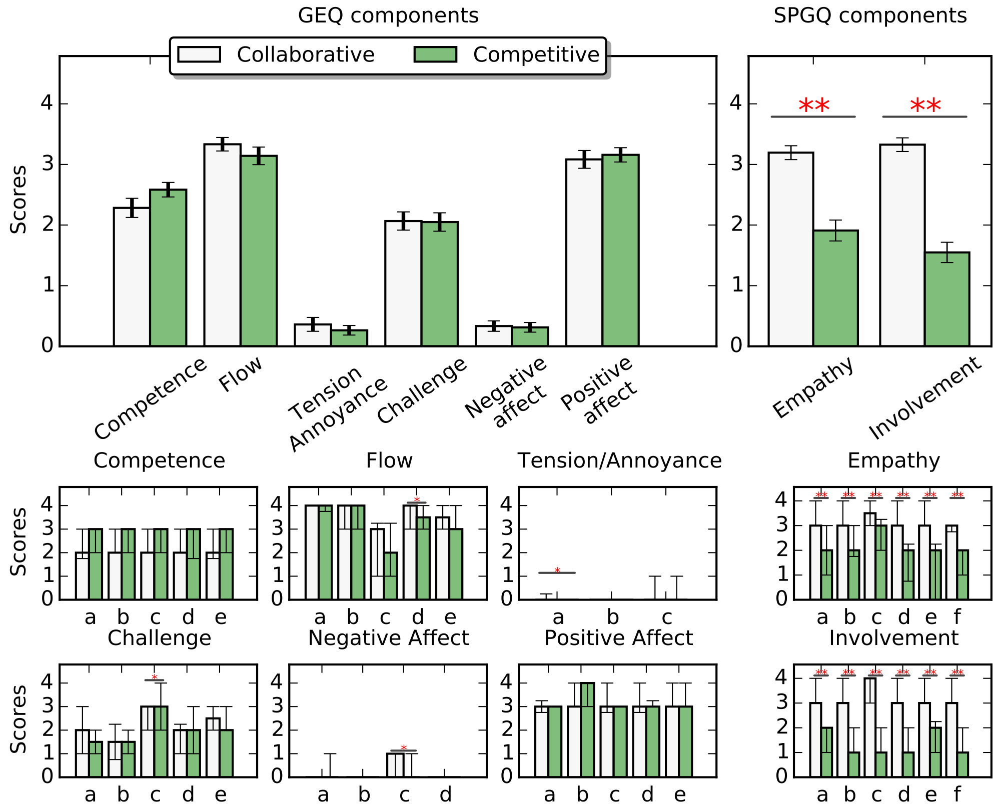


Figure 8.11: GEQ and SPGQ results: Game Engagement Questionnaire and Social Presence in Gaming Questionnaire results. Bars reports 25th and 75th percentiles. $*p \leq 0.05$, $**p \leq 0.01$.

Participants have been more concentrated while collaborating.

Challenge has been medium-rated by participants with no significant difference between the two games. Nonetheless single question d, "*I felt challenged*", shows a greater value when competing ($W = 9.0, p = 0.0455$).

Low values for tension/annoyance and negative affects have been recorded for both games. A slightly higher value of annoyance ($W = 0.0, p = 0.025$, "*I felt annoyed*", question a) and negative affects ($W = 4.5, p = 0.034$, "*I found it tiresome*", question c) have been registered when collaborating.

Players have reported a high positive affects in both games.

PREFERENCES AND SATISFACTION To the question *"Which kind of game do you prefer?"* of the ExQ, 15 players ($\sim 62.5\%$) have answered the competitive one. In the open questions of the ExQ and during the final debriefing, experimenters have asked to the players to explain their choice. Most of the players have reported that the competitive game has been easier and straightforward when compared to the collaborative one. Many

Table 8.12: GEQ results for collaborative and competitive game sessions

Factor	**Game**	**Mean**	**SEM**	**W**	***p***
Competence	Coll.	2.283	0.158	78.00	0.067
	Comp.	2.583	0.120		
Flow	Coll.	3.333	0.112	44.50	0.073
	Comp.	3.142	0.145		
Tension Annoyance	Coll.	0.361	0.114	32.50	0.359
	Comp.	0.264	0.079		
Challenge	Coll.	2.067	0.151	92.00	0.627
	Comp.	2.050	0.152		
Negative affect	Coll.	0.333	0.087	49.50	0.847
	Comp.	0.312	0.080		
Positive affect	Coll.	3.083	0.147	86.00	0.476
	Comp.	3.158	0.118		

participants have asserted also to thoroughly like the challenge provided by competition. The greater commitment required to play collaboratively has been reported by almost all the participants. Nonetheless the majority of the players asserted that they have been positively impressed by the collaborative game and they would like to play it again. Few subjects have found the dependency to the other during collaboration frustrating.

The question "*Please rate your overall satisfaction*" has reported a high score in both games (see Fig 8.9d). As shown in Tab. 8.10, a slightly higher satisfaction has been conferred to the collaborative game $(3,458 \pm 0,658)$ with respect to the competitive one $(3,375 \pm 0,647)$.

8.4 DISCUSSION

In this research we have investigated the impact of emerging immersive visualization technologies combined with NUI on player engagement, awareness and social presence when playing social games. We also assessed how different social game mechanics exploit or benefit of the new interaction metaphor.

OCULUS & NUI VS. KEYBOARD & MOUSE Almost all the players have enjoyed the OU metaphor and most of them (66.7%) have found it preferable to the classic Keyboard & Mouse interface even if it has resulted being challenging. The playing experience with the Oculus has been perceived as more engaging and entertaining. Almost all the participants who have preferred the KM metaphor have appreciated the lower complexity of the interface, which results more familiar and comfortable for people who daily use computers. They have been able to become proficient in a shorter period, while the OU configuration have required more time to get used to. Most participants (71%), during the final debriefing, have appreciated the natural interaction because it makes the experience more similar to the reality (*"I felt like I was really playing with him a real puzzle game!"*).

During KM sessions, while mouse indication has been the preferred interaction method, verbal communication has been nonetheless extremely important to support it. Players described the visual features of the tiles to grab (e.g. *"Take the big red bird!"*) and sometimes indicated the destination placeholder numbering its position (e.g. *"Put it on the third column, four cells up."*).

Participants have extensively used all the available communication channels provided in order to complete the task when using the OU setup. Experimenters have observed that most of the players have preferred using hands gesture to interact with the other (see figure 8.3): "*Wow, I can point out them!*". Verbal communication has been nonetheless extremely important to support gestures. Also in this case many participants have described the tiles to grab or the action to undertake verbally in addition to using their hands.

All the subjects have reported that the NUI has been more challenging with respect to the KM metaphor, however only one player has found the former metaphor too complex to be used and not enjoyable. Keyboard & Mouse have been considered to be more immediate and faster by majority of players (75%). Nonetheless the experimenters have observed that even during the KM sessions many players have tried to use body language and gestures to interact with the other—like pointing at the monitor with the hands—even if these communication channels were not available. The same attempt has also been reported in (Sajjadi, Cebolledo Gutierrez, Trullemans, and De Troyer, 2014). Hence providing a natural interaction seems to be important during social activities. Adopting a transparent NUI, by removing any artificial medium between the user and the social sphere, would probably lead to an easier and more enjoyable communication with respect to a mediated one (e.g. Keyboard & Mouse).

The game environment has been designed in order to maximize the space needed by a player during a game session and stimulate participant's movements in order to evaluate the spatial awareness in a social shared environment. In the OU setup, due to the nature of the technology and to the choices made, players have been able to see only part of the scene at once and therefore they have been forced to walk and move the head in order to play. On the contrary, when playing using the KM setup, player have not been subject to physical restraints. They have been able to move backward enough to have a global view of the entire scene and still being able to interact with it. In this way they have not been forced to continuously move, saving precious time. This aspect may have influenced the interfaces comparison in terms of usability penalizing the OU interface. Adopting a different game design forcing the players to move and rotate the view in the KM as in the immersive setup, would probably generate different results. It would be interesting to evaluate this different condition.

Thanks to the high immersion and embodiment induced by the technology, participants during OU session have perceived the proxy of the other more as a physical presence rather than a virtual representation. This has made the experience more engaging. At the same time, the absence of any physical feedback and the possibility to pass through the representation of the partner has been perceived by some players odd and sometimes a bit annoying for the purpose of the game, while cheerful by others. The essential KM interface has resulted to be more functional to the task but, as observed by the experimenter and highlighted in the open questions, less funny and more impersonal.

Differently from what presumed, there have been no significant differences between the two configurations in terms of awareness of the partner. Even if the OU metaphor provides more detailed and richer information on what the other user is doing, players have been able to equally distinguish other's intentions and actions in both setups. This can be mainly justified by the heavy usage of verbal communication to coordinate the team actions. Partner location awareness has obtained similar results. Players focusing on the puzzle completion do not require rich and accurate information about partner location and consider equally satisfying the two modalities.

Almost all the participants have been Italian native speakers with two exceptions. The experimenters have noticed that players speaking different languages have benefited more from the NUI. Language misunderstandings have been compensated by gestures. Due to the small number of non Italian native speaker it has not been possible to evaluate the real impact on the user engagement. It would be interesting to further investigate the impact of linguistic differences on player engagement.

The sample used in the experiment has been composed of people who daily use computers. All the participants have been at ease with the use of Keyboard & Mouse. It would be interesting to evaluate the impact of the two metaphors on a more variegated sample to highlight possible differences related to different personal skills.

The study results show a significant increment in the participants flow experience in the OU configuration. The deep immersion provided by the technological solution together with the natural interaction have led to a greater absorption in the game. During the debriefing session, a player talking about the OU experience said: "*The interruption has a much stronger impact; the break is much clearer*". Sweetser and Wyeth (2005) speculate on the effect of the social and flow components in games. They assert that reaching the flow mental state is impeded by the social activity which establish a link between the player and the real world. In the OU setup using the NUI, being the social interaction fused into the virtual environment, players do not need any more to "leave" the game in order to interact. The link with the real world requested by the social communication is therefore broken. In order to investigate this effect, it would be interesting to develop a single-player version of the game using the same immersive setup with the NUI. If the provided communication channels would be expressive and transparent enough to make the social interaction fused in the virtual world, the comparison between the single and multi-player game should not highlight any relevant difference in terms of flow experience. It would be even possible to observe a significant flow increment due to the greater engagement provided by the social component.

COLLABORATIVE VS. COMPETITIVE During competitive sessions users have mostly focused on accomplishing task as fast as possible and have been inclined to overlook, and sometimes purposely ignore, the other. The interaction between the opponents has almost entirely consisted in short occasional duels for grasping or positioning a tile. Verbal interaction has been limited at utterance and laughs happening usually during or after disputes. "*Ahh!! You are also here just to annoy me!*". On the contrary, in the collaborative scenario participants have extensively used all the available communication channels provided in order to complete the task. Experimenters have observed that, among the available communication modalities, most of the players have preferred using gesture to collaborate: "*Wow, I can point out them!*".

Nonetheless verbal communication has been important for team coordination and to support body language especially during the first part of the game while players have to get used to the immersive visualization and the NUI. Many participants have described the tiles to grab or the action to undertake verbally in addition to using their hands.

Awareness of the partner has been considerably higher in the collaborative scenario: the rich cooperation between players has led to greater consciousness of the other. When collaborating players need to focus on the partner in order to coordinate their actions. On the contrary when competing the rush induced by the competition mostly limits the social interaction; the goal commitment has outweighed partner awareness. The same difference has been observed also in the attention players have paid to avoid the proxy of the partner ("*Sometimes it was like being in the body of the other*"). When the users' proxies have collided in the collaborative scenario, the experimenters have observed more ailment. Players have been aware of the partner and have tried to avoid bumps. When competing instead collisions between the proxy generates much less complaints; users sometimes have intentionally crossed the "body" of the partner to grab a tile or reach a position.

The competitive goal structure has resulted to be more challenging (see figure 8.11, question c of Challenge item). The higher challenge can be explained by the recurrent proxies occlusions happening during the game. Players in fact have not cared about avoiding collisions in order to take advantage of a better position or conquer a tile. The hypothesis is supported by the statements registered during the debriefing session. Most of the player have indeed found the proxies collision annoying. In order to address this issue, it would be interesting to evaluate virtual stimuli to be integrated in the interface able to alert the user about the proximity of the partner or opponent.

As reported in the results (see Fig 8.9d), a slightly higher satisfaction has been registered during collaborative sessions. At the same time in the ExQ questionnaire more participants ($\sim$ 62.5%) have asserted to prefer the competitive goal structure. The two results can be likely explained by personal attitudes. During the debriefing session, in fact, many participants have expressed their overall preference for competition in general. Players have found the mechanics of the competitive game more straightforward and easier to understand. At the same time most of them have seen great potential in collaborating in a shared VE through natural interaction. Some of them have also suggested that this technology could also improve team work skills. Many participants preferring competition have expressed the wish to play more the game in order to fully appreciate the game mechanics they have found extremely enjoying.

Players need more time to become “productive” in the collaborative scenario, as also reported in (Plass, O’Keefe, Homer, Case, Hayward, Stein, and Perlin, 2013).

As expected, when competing players have performed considerably better. The results are aligned with previous researches (Peng and Hsieh, 2012; Plass, O’Keefe, Homer, Case, Hayward, Stein, and Perlin, 2013; Siu, Zook, and Riedl, 2014). When collaborating players have taken time to find the best way for them to communicate and plan a strategy. Differences in performances can be attributed to the overhead in terms of coordination. At the same time, users have resulted to be more deeply concentrated when collaborating than when operating autonomously. The intense interaction and the higher complexity required to coordinate the partners actions have led to a deeper concentration and therefore slightly greater value of flow (see Fig 8.11, question d Flow item).

Higher values of annoyance and negative affects have been registered during collaboration (see figure 8.11, question a of Tension/Annoyance item and question c of Negative affect item). The main reason that has led to these results has been related to the higher complexity required to coordinate the players. Participants during the debriefing session have indeed reported that the strong interdependence between the partners have been sometimes perceived as annoying.

As expected, the social components have been much more relevant during collaboration than when competing. Unlike the competitive game where each user is autonomous, in the collaborative one players have relied on each other in order to solve the puzzle. Players have indeed felt more connected when collaborating. Results suggests that empathy have been fostered more when working together on a common goal, than when competing against an opponent. As expected, users behaviors are reciprocally more influenced when collaborating than when competing. These results are reported in Tab. 8.11.

A different usage of the space has been found in the results. When playing collaborative games, players have tended to occupy distinct regions of the virtual space. Experimenters have noticed that once reached a stable configuration allowing each player to reach a set of tiles, the positions have been maintained over the time. Each subject tended to move the unreachable tiles closer to the partner allowing him to maintain its position. On the contrary during competitive sessions each player has tended to occupy a central position from where all the tiles have been reachable. The distance among the players has been significantly lower during these games. Furthermore the participants have moved significantly faster in order to beat the opponent and grasp more tiles.

Almost all the participants were Italian native speakers with few exceptions. During the game sessions experimenters have noticed that collaborative mechanic has resulted to be more demanding and sometimes frustrating for pairs speaking different languages. After repeated attempts at verbally coordinating their action, the availability of gesture have resulted to be fundamental to perform the task. Players have been able to manage team coordination by means of body language. Due to the small number of non Italian native speaker it has not been possible to further investigate the impact on the user engagement of the linguistic factor. It would be interesting to investigate the impact of linguistic differences on collaborative applications.

As highlighted in (Tauer and Harackiewicz, 2004), competition and collaboration can be combined to create more complex social scenario in which team of person cooperate against others: intergroup competition. The authors examined the effects of pure cooperation, pure competition, and intergroup competition on performance and intrinsic motivation in a sport setting. This kind of social setting can be already found in different video games, from Capture the Flag(CTF) to MMORPGs. It would be interesting to verify if the same difference evaluated in (Tauer and Harackiewicz, 2004) between pure cooperative/competitive and inter-group competition can be observed also in player engagement in video games.

9

CONCLUSIONS

The presented work investigated the challenges faced while developing fully immersive virtual environment systems. The main technological challenges provided by IVEs can be grouped in four areas: present a visual feedback, estimating the user's head/body pose, minimizing the motion-to-photon and action-reaction latencies and finally providing effective interaction mechanisms. The two most adopted IVEs solutions are spatially immersive displays and head-mounted displays. AR typically exploit the latter technology.

Four immersive VEs systems were developed to the purpose of this research and we presented the challenges that we faced during the development and the solutions we adopted. A big CAVE system exploiting a 18 projectors, a 16 m^2 floor and 3 walls was developed. The rendering of complex VEs is a very computational demanding task. In CAVE system each wall require a dedicated stereoscopic rendering of the scene from a different perspective, hence the task is ever more demanding. Our solution to this problem consisted in a cluster rendering architecture which takes advantage of a network of calculators to achieve computational performances that are beyond the capability of any single workstation. Our architecture exploit a master-slaves configuration: the graphical application runs on the master node intercepting the graphical commands that are distributed among the cluster of slave nodes. The system turned out to be an efficient and flexible solution, which allows to control multiple output devices and workstations—even with different specifications—and allowing to efficiently distribute the workload among the cluster. The final system performances are strictly related to the characteristics of each application. The architecture is optimized for the use with immersive visualization systems allowing us to build the above described complex CAVE system **(RA1)**. IVEs systems typically take advantage of the tracking of the user's movements to implement direct short-range navigation metaphors. However due to the typically limited range of the tracking systems the long-range navigation exploits the use of input devices. Many common navigation tools require at least one hand to be operated thus limiting the possible bi-manual interaction with the environment. In order to be able to navigate large VE we developed a hands-free interaction device. The device is a special carpet provided with an array of pressure sensitive cells allowing to analyse the body pressure distribution and the recognition of foot-based gestures **(RA1)**.

A second fully IVR system exploiting a head-mounted display was developed and subsequently used as virtual training platform. We didn't propose any novel solution, however we adopted state-of-the-art devices and developed cutting edge algorithms and software providing guidelines to anyone interested in doing the same. The system combines a commodity HMD, with custom head-tracking and body-tracking technologies as well as multi-player capabilities. One ore more trainers and a trainees are able to share the same VE—even if not physically co-located—to perform training sessions on realistically simulated environments **(RA1)**.

The third system developed is an AR see-through system with real-world occlusion capabilities. Nowadays, most commodity AR headset exploit the use of optical see-through display to augment the real world. This kind of displays have obvious advantages over video see-through systems in terms of real-world perception. A direct perception of the real world cannot be achieved by any display in terms of resolution, colours fidelity and latency. However see-through displays are not perfect. Currently, one of the major drawbacks is the capability lack of selectively block-out the real environment; as results of this lack the augmented overlay is affected by the real world lighting. Furthermore, differently from what is possible using video see-through AR headsets, it is not possible to replace or mask real objects **(RA1)**. Our approach to this problem, even if not perfect or applicable in all conditions, consists in replacing standard light sources with projectors in order to project occlusion masks for the augmented objects over the real environment. This approach allows both to improve the display fidelity as well as to reshape the real world. We developed an open framework, based on Unity, to provide occlusion capabilities to commodity AR headsets; the Microsoft Hololens was used in our case **(RA1)**. This platform allowed us to investigate the importance of this capability. Our studies reported that, when the real environment is not too simple, real-world occlusion is important to enhance the user perception of virtual objects, furthermore potential depth conflicts between the two worlds—real and virtual—can be properly solved. On the other side, if the real environment is very simple the merging of the virtual objects with the real world results to be effective enough without requiring the occlusion capability. In a simple environment, while performing tasks where a very accurate estimation of the environments is fundamental, providing some sort of virtual geometrical aids turned out to be very effective. Our target scenario was in fact a surgical procedure an accurate position estimation of some interest points inside the patient's body is extremely important to the surgeon. Providing a virtual guidance with no occlusion aid proved to improve the surgeon's performances **(RQ1, RQ2)**.

Lastly a fully immersive mixed reality system exploiting the use of an HMD and a 3D camera allowing the user to see his own hand embedded in the VE was developed. The user is able to naturally interact with the virtual objects. Showing the photo-realistic capture of user's hands in a coherently rendered virtual scenario induces in the user a strong feeling of embodiment without the need of a virtual avatar as a proxy. Furthermore, manipulating virtual objects with own real hands and navigating in the virtual space using own body not only provides an intuitive user interaction experience, but also improves the spatial understanding and self perception inside the VE **(RA1)**. For this reasons we believe that the system could be an effective learning and training platform.

Virtual Reality is an extremely promising solution for industrial training purposes, as it allows to perform simulated hands-on activities in a controlled and safe environment. A VR-based training would allow to challenge operators with dynamic cases in order to train them to respond quickly in unusual situations which cannot be easily or safely simulated in real life, thus reducing costs and risks associated to these activities. We conducted experiments using fully IVE systems as learning platforms in industrial context. Workers were instructed on maintenance and assembly/disassembly procedures as well as on safety procedures showing them the possible negative outcomes of unsafe operations. When the task assigned to the worker involves the assembly of a machinery, VR training resulted to be more effective in terms of knowledge transfer than practicing on a real machinery. The user better remember the correct sequence and accomplish the task with an higher success rate **(RQ4)**. The VR training resulted to be as effective as the traditional training approach when learning theoretical notions (like the safety rules they must respect) The lack of significant differences between the two training approaches suggests that methods are interchangeable **(RQ6, RQ5)**. Thanks to the high level of immersion provided by the technological solutions adopted—both VR and MR systems—an high level of presence was reported. The high levels of presence is an important driver of user engagement. Indeed, the VR experience resulted in a greater involvement compared to traditional classroom-based training, meaning that the VR course have resulted to be much more engaging and, consequently, positively impacting on motivation, attention and training efficacy **(RQ3)**. Looking at the workers behaviour while performing the assigned procedures, the VR training resulted to be significantly much more effective than the traditional approach **(RQ6)**. The users performed better both in terms of adherence to the safety rules and in terms of correctly performing the assigned tasks **(RQ5)**. VR training could indeed raise the level of reception and memorization of physical procedures, but further investigation to specifically investigate this effects are needed.

Collaboration is fundamental in our lives and permeates most of the activities we perform everyday. Nowadays the hardware evolution has lead to powerful solutions able to substantially improve the interaction of players with the VE. We are slightly moving from an abstract and mediated interaction to a more physical and embodied interaction. This big revolution in the way we interact with and inside VEs offers great possible improvements for the social interaction. In this work we have investigated the effects that IVEs and natural interaction technologies could produce on our social life and social behaviour, and vice versa, how our social habits influence the way we use a new technology and what are the user's expectations. We conducted the investigation using collaborative virtual environments in the entertainment field. Given the relevance that video-games have in our society and their deep correlation with technology, entertaining is an important test-case to study the impact of the emerging technologies. Furthermore, multi-player video-games fostering social interactions are perfect to study the impact on our social behaviour. In this context, a fully immersive mixed reality multi-player puzzle game exploiting natural interaction was developed. The impact of the new immersive technologies (HMD and natural user interface) have been assessed in comparison with traditional gaming setups (monitor, keyboard and mouse).

A high level of social presence was reported in game. However the immersion provided by the system and the natural interaction didn't influenced the social presence level that resulted to be the same on both configurations. Thanks to the high immersion and embodiment induced by the new technologies, participants have perceived the proxy of the other more as a physical presence rather than a virtual representation. This has made the experience more engaging. The most noteworthy result is a significantly greater flow experience when playing the immersive game. In contrast with the common hardware setups in which the communication between the players is not natural and mediated by some artefacts, the immersive solution has allowed the users to experience ways of communication similar to what they use in real-life **(RQ7)**. We also found that the strong physical presence in the VE and the ability to naturally interact with the partner reversed the wrong correlation between flow and social presence. Previous works have supposed a negative influence of social interaction on the flow experience. We believe that this effects was induced by the technologies adopted to communicate. Encouraging sociability, but forcing players to communicate by typing on a keyboard or using devices "external" to the VE, the player is drawn out of the game environment, back into the real world, breaking the flow state. In this case the social interaction among players represents a link between players and the real world. We believe that making the interfaces totally transparent to the user, like in the case of the immersive setup, this correlation could be totally broken **(RQ8)**.

We also analysed also how different social mechanics—collaboration and competition—exploit the new opportunities provided by the new interfaces. The research proved that collaboration is particularly indicated to foster awareness of the partner, behavioural involvement and empathy creating stronger connections between players, while competition usually improves the performances of the single player. Slightly higher values of concentration have been reached in collaborative games suggesting that this mechanic could facilitate the attainment of the flow state. If on one hand the proposed immersive system proved to improve the social interaction in games, on the other hand new issues arise. Although the proxy representation of the player have greatly contributed to improve the awareness and social connection between participants, the virtual presence of the players in a shared VE and the absence of physical feedback lead to frequent virtual collisions between users. When collaborating the greater awareness of the partner allowed the players to reduce the collisions, and the sporadic bumps generated more ailment. On the contrary the rush induced by the competition causes the player to forget about the other. The absence of physical feedback or of any sort of virtual warnings about the presence of the opponent have penalized more the competitive game, where more frequent collisions have increased the challenge. New expedients have to be developed in order to reduce bumps or make them less annoying and to further improve the awareness of the partner **(RQ8, RQ9)**. With this comparisons we have contributed to deepen our knowledge about the relation between social interaction and technology exploring potential synergies in different contexts.

According to the results obtained during this research we believe that the presented technologies have immense potential to improve our lives both as individuals as well as as social entities. Immersive VEs have also demonstrated to be effective training tools leading to better knowledge transfer as well as guaranteeing pleasant and engaging learning experience.

9.1 FUTURE WORK

In this work we tackled just a tiny part of the immensely vast field of virtual environments, nevertheless I hope it could result useful to someone getting an overall idea of what are the problems faced when dealing with immersive virtual environments.

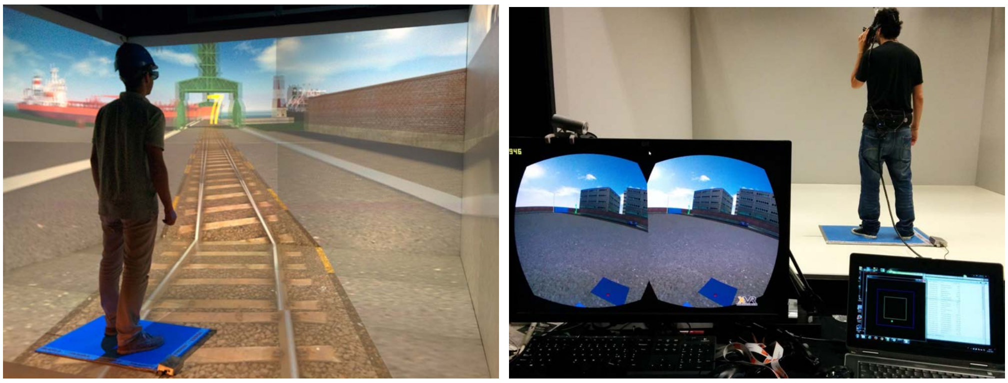
We found extremely funny and rewarding to work on collaborative applications and we believe that the next major step in VEs will involve the social aspect. For this reason, we aim in the future to work on pushing the boundaries of social interaction in virtual environments. In particular we'd like to investigate techniques to improve the reciprocal users' awareness, and which are the methodologies which can be adopted to improve the collaboration between distant people allowing them to share the same virtual spaces.

We found evidence of the effectiveness of using VR systems to train operators in performing different tasks. In this field we'd like to conduct long-term studies in collaboration with some companies in order to investigate the actual impact in a real work environment, and to find methodologies which could further improve operators' performances. We believe that also in this scenarios the collaboration plays a major role, so in the future we aim at investigating how VR training systems could benefit of shared experiences.

Appendix

A

THE FOOT CONTROLLER



(a) (b)

Figure A.1: The Foot Controller in the CAVE (a) and wearing an HMD (b).

The proposed interaction device is a pressure sensitive carpet exploiting resistive technology to achieve variable pressure sensing. It is built combining flexible and protective materials with actual pressure sensing circuitry sandwiched between them (see figure A.2).

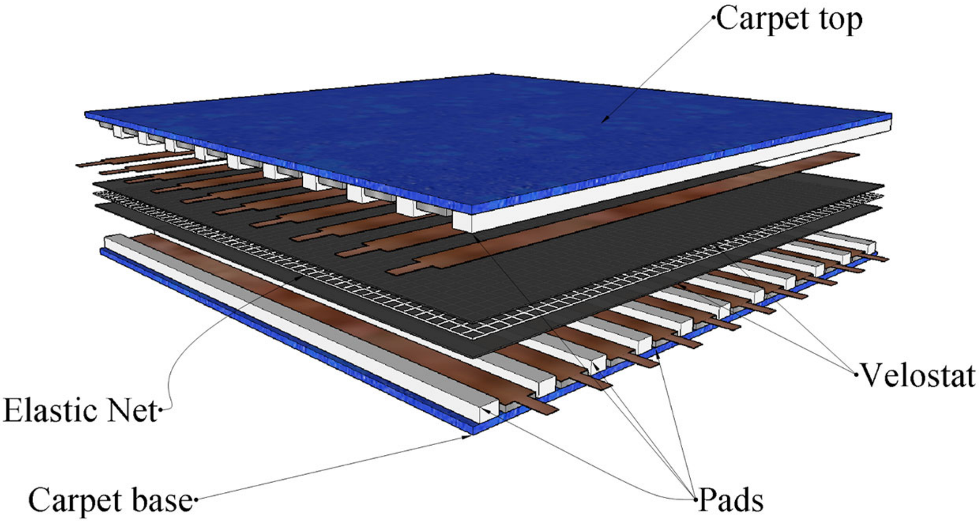


Figure A.2: Composition of the Foot Controller.

The protective materials has been selected to be strong and durable enough to carry the weight of a person, yet flexible enough to transmit the applied pressure to the internal sensors and also recover quickly the original shape when the body weight is removed. Foam or resin based materials have the desired properties; in fact the best materials identified in our tests were thick yoga mats which are very durable and recover their structure quickly after pressure is removed. A first internal layer consists of parallel and equally spaced copper lines which are glued to the bases with bottom layer lines being orthogonal to the top layer's ones. The innermost conductive layer consists of two Velostat™sheets which allow to measure the amount of pressure applied, point wise. Velostat™[1] is a variable resistance material which changes its conductivity as a function of the applied pressure. In order to further reduce leakage current resulting from large surface of the touching sheets, they are separated by an elastic net which physically decouples the conductive material unless certain pressure is applied. As a result of this structure, the Foot Controller is a pressure-sensitive carpet suitable for any type of footwear (including no-shoes) except for high-heels shoes. Several functioning prototypes were developed, sized from $45x60cm$ ($15x22$ sensels, Figure 3) up to $160x160cm$ ($48x48$ sensels). Pressure data is gathered by means of a dedicated microcontroller from the Arduino family that continuously scans the sensels, and encodes them into custom RLE-based encoding to save transfer time. The controller connects to the PC via USB-to-Serial adapter.

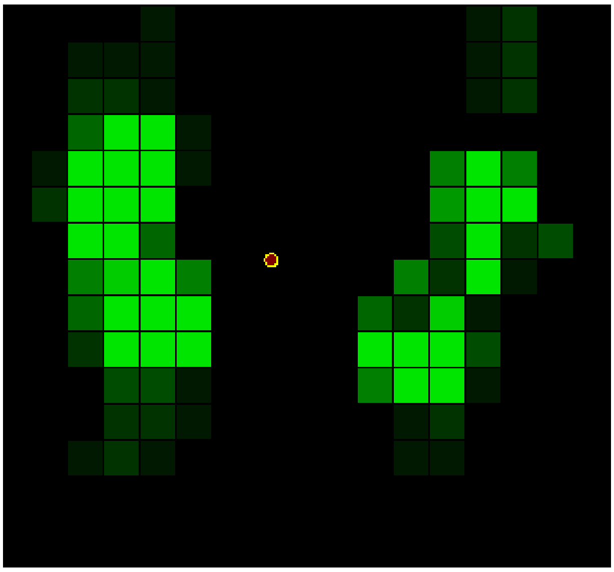


Figure A.3: Example of pressure distribution on sensels while a user is standing still on the mat: lighter areas receives higher pressure.

---

[1] 3-M Company, Maplewood, Minnesota, U.S.

Scanned sensels produce images that can be processed in order to retrieve gesture information, following an algorithm articulated into the following steps: first of all there is a noise reduction stage; then the image is thresholded and segmented into BLOBs. Depending on their relative position, size, shape and movement from previous frames, each is attributed to a foot with a certain level of confidence, or marked as noise. Finally, the algorithm uses the sequence of processed images in order to appropriately detect gestures. Our initial strategy for exploiting the pressure sensing capabilities of the Foot Controller for foot gesture recognition is mapping per-foot center of mass to conventional 2D gestures. While the latter has well-established algorithms and approaches for reliable recognition, the former is a novel strategy. This resulted to be simple and robust to implement, as some of the user actions do not require any special machine-learning training and can be immediately recognized and in particular:

- the user steps on the mat;
- the user leaves the mat;
- the user performs left and right (fore) foot taps;
- the user performs left and right backtaps;

These information can be either used directly as commands or they can be used as an alphabet and recognize more complex gestures as a grammar. Other gestures can rely on more advanced information, such as the distribution of pressure. As an example, the Foot Controller can be used in the "Human Joystick" mode. In this mode a Segway-like navigation can be implemented. When enabled, the mode is automatically turned on once the distance between the feet is more than foot length ("feet shoulder width apart"). Leaning on one of the sides of the mat produces values, similar to those of an analog joystick that can be used, for instance, as an input to a navigation metaphor. Combining this mode with the regular gesture recognition mode, however, is generally difficult because users often perform many unconscious movements when trying to make a tap while not in balance and vice versa, leading to potential confusion in recognition.

In order to perform a preliminary evaluation of the Foot Controller as a navigation interface, we have setup a basic interaction metaphor. Based on the Human Joystick mode, a similar metaphor has been already previously used by Haan, Griffith, and Post (2008) and other works using the Wii Balance Board. In particular, up/down values are used to produce forward/backward movements and left/right values (see figure A.4) are used to rotate clockwise/counter-clockwise about the vertical axis. Based on previous work of Hilsendeger, Brandauer, Tolksdorf, and Fröhlich, 2009 and on a battery of preliminary tests we decided to use the displacement of the user's center of mass in order to modulate acceleration rather than directly speed, as this allows for smoother functioning even in case of noisy data. The presence of many sensors allows to automatically detect the user orientation and make the stationary position transparent to it, whilst, for instance, simpler devices like the Wii Balance Board bound that position to one of the four main orientations. Furthermore the use of a large controller enables mixing natural navigation via direct walking with the use of metaphors where the user assumes a chosen stationary pose.

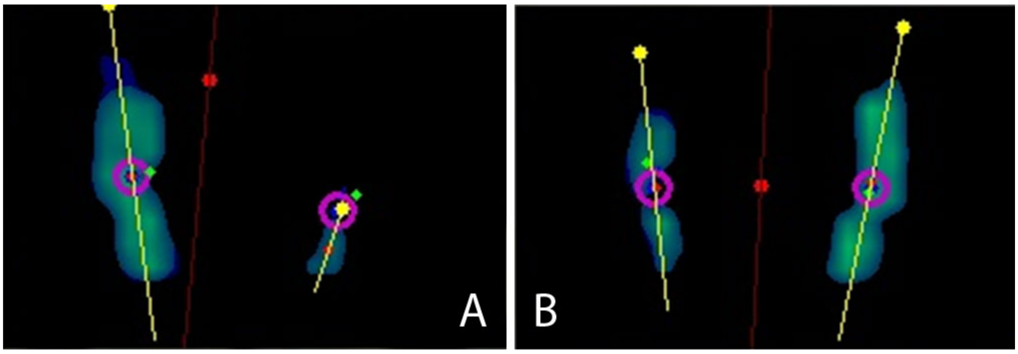


Figure A.4: Pressure distribution while leaning on the left foot (a) and on the right foot (b).

The task is a typical navigation task where users have to explore the environments by freely finding a path across landmarks identified by numbers increasing from 0 to 7. Landmarks are placed so that each landmark is visible from the previous landmark, in order to ensure that time-on-task is spent mostly on navigation rather than on bearing. However, some landmarks are placed in order to force ample rotations of the forward direction, thus soliciting all the gestures detected by the mat. The pilot test has been conducted on a group of 8 users, whose ages range from 25 to 44 (average 31.5). All of them are familiar with the use of VR devices. Users have been divided in two subgroups: the first has accomplished the task in a CAVE-like environment, the second using the Oculus Rift HMD. Both visualization systems run by means of the XVR technology (Tecchia, Carrozzino, Bacinelli, Rossi, Vercelli, Marino, Gasparello, and Bergamasco, 2010) which allows, with minimal configuration changes, to easily switch from one device to the other. Both systems make use of an optical system for head tracking. While in the CAVE the user can see the real mat, this is obviously not possible using the HMD. In order to establish similar experimental conditions, in the case of HMD-based visualization a virtual mat is drawn, spatially corresponding to the real one, so that users are able to know if they are correctly positioned on the mat without the need of removing the HMD.

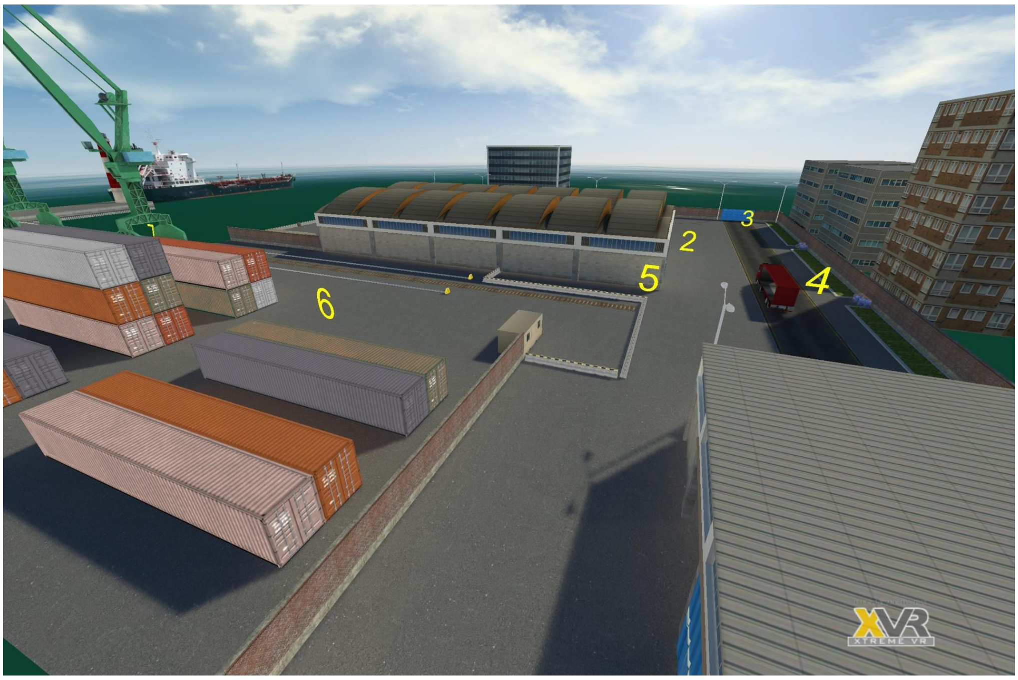


Figure A.5: Virtual Scenario used in the Pilot Study.

The test is divided in two phases. First each user has to practice with the navigation metaphor on a simple scene for 5 minutes; then he is asked to complete the task. We have measured the time-on-task both for the complete path and between each pair of consecutive landmarks. After the task, users were asked to answer a questionnaire consisting in 6 questions aimed at evaluating the effort needed to master the use of the Foot Controller and its usability. The total length of the virtual path is $280m$, which is the minimum distance each user had to travel. The allowed maximum translation speed is $13km/h$, while the rotational one is limited to $0.47rad/s$; these values have been selected based on the results of the preliminary tests.

RESULTS All users have been able to perform the whole task. The average completion time is 269 seconds (SD = 142), and the average speed is $4.66km/h$ (SD = 2.26). The achieved average speed over an allowed maximum of $13km/h$ can be considered high enough to say that the Foot Controller and the chosen metaphor are effective enough to allow exploring comfortably the VE. This can be considered a baseline to be compared with future improvements on both the device and other navigation metaphors. After the experiment, users were asked to fill a short questionnaire made up of six questions, rating from 0 to 6 on a 7-point Likert scale, similar to the one presented by Hilsendeger, Brandauer, Tolksdorf, and Fröhlich, 2009:

1. It was easy to learn to navigate.
2. I did not need to think after a while.
3. I was able to move wherever and however I wanted.
4. I was able to stop wherever and however I wanted.
5. I did not have a feeling of limited freedom of movement.
6. I felt tired after navigation.

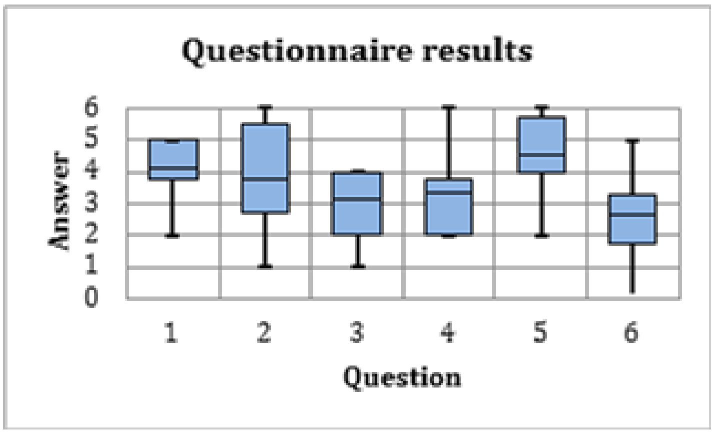


Figure A.6: Questionnaire results.

The questionnaire results show that users found easy to learn how to navigate and they felt free to move in space much wider than the real one where the task was performed. Some users reported they still had to think about how to move. Performing the task resulted to be physically tiring for some users. Additionally, a basic elementary test focusing on the learning curve in the use of the mat has been performed. A user was left free to use the mat for half an hour, after that the experiment task has been repeated, resulting in a completion time of 110 seconds instead of 230, and an increase in the average speed of $9.15 km/h$. Results suggest that using the Foot Controller can be very effective in navigating VEs. Nevertheless the use of different metaphors should be investigated in order to improve the learning curve and to setup more comfortable conditions for users.

B

SOLIDAR FRAMEWORK

The open framework we developed, *SolidAR*, exploits the use of stereoscopic projectors do add real-world occlusion capabilities to optical see-through AR headsets (like the Microsoft HoloLens). The framework is freely distributed in the form of a Unity package, allowing to develop a wide range of real-time AR applications and scenarios to anyone interested in conducting studies in which the mutual occlusion matters. The proposed system exploits only commodity hardware that can easily be bought. The choices of using only commodity hardware and a widespread graphical framework have been taken in order to allow a large number of developers to experiences a convincing AR experience with a minimal effort.

It includes a series of prefabs, materials, shaders and scripts to handle tracking, stereoscopy, network streaming, speech input, calibration and synchronization allowing to easily integrate fine lighting-control inside any Unity AR application. An overall working scheme is depicted in figure B.1.

Two types of prefab cameras are available to the user:

- *HoloCamera*: represents the HoloLens stereo camera entity. It is responsible of streaming the head pose and the HoloLens's perspective matrices and render onto the depth buffers needed for the occlusion shadows' computation. Only the contents tagged as *RemoteScene* are visible to this camera.

- *ProjectorCamera*: represents a projector's stereo camera entity. It is in charge of computing and displaying the stereo occlusion masks. The projector's pose is estimated using the calibration procedure. This camera renders only the contents tagged as *LocalScene*.

The tagging mechanism allows to select the nature of each object: real world objects must be tagged as LocalScene, while virtual ones must be tagged as RemoteScene. In this way it is possible to compute occlusion shadows according to the pseudo-algorithm described in section 6.1. A post-process shader applied to each ProjectorCamera is in charge of actually performing the occlusion masks calculation.

The steps needed for a developer to integrate the lighting-control into his AR applications are the following:

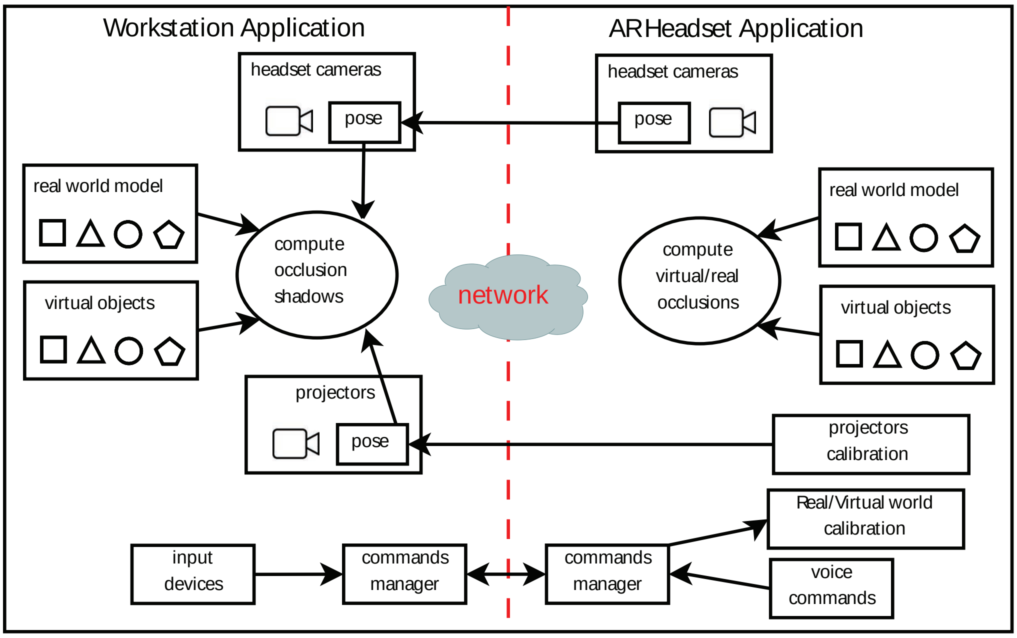


Figure B.1: An overall scheme of framework.

1. Adding a HoloCamera to the project.
2. Adding one or more ProjectorCameras prefabs to the project according to the system's actual configuration and assign to it the intrinsic projector's calibration.
3. Configure the *HoloSender* script attached to the HoloCamera with the parameters of the network. This component performs the streaming of the HoloLens tracking data.
4. Configure the *HoloReceiver* script attached to the each ProjectorCamera with the parameters of the network. This component receiving the HoloLens tracking data allows the occlusion mask calculation.
5. Assign the virtual objects to to the RemoteScene layer.
6. Assign all the real objects to the LocalScene layer.
7. Deploy the application to the HoloLens.
8. Deploy the application to the each workstation connected to one or more projectors.

The framework provides further scripts and prefabs to help in the development of AR applications; a full updated list can be found in the framework's documentation. *SolidAR* is released under Creative Commons.

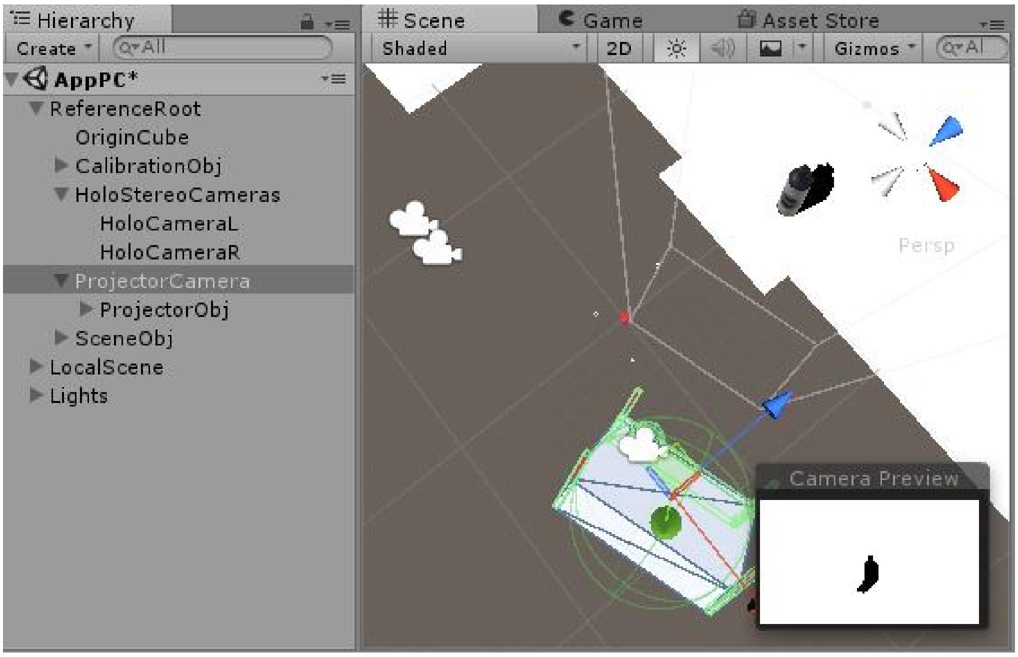


Figure B.2: A screenshot of a simple Unity scene using *SolidAR* for Unity. In white the geometry of the real environment onto which the virtual object's occlusion mask is projected.

Even if the system, due to the necessary instrumentation of the environment, is not practical enough to be used in any condition, there are some specific scenarios – like the needle biopsy scenario that we presented – where the benefits obtained by the improved capabilities overcame the high system complexity disadvantage. Although the presented framework is ready to be used, some components could be improved. Support for additional external tracking systems can be added, as well as a better handling of multiple overlapping projectors and multiple workstations can be implemented.

B.1 PERFORMANCES

The framework is targeted at developing real-time AR applications. However, actual system performances depend on several factors like the complexity of the scenarios, the computational load demanded to the HoloLens and to the workstations and the projectors' and network's latencies.

Our setup consists of a workstation equipped with two Xeon E5-2630 CPUs, 64GB of ram and a AMD FirePro W8000 GPU. The FirePro graphic card is provided with a DIN output needed to drive the shutter glasses. An Optoma GT750 stereo projector and a Optoma 3D-RF shutter glasses are used. The projector's measured latency is $43ms$ which is slightly high for AR applications that involve fast head movements; additional network latency and jitter affecting the data streaming need to be taken into account. However, considering that these latencies affect the occlusion mask but not the AR overlay we found them acceptable for our experimental setup. Latencies could in some cases be reduced by using hardware with better specifications and tracking prediction algorithms. The setup we used to perform the experiments exploits a single projector and a single workstation, however the framework already support multiple projectors and workstations. All the experiments were conducted with a video refresh rate of $60Hz$ (HoloLens maximum supported framerate).

C

LIST OF PUBLICATIONS

- Giovanni Avveduto, Franco Tecchia, and Henry Fuchs (2017). “Evaluating the Importance of Real-world Occlusion in Optical See-through AR Displays”. In: *23th ACM Symposium on Virtual Reality Software and Technology*. **Paper submitted**. ACM

- Giovanni Avveduto, Camilla Tanca, Cristian Lorenzini, Franco Tecchia, Marcello Carrozzino, and Massimo Bergamasco (2017). “Safety Training Using Virtual Reality: A Comparative Approach”. In: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, pp. 148–163

- Giovanni Avveduto, Franco Tecchia, Marcello Carrozzino, and Massimo Bergamasco (2016). “A Scalable Cluster-Rendering Architecture for Immersive Virtual Environments”. In: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer International Publishing, pp. 102–119

- Raffaello Brondi, Leila Alem, Giovanni Avveduto, Claudia Faita, Marcello Carrozzino, Franco Tecchia, and Massimo Bergamasco (2015). “Evaluating the impact of highly immersive technologies and natural interaction on player engagement and flow experience in games”. In: *International Conference on Entertainment Computing*. Springer International Publishing, pp. 169–181

- Raffaello Brondi, Giovanni Avveduto, Leila Alem, Claudia Faita, Marcello Carrozzino, Franco Tecchia, Yusuf Pisan, and Massimo Bergamasco (2015). “Evaluating the effects of competition vs collaboration on user engagement in an immersive game using natural interaction”. In: *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. ACM, pp. 191–191

- Raffaello Brondi, Giovanni Avveduto, Marcello Carrozzino, Franco Tecchia, Leila Alem, and Massimo Bergamasco (2015). “Immersive Technologies and Natural Interaction to Improve Serious Games Engagement”. In: *International Conference on Games and Learning Alliance*. Springer International Publishing, pp. 121–130

- Daniele Sportillo, Giovanni Avveduto, Franco Tecchia, and Marcello Carrozzino (2015). “Training in VR: A Preliminary Study on Learning Assembly/Disassembly Sequences”. In: *International Conference on Augmented and Virtual Reality.* Springer, pp. 332–343

- Marcello Carrozzino, Giovanni Avveduto, Franco Tecchia, Pavel Gurevich, and Benjamin Cohen (2014). “Navigating immersive virtual environments through a foot controller”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology.* ACM, pp. 23–26

- Franco Tecchia, Giovanni Avveduto, Marcello Carrozzino, Raffaelo Brondi, Massimo Bergamasco, and Leila Alem (2014). “Interacting with your own hands in a fully immersive MR system”. In: *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on.* IEEE, pp. 313–314

- Franco Tecchia, Giovanni Avveduto, Raffaello Brondi, Marcello Carrozzino, Massimo Bergamasco, and Leila Alem (2014). “I’m in VR!: using your own hands in a fully immersive MR system”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology.* ACM, pp. 73–76

BIBLIOGRAPHY

Abe, Norihiro, Jiang Y Zheng, Kazuaki Tanaka, and Hirokazu Taki (1996). “A training system using virtual machines for teaching assembling/disassembling operation to novices”. In: *Systems, Man, and Cybernetics, 1996., IEEE International Conference on*. Vol. 3. IEEE, pp. 2096–2101.

Alliger, George M and Elizabeth A Janak (1989). “Kirkpatrick’s levels of training criteria: Thirty years later”. In: *Personnel psychology* 42.2, pp. 331–342.

Avveduto, Giovanni, Camilla Tanca, Cristian Lorenzini, Franco Tecchia, Marcello Carrozzino, and Massimo Bergamasco (2017). “Safety Training Using Virtual Reality: A Comparative Approach”. In: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, pp. 148–163.

Avveduto, Giovanni, Franco Tecchia, Marcello Carrozzino, and Massimo Bergamasco (2016). “A Scalable Cluster-Rendering Architecture for Immersive Virtual Environments”. In: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer International Publishing, pp. 102–119.

Avveduto, Giovanni, Franco Tecchia, and Henry Fuchs (2017). “Evaluating the Importance of Real-world Occlusion in Optical See-through AR Displays”. In: *23th ACM Symposium on Virtual Reality Software and Technology*. **Paper submitted**. ACM.

Azuma, Ronald, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre (2001). “Recent advances in augmented reality”. In: *IEEE computer graphics and applications* 21.6, pp. 34–47.

Azuma, Ronald T (1997). “A survey of augmented reality”. In: *Presence: Teleoperators and virtual environments* 6.4, pp. 355–385.

Bajura, Michael, Henry Fuchs, and Ryutarou Ohbuchi (1992). “Merging virtual objects with the real world: Seeing ultrasound imagery within the patient”. In: *ACM SIGGRAPH Computer Graphics*. Vol. 26. 2. ACM, pp. 203–210.

Bajura, Michael and Ulrich Neumann (1995). “Dynamic registration correction in video-based augmented reality systems”. In: *IEEE Computer Graphics and Applications* 15.5, pp. 52–60.

Barsalou, Lawrence W (2008). “Grounded cognition”. In: *Annu. Rev. Psychol.* 59, pp. 617–645.

Bergamasco, Massimo, S Perotti, Carlo Alberto Avizzano, Marcello Angerilli, Marcello Carrozzino, and Emanuele Ruffaldi (2005). “Fork-lift truck simulator for training in industrial environment”. In: *Emerging*

Technologies and Factory Automation, 2005. ETFA 2005. 10th IEEE Conference on. Vol. 1. IEEE, 5–pp.

Billinghurst, Mark, Adrian Clark, Gun Lee, et al. (2015). “A survey of augmented reality”. In: *Foundations and Trends® in Human–Computer Interaction* 8.2-3, pp. 73–272.

Bimber, Oliver and Bemd Frohlich (2002). “Occlusion shadows: using projected light to generate realistic occlusion effects for view-dependent optical see-through displays”. In: *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on.* IEEE, pp. 186–319.

Bimber, Oliver and Ramesh Raskar (2005). *Spatial augmented reality: merging real and virtual worlds.* CRC press.

Biocca, Frank, Chad Harms, and Judee Burgoon (2003). “Toward a more robust theory and measure of social presence: Review and suggested criteria”. In: *Presence* 12.5, pp. 456–480.

Biocca, Frank and Mark R Levy (2013). *Communication in the age of virtual reality.* Routledge.

Blake, Jonathan and Hakan B Gurocak (2009). “Haptic glove with MR brakes for virtual reality”. In: *IEEE/ASME Transactions On Mechatronics* 14.5, pp. 606–615.

Bowman, Doug, Chadwick Wingrave, Joshua Campbell, and Vinh Ly (2001). “Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments”. In:

Bradley, L, G Walker, and A McGrath (1996). *Shared Spaces' British Telecommunications Engineering.*

Brockmyer, Jeanne H, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny (2009). “The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing”. In: *Journal of Experimental Social Psychology* 45.4, pp. 624–634.

Brondi, Raffaello, Leila Alem, Giovanni Avveduto, Claudia Faita, Marcello Carrozzino, Franco Tecchia, and Massimo Bergamasco (2015). “Evaluating the impact of highly immersive technologies and natural interaction on player engagement and flow experience in games”. In: *International Conference on Entertainment Computing.* Springer International Publishing, pp. 169–181.

Brondi, Raffaello, Giovanni Avveduto, Leila Alem, Claudia Faita, Marcello Carrozzino, Franco Tecchia, Yusuf Pisan, and Massimo Bergamasco (2015). “Evaluating the effects of competition vs collaboration on user engagement in an immersive game using natural interaction”. In: *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology.* ACM, pp. 191–191.

Brondi, Raffaello, Giovanni Avveduto, Marcello Carrozzino, Franco Tecchia, Leila Alem, and Massimo Bergamasco (2015). “Immersive Technologies and Natural Interaction to Improve Serious Games Engagement”. In:

International Conference on Games and Learning Alliance. Springer International Publishing, pp. 121–130.

Brooks Jr, Frederick, U Marcus Brown, Tuscaloosa Chris Burbeck, Nat Durlach, MIT Steve Ellis, James Lackner, Warren Robinett, Mandayam Srinivasan, MIT Ivan Sutherland, Dick Urban, et al. (1992). "Research directions in virtual environments". In: *Computer Graphics* 26.3, p. 153.

Brooks Jr, Frederick P (1996). "The computer scientist as toolsmith II". In: *Communications of the ACM* 39.3, pp. 61–68.

Bruder, Gerd, Frank Steinicke, Kai Rothaus, and Klaus Hinrichs (2009). "Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras". In: *CyberWorlds, 2009. CW'09. International Conference on*. IEEE, pp. 43–50.

Bryson, Alex (2016). "Health and safety risks in Britain's workplaces: where are they and who controls them?" In: *Industrial Relations Journal* 47.5-6, pp. 547–566.

Bryson, Steve (1996). "Virtual reality in scientific visualization". In: *Communications of the ACM* 39.5, pp. 62–71.

Buchmann, Volkert, Stephen Violich, Mark Billinghurst, and Andy Cockburn (2004). "FingARtips: gesture based direct manipulation in Augmented Reality". In: *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. ACM, pp. 212–221.

Burguillo, Juan C (2010). "Using game theory and competition-based learning to stimulate student motivation and performance". In: *Computers & Education* 55.2, pp. 566–575.

Cairns, Paul, Anna L Cox, Matthew Day, Hayley Martin, and Thomas Perryman (2013). "Who but not where: The effect of social play on immersion in digital games". In: *International Journal of Human-Computer Studies* 71.11, pp. 1069–1077.

Caplan, Scott, Dmitri Williams, and Nick Yee (2009). "Problematic Internet use and psychosocial well-being among MMO players". In: *Computers in Human Behavior* 25.6, pp. 1312–1319.

Cardoso, Alexandre, Paulo R Prado, Gerson FM Lima, and Edgard Lamounier (2017). "A Virtual Reality Based Approach to Improve Human Performance and to Minimize Safety Risks When Operating Power Electric Systems". In: *Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries*. Springer, pp. 171–182.

Carrozzino, Marcello, Giovanni Avveduto, Franco Tecchia, Pavel Gurevich, and Benjamin Cohen (2014). "Navigating immersive virtual environments through a foot controller". In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. ACM, pp. 23–26.

Carrozzino, Marcello, Chiara Evangelista, R Brondi, C Lorenzini, and Massimo Bergamasco (2012). "Social networks and web-based serious games as novel educational tools". In: *Procedia Computer Science* 15, pp. 303–306.

Caudell, Thomas P and David W Mizell (1992). "Augmented reality: An application of heads-up display technology to manual manufacturing processes". In: *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on.* Vol. 2. IEEE, pp. 659–669.

Chang, George, Patricia Morreale, and Padmavathi Medicherla (2010). "Applications of augmented reality systems in education". In: *Proceedings of society for information technology & teacher education international conference.* Vol. 1. AACE Chesapeake, VA, pp. 1380–1385.

Chao, Kuo-Jen, Hui-Wen Huang, Wei-Chieh Fang, and Nian-Shing Chen (2013). "Embodied play to learn: Exploring Kinect-facilitated memory performance". In: *British Journal of Educational Technology* 44.5, E151–E155.

Chen, Jenova (2007). "Flow in games (and everything else)". In: *Communications of the ACM* 50.4, pp. 31–34.

Cheng, Dewen, Yongtian Wang, Hong Hua, and Jose Sasian (2011). "Design of a wide-angle, lightweight head-mounted display using free-form optics tiling". In: *Optics letters* 36.11, pp. 2098–2100.

Cole, Helena and Mark D Griffiths (2007). "Social interactions in massively multiplayer online role-playing gamers". In: *CyberPsychology & Behavior* 10.4, pp. 575–583.

Colman, Andrew M (2013). *Game theory and its applications: In the social and biological sciences.* Psychology Press.

Corrêa, Wagner T, James T Klosowski, and Cláudio T Silva (2003). "Out-of-core sort-first parallel rendering for cluster-based tiled displays". In: *Parallel Computing* 29.3, pp. 325–338.

Crockett, Thomas W (1997). "An introduction to parallel rendering". In: *Parallel Computing* 23.7, pp. 819–843.

Cruz-Neira, Carolina, Jason Leigh, Michael Papka, Craig Barnes, Steven M Cohen, Sumit Das, Roger Engelmann, Randy Hudson, Trina Roy, Lewis Siegel, et al. (1993). "Scientists in wonderland: A report on visualization applications in the CAVE virtual reality environment". In: *Virtual Reality, 1993. Proceedings., IEEE 1993 Symposium on Research Frontiers in.* IEEE, pp. 59–66.

Cruz-Neira, Carolina, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart (1992). "The CAVE: audio visual experience automatic virtual environment". In: *Communications of the ACM* 35.6, pp. 64–73.

Csikszentmihalyi, Mihaly (2014). *Flow.* Springer.

Darken, Rudolph P, William R Cockayne, and David Carmein (1997). "The omni-directional treadmill: a locomotion device for virtual worlds". In: *Proceedings of the 10th annual ACM symposium on User interface software and technology.* ACM, pp. 213–221.

De Sa, Antonino Gomes and Gabriel Zachmann (1999). “Virtual reality as a tool for verification of assembly and maintenance processes”. In: *Computers & Graphics* 23.3, pp. 389–403.

Dinh, Huong Q, Neff Walker, Larry F Hodges, Chang Song, and Akira Kobayashi (1999). “Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments”. In: *Virtual Reality, 1999. Proceedings., IEEE.* IEEE, pp. 222–228.

Ducheneaut, Nicolas and Robert J Moore (2004). “The social side of gaming: a study of interaction patterns in a massively multiplayer online game”. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work.* ACM, pp. 360–369.

Ducheneaut, Nicolas, Nicholas Yee, Eric Nickell, and Robert J Moore (2006). “Alone together?: exploring the social dynamics of massively multiplayer online games”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems.* ACM, pp. 407–416.

Ellis, Stephen R and Brian M Menges (1998). “Localization of virtual objects in the near visual field”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40.3, pp. 415–431.

Enayati, Nima, Elena De Momi, and Giancarlo Ferrigno (2015). “A quaternion-based unscented kalman filter for robust optical/inertial motion tracking in computer-assisted surgery”. In: *IEEE Transactions on Instrumentation and Measurement* 64.8, pp. 2291–2301.

Feasel, Jeff, Mary C Whitton, and Jeremy D Wendt (2008). “LLCM-WIP: Low-latency, continuous-motion walking-in-place”. In: *3D User Interfaces, 2008. 3DUI 2008. IEEE Symposium on.* IEEE, pp. 97–104.

Fiore, Loren Puchalla and Victoria Interrante (2012). “Towards achieving robust video selfavatars under flexible environment conditions”. In: *International Journal of Virtual Reality* 11.3, pp. 33–41.

Fjeld, Morten and Benedikt M Voegtli (2002). “Augmented chemistry: An interactive educational workbench”. In: *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on.* IEEE, pp. 259–321.

Freitas, Rubina and Pedro Campos (2008). “SMART: a SysteM of Augmented Reality for Teaching 2 nd grade students”. In: *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2.* British Computer Society, pp. 27–30.

Frostling-Henningsson, Maria (2009). “First-person shooter games as a way of connecting to people:“Brothers in blood””. In: *CyberPsychology & Behavior* 12.5, pp. 557–562.

Fuchs, Henry (1990). “Systems for display of three-dimensional medical image data”. In: *3D imaging in medicine.* Springer, pp. 315–331.

Fuchs, Henry, Mark Livingston, Ramesh Raskar, D’nardo Colucci, Kurtis Keller, Andrei State, Jessica Crawford, Paul Rademacher, Samuel Drake, and Anthony Meyer (1998). “Augmented reality visualization for laparoscopic surgery”. In: *Medical Image Computing and Computer-Assisted Interventation—Miccai’98*, pp. 934–943.

Gajadhar, Brian J, Yvonne AW De Kort, and Wijnand A Ijsselsteijn (2008). “Shared fun is doubled fun: player enjoyment as a function of social setting”. In: *Fun and games*. Springer, pp. 106–117.

Gao, Chunyu, Yuxiang Lin, and Hong Hua (2012). “Occlusion capable optical see-through head-mounted display using freeform optics”. In: *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*. IEEE, pp. 281–282.

Gavish, Nirit, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia (2015). “Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks”. In: *Interactive Learning Environments* 23.6, pp. 778–798.

Gibson, James Jerome (1966). “The senses considered as perceptual systems.” In:

Gonzalez-Franco, Mar, Rodrigo Pizarro, Julio Cermeron, Katie Li, Jacob Thorn, Windo Hutabarat, Ashutosh Tiwari, and Pablo Bermell-Garcia (2017). “immersive Mixed reality for Manufacturing Training”. In: *Frontiers* 4.3, p. 1.

Haan, Gerwin de, Eric J Griffith, and Frits H Post (2008). “Using the Wii Balance Board™ as a low-cost VR interaction device”. In: *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. ACM, pp. 289–290.

Hamdi, Mustapha, Antoine Ferreira, Gaurav Sharma, and Constantinos Mavroidis (2008). “Prototyping bio-nanorobots using molecular dynamics simulation and virtual reality”. In: *Microelectronics Journal* 39.2, pp. 190–201.

He, Changyu, H Tutkun Şen, Sungmin Kim, Praneeth Sadda, and Peter Kazanzides (2014). “Fusion of inertial sensing to compensate for partial occlusions in optical tracking systems”. In: *Workshop on Augmented Environments for Computer-Assisted Interventions*. Springer, pp. 60–69.

Heeter, Carrie (1992). “Being there: The subjective experience of presence”. In: *Presence: Teleoperators & Virtual Environments* 1.2, pp. 262–271.

Henderson, Steven J and Steven Feiner (2009). “Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret”. In: *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, pp. 135–144.

Hilsendeger, Andre, Stephan Brandauer, Julia Tolksdorf, and Christian Fröhlich (2009). “Navigation in virtual reality with the wii balance board”. In: *6th Workshop on Virtual and Augmented Reality*.

Hoffman, Hunter G, Ari Hollander, Konrad Schroder, Scott Rousseau, and Tom Furness (1998). “Physically touching and tasting virtual objects enhances the realism of virtual experiences”. In: *Virtual Reality* 3.4, pp. 226–234.

Hromek, Robyn and Sue Roffey (2009). “Promoting social and emotional learning with games:” It’s fun and we learn things””. In: *Simulation & Gaming*.

Hu, Xinda and Hong Hua (2012). “Design of an optical see-through multi-focal-plane stereoscopic 3d display using freeform prisms”. In: *Frontiers in Optics*. Optical Society of America, FTh1F–2.

Huang, Weidong and Leila Alem (2013). “HandsinAir: a wearable system for remote collaboration on physical tasks”. In: *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, pp. 153–156.

Hudson, Matthew and Paul Cairns (2014). “Measuring social presence in team-based digital games”. In: *Interacting with Presence: HCI and the Sense of Presence in Computer-mediated Environments*, p. 83.

Huizinga, Johan (1944). *Homo ludens: Versuch einer bestimmung des spielelementes der kultur*. Akad. Verlag-Anst. Pantheon.

Humphreys, Greg, Matthew Eldridge, Ian Buck, Gordan Stoll, Matthew Everett, and Pat Hanrahan (2001). “WireGL: a scalable graphics system for clusters”. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, pp. 129–140.

Humphreys, Greg, Mike Houston, Ren Ng, Randall Frank, Sean Ahern, Peter D Kirchner, and James T Klosowski (2002). “Chromium: a stream-processing framework for interactive rendering on clusters”. In: *ACM transactions on graphics (TOG)* 21.3, pp. 693–702.

Hunt, James Wayne and MD MacIlroy (1976). *An algorithm for differential file comparison*. Citeseer.

Iwata, Hiroo (1999). “Walking about virtual environments on an infinite floor”. In: *Virtual Reality, 1999. Proceedings., IEEE*. IEEE, pp. 286–293.

Iwata, Hiroo, Hiroaki Yano, and Fumitaka Nakaizumi (2001). “Gait master: A versatile locomotion interface for uneven virtual terrain”. In: *Virtual Reality, 2001. Proceedings. IEEE*. IEEE, pp. 131–137.

Jayaram, Sankar, Hugh I Connacher, and Kevin W Lyons (1997). “Virtual assembly using virtual reality techniques”. In: *Computer-aided design* 29.8, pp. 575–584.

Jennett, Charlene, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton (2008). “Measuring and defining the experience of immersion in games”. In: *International journal of human-computer studies* 66.9, pp. 641–661.

Juan, M Carmen, Rosa Baños, Cristina Botella, David Pérez, Mariano Alcaníiz, and Carlos Monserrat (2006). “An augmented reality system for the treatment of acrophobia: the sense of presence using immersive photography”. In: *Presence: Teleoperators and virtual environments* 15.4, pp. 393–402.

Katz, Brian FG, Dalai Q Felinto, Damien Touraine, David Poirier-Quinot, and Patrick Bourdot (2015). “BlenderVR: Open-source framework for interactive and immersive VR". In: *Virtual Reality (VR), 2015 IEEE*. IEEE, pp. 203–204.

Kauko, Jarmo and Jonna Häkkilä (2010). "Shared-screen social gaming with portable devices". In: *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, pp. 317–326.

Kimura, Makoto, Masaaki Mochimaru, and Takeo Kanade (2007). "Projector calibration using arbitrary planes and calibrated camera". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1–2.

Kirkpatrick, Donald L (1967). "Evaluation of training". In:

— (1975). *Evaluating training programs*. Tata McGraw-Hill Education.

— (1979). "Techniques for evaluating training programs". In: *Training and development journal*.

Kiyokawa, Kiyoshi, Yoshinori Kurata, and Hiroyuki Ohno (2001). "An optical see-through display for mutual occlusion with a real-time stereovision system". In: *Computers & Graphics* 25.5, pp. 765–779.

Kort, Yvonne AW de, Wijnand A IJsselsteijn, and Karolien Poels (2007). "Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ)". In: *Proceedings of PRESENCE*, pp. 195–203.

Kuntz, Sebastien (2015). "Middlevr a generic vr toolbox". In: *Virtual Reality (VR), 2015 IEEE*. IEEE, pp. 391–392.

Lapeer, R, MS Chen, G Gonzalez, A Linney, and G Alusi (2008). "Image-enhanced surgical navigation for endoscopic sinus surgery: evaluating calibration, registration and tracking". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 4.1, pp. 32–45.

Lazzaro, Nicole (2004). "Why we play games: Four keys to more emotion without story". In:

Lee, Kangdon (2012). "Augmented reality in education and training". In: *TechTrends* 56.2, pp. 13–21.

Lego® Creator Sea Plane. `http://shop.lego.com/en-US/Sea-Plane-31028`. Accessed: 2015-05-08.

Li, Hairong, Terry Daugherty, and Frank Biocca (2002). "Impact of 3-D advertising on product knowledge, brand attitude, and purchase intention: The mediating role of presence". In: *Journal of advertising* 31.3, pp. 43–57.

Li, Jing-Rong, Li Pheng Khoo, and Shu Beng Tor (2003). "Desktop virtual reality for maintenance training: an object oriented prototype system (V-REALISM)". In: *Computers in Industry* 52.2, pp. 109–125.

Lindley, Siân E, James Le Couteur, and Nadia L Berthouze (2008). "Stirring up experience through movement in game play: effects on engagement and social behaviour". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 511–514.

Livingston, Mark A, Arindam Dey, Christian Sandor, and Bruce H Thomas (2013). “Pursuit of “X-ray vision” for augmented reality”. In: *Human Factors in Augmented Reality Environments*. Springer, pp. 67–107.

Livingston, Mark A, Lawrence J Rosenblum, Simon J Julier, Dennis Brown, Yohan Baillot, II Swan, Joseph L Gabbard, Deborah Hix, et al. (2002). *An augmented reality system for military operations in urban terrain*. Tech. rep. DTIC Document.

Livingston, Mark A, J Edward Swan II, Joseph L Gabbard, Tobias H Höllerer, Deborah Hix, Simon J Julier, Yohan Baillot, and Dennis Brown (2003). “Resolving multiple occluded layers in augmented reality”. In: *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, p. 56.

Lok, Benjamin, Samir Naik, Mary Whitton, and Frederick P Brooks (2003). “Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments”. In: *Presence: Teleoperators and Virtual Environments* 12.6, pp. 615–628.

Lombard, Matthew and Theresa Ditton (1997). “At the heart of it all: The concept of presence”. In: *Journal of Computer-Mediated Communication* 3.2, pp. 0–0.

Magee, D, Y Zhu, Rish Ratnalingam, P Gardner, and David Kessel (2007). “An augmented reality simulator for ultrasound guided needle placement training”. In: *Medical & biological engineering & computing* 45.10, pp. 957–967.

Maggioni, Christoph (1993). “A novel gestural input device for virtual reality”. In: *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*. IEEE, pp. 118–124.

Maimone, Andrew and Henry Fuchs (2013). “Computational augmented reality eyeglasses”. In: *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*. IEEE, pp. 29–38.

Maimone, Andrew, Xubo Yang, Nate Dierk, Andrei State, Mingsong Dou, and Henry Fuchs (2013). “General-purpose telepresence with head-worn optical see-through displays and projector-based lighting”. In: *2013 IEEE Virtual Reality (VR)*. IEEE, pp. 23–26.

Manca, Davide, Sara Brambilla, and Simone Colombo (2013). “Bridging between virtual reality and accident simulation for training of process-industry operators”. In: *Advances in Engineering Software* 55, pp. 1–9.

Mann, James Fung Steve and James Fung (2001). “Video orbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene”. In: *EYE* 3, P3.

Marino, Giuseppe, Paolo Simone Gasparello, Davide Vercelli, Franco Tecchia, and Massimo Bergamasco (2010). “Network streaming of dynamic 3D content with on-line compression of frame data”. In: *Virtual Reality Conference (VR), 2010 IEEE*. IEEE, pp. 285–286.

Marino, Giuseppe, Davide Vercelli, Franco Tecchia, and Paolo Simone Gasparello (2007). “Description and performance analysis of a distributed rendering architecture for virtual environments”. In: *Artificial Reality and Telexistence, 17th International Conference on.* IEEE, pp. 234–241.

Martynov, Ivan, Joni-Kristian Kamarainen, and Lasse Lensu (2011). “Projector calibration by “inverse camera calibration””. In: *Scandinavian Conference on Image Analysis.* Springer, pp. 536–544.

McGrath, A, A Oldroyd, and G Walker (1998). “The mirror: Reflections on inhabited TV”. In: *Proceedings of ACM Conference/Video Programme on Computer Supported Cooperative Work*, pp. 14–18.

Mead, George Herbert (1934). “Mind, Self and Society from the standpoint of a social behaviorist.[Edited and with an Introduction by Charles W. Morris]”. In: *University of Chicago Press, Chicago, London* 18, p. 1972.

Meehan, Michael, Sharif Razzaque, Mary C Whitton, and Frederick P Brooks (2003). “Effect of latency on presence in stressful virtual environments”. In: *virtual reality, 2003. Proceedings. IEEE.* IEEE, pp. 141–148.

Milgram, Paul and Fumio Kishino (1994). “A taxonomy of mixed reality visual displays”. In: *IEICE TRANSACTIONS on Information and Systems* 77.12, pp. 1321–1329.

Molnar, Steven, Michael Cox, David Ellsworth, and Henry Fuchs (1994). “A sorting classification of parallel rendering”. In: *Computer Graphics and Applications, IEEE* 14.4, pp. 23–32.

Molnar, Steven, John Eyles, and John Poulton (1992). “PixelFlow: high-speed rendering using image composition”. In: *ACM SIGGRAPH Computer Graphics.* Vol. 26. 2. ACM, pp. 231–240.

Mortensen, Jesper, Insu Yu, Pankaj Khanna, Franco Tecchia, Bernhard Spanlang, Giuseppe Marino, and Mel Slater (2008). “Real-time global illumination for vr applications”. In: *IEEE Computer Graphics and Applications* 6, pp. 56–64.

Muhanna, Muhanna A (2015). “Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions”. In: *Journal of King Saud University-Computer and Information Sciences* 27.3, pp. 344–361.

Normand, Jean-Marie, Bernhard Spanlang, Franco Tecchia, Marcello Carrozzino, David Swapp, and Mel Slater (2012). “Full body acting rehearsal in a networked virtual environment—A case study”. In: *Presence: Teleoperators and Virtual Environments* 21.2, pp. 229–243.

O’Regan, J Kevin and Alva Noë (2001). “A sensorimotor account of vision and visual consciousness”. In: *Behavioral and brain sciences* 24.05, pp. 939–973.

Parsons, Thomas D and Albert A Rizzo (2008). “Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis”. In: *Journal of behavior therapy and experimental psychiatry* 39.3, pp. 250–261.

Peng, Wei and Gary Hsieh (2012). “The influence of competition, cooperation, and player relationship in a motor performance centered computer game”. In: *Computers in Human Behavior* 28.6, pp. 2100–2106.

Pérez Marcos, Daniel, Massimiliano Solazzi, William Steptoe, Oyewole Oyekoya, Antonio Frisoli, Tim Weyrich, Anthony Steed, Franco Tecchia, Mel Slater, and María Victoria Sánchez-Vives (2012). “A fully immersive set-up for remote interaction and neurorehabilitation based on virtual body ownership”. In: *Frontiers in Neurology, 2012, vol. 3, p. 110.*

Phillips, Jack J (1996). “How much is the training worth?” In: *Training & Development* 50.4, pp. 20–25.

Plass, Jan L, Paul A O’Keefe, Bruce D Homer, Jennifer Case, Elizabeth O Hayward, Murphy Stein, and Ken Perlin (2013). “The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation.” In: *Journal of Educational Psychology* 105.4, p. 1050.

Popescu, V, Grigore Burdea, and Mourad Bouzit (1999). “Virtual reality simulation modeling for a haptic glove”. In: *Computer Animation, 1999. Proceedings.* IEEE, pp. 195–200.

Popescu, Viorel G, Grigore C Burdea, Mourad Bouzit, and Vincent R Hentz (2000). “A virtual-reality-based telerehabilitation system with force feedback”. In: *IEEE transactions on Information Technology in Biomedicine* 4.1, pp. 45–51.

Qin, Hua, Pei-Luen Patrick Rau, and Gavriel Salvendy (2009). “Measuring player immersion in the computer game narrative”. In: *Intl. Journal of Human–Computer Interaction* 25.2, pp. 107–133.

Raskar, Ramesh, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs (1998). “The office of the future: A unified approach to image-based modeling and spatially immersive displays”. In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques.* ACM, pp. 179–188.

Rettie, Ruth (2003). “Connectedness, awareness and social presence”. In:

Rocha, José Bernardo, Samuel Mascarenhas, and Rui Prada (2008). “Game mechanics for cooperative games”. In: *ZDN Digital Game*, pp. 73–80.

Rolland, Jannick P (2000). “Wide-angle, off-axis, see-through head-mounted display”. In: *Optical engineering* 39.7, pp. 1760–1767.

Rolland, Jannick P and Henry Fuchs (2000). “Optical versus video see-through head-mounted displays in medical visualization”. In: *Presence: Teleoperators and Virtual Environments* 9.3, pp. 287–309.

Romero, Margarida, Mireia Usart, Michela Ott, Jeffrey Earp, and Sara de Freitas (2012). “Learning through playing for or against each other? Promoting collaborative learning in digital game based learning”. In: *Learning* 5, pp. 15–2012.

Rosenthal, Michael, Andrei State, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta Pisano, Michael Jiroutek, Keith Muller, and Henry Fuchs (2001). “Augmented reality guidance for needle biopsies:

A randomized, controlled trial in phantoms". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*. Springer, pp. 240–248.

Sajjadi, Pejman, Edgar Omar Cebolledo Gutierrez, Sandra Trullemans, and Olga De Troyer (2014). "Maze commander: a collaborative asynchronous game using the oculus rift & the sifteo cubes". In: *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*. ACM, pp. 227–236.

Santos, Pedro, Thomas Gierlinger, Oliver Machui, and André Stork (2008). "The daylight blocking optical stereo see-through HMD". In: *Proceedings of the 2008 workshop on Immersive projection technologies/Emerging display technologiges*. ACM, p. 4.

Schmierbach, Mike, Qian Xu, Anne Oeldorf-Hirsch, and Frank E Dardis (2012). "Electronic friend or virtual foe: exploring the role of competitive and cooperative multiplayer video game modes in fostering enjoyment". In: *Media Psychology* 15.3, pp. 356–371.

Seif El-Nasr, Magy, Bardia Aghabeigi, David Milam, Mona Erfani, Beth Lameman, Hamid Maygoli, and Sang Mah (2010). "Understanding and evaluating cooperative games". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 253–262.

Seth, Abhishek, Judy M Vance, and James H Oliver (2011). "Virtual reality for assembly methods prototyping: a review". In: *Virtual reality* 15.1, pp. 5–20.

Sheridan, Thomas B (1992). "Musings on telepresence and virtual presence". In: *Presence: Teleoperators & Virtual Environments* 1.1, pp. 120–126.

Sherman, William R and Alan B Craig (2002). *Understanding virtual reality: Interface, application, and design*. Elsevier.

Short, John, Ederyn Williams, and Bruce Christie (1976). "The social psychology of telecommunications". In:

Shreiner, Dave, Bill The Khronos OpenGL ARB Working Group, et al. (2009). *OpenGL programming guide: the official guide to learning OpenGL, versions 3.0 and 3.1*. Pearson Education.

Siu, Kristin, Alexander Zook, and Mark O Riedl (2014). "Collaboration versus competition: Design and evaluation of mechanics for games with a purpose". In: *Proceedings of the 9th International Conference on the Foundations of Digital Games*.

Siwek, Stephen E (2014). "Video Games in the 21st Century
The 2014 report". In: *Entertainment software association*.

Slater, Mel (2009). "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535, pp. 3549–3557.

Slater, Mel, Antonio Frisoli, Franco Tecchia, Christoph Guger, Beau Lotto, Anthony Steed, Gert Pfurtscheller, Robert Leeb, Miriam Reiner, Maria V Sanchez-Vives, et al. (2007). "Understanding and realizing presence in the Presenccia project”. In: *IEEE computer graphics and applications* 27.4, pp. 90–93.

Slater, Mel and Maria V Sanchez-Vives (2016). “Enhancing our lives with immersive virtual reality”. In: *Frontiers in Robotics and AI* 3, p. 74.

Slater, Mel, Martin Usoh, and Anthony Steed (1995). “Taking steps: the influence of a walking technique on presence in virtual reality”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3, pp. 201–219.

Souman, Jan L, P Robuffo Giordano, Martin Schwaiger, Ilja Frissen, Thomas Thümmel, Heinz Ulbrich, A De Luca, Heinrich H Bülthoff, and Marc O Ernst (2011). “CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments”. In: *ACM Transactions on Applied Perception (TAP)* 8.4, p. 25.

Sportillo, Daniele, Giovanni Avveduto, Franco Tecchia, and Marcello Carrozzino (2015). “Training in VR: A Preliminary Study on Learning Assembly/Disassembly Sequences”. In: *International Conference on Augmented and Virtual Reality*. Springer, pp. 332–343.

Stark, Lawrence W (1995). “How virtual reality works: illusions of vision in real and virtual environments”. In: *IS&T/SPIE’s Symposium on Electronic Imaging: Science & Technology*. International Society for Optics and Photonics, pp. 277–287.

State, Andrei, Mark A Livingston, William F Garrett, Gentaro Hirota, Mary C Whitton, Etta D Pisano, and Henry Fuchs (1996). “Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies”. In: *Proceedings of the 23rd annual conference on computer graphics and interactive techniques*. ACM, pp. 439–446.

Steed, Anthony, Franco Tecchia, Massimo Bergamasco, Mel Slater, William Steptoe, Wole Oyekoya, Fabrizio Pece, Tim Weyrich, Jan Kautz, Doron Friedman, et al. (2012). “Beaming: an asymmetric telepresence system”. In: *IEEE computer graphics and applications* 6, pp. 10–17.

Steuer, Jonathan (1992). “Defining virtual reality: Dimensions determining telepresence”. In: *Journal of communication* 42.4, pp. 73–93.

Stoll, Gordon, Matthew Eldridge, Dan Patterson, Art Webb, Steven Berman, Richard Levy, Chris Caywood, Milton Taveira, Stephen Hunt, and Pat Hanrahan (2001). “Lightning-2: a high-performance display subsystem for PC clusters”. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, pp. 141–148.

Suma, Evan A, David M Krum, and Mark Bolas (2011). “Sharing space in mixed and virtual reality environments using a low-cost depth sensor”. In: *VR Innovation (ISVRI), 2011 IEEE International Symposium on*. IEEE, pp. 349–350.

Sutherland, Ivan E (1968). “A head-mounted three dimensional display”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. ACM, pp. 757–764.

Sweetser, Penelope and Peta Wyeth (2005). “GameFlow: a model for evaluating player enjoyment in games”. In: *Computers in Entertainment (CIE)* 3.3, pp. 3–3.

Tauer, John M and Judith M Harackiewicz (2004). “The effects of cooperation and competition on intrinsic motivation and performance.” In: *Journal of personality and social psychology* 86.6, p. 849.

Tecchia, Franco, Leila Alem, and Weidong Huang (2012). “3D helping hands: a gesture based MR system for remote collaboration”. In: *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry.* ACM, pp. 323–328.

Tecchia, Franco, Giovanni Avveduto, Raffaello Brondi, Marcello Carrozzino, Massimo Bergamasco, and Leila Alem (2014). “I’m in VR!: using your own hands in a fully immersive MR system”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology.* ACM, pp. 73–76.

Tecchia, Franco, Giovanni Avveduto, Marcello Carrozzino, Raffaelo Brondi, Massimo Bergamasco, and Leila Alem (2014). “Interacting with your own hands in a fully immersive MR system”. In: *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on.* IEEE, pp. 313–314.

Tecchia, Franco, Marcello Carrozzino, Sandro Bacinelli, Fabio Rossi, Davide Vercelli, Giuseppe Marino, Paolo Gasparello, and Massimo Bergamasco (2010). “A flexible framework for wide-spectrum VR development”. In: *Presence: Teleoperators and Virtual Environments* 19.4, pp. 302–312.

Templeman, James N, Patricia S Denbrook, and Linda E Sibert (1999). “Virtual locomotion: Walking in place through virtual environments”. In: *Presence: teleoperators and virtual environments* 8.6, pp. 598–617.

University of Minnesota. *PowerWall.* `http://www.lcse.umn.edu/research/powerwall/powerwall.html`. [Online; accessed 21-February-2016].

Van Krevelen, DWF and R Poelman (2010). “A survey of augmented reality technologies, applications and limitations”. In: *International Journal of Virtual Reality* 9.2, p. 1.

Van Wyk, Etienne and Ruth De Villiers (2009). “Virtual reality training applications for the mining industry”. In: *Proceedings of the 6th international conference on computer graphics, virtual reality, visualisation and interaction in Africa.* ACM, pp. 53–63.

Vlahakis, Vassilios, John Karigiannis, Manolis Tsotros, Michael Gounaris, Luis Almeida, Didier Stricker, Tim Gleue, Ioannis T Christou, Renzo Carlucci, and Nikos Ioannidis (2001). “Archeoguide: first results of an augmented reality, mobile computing system in cultural heritage sites”. In: *Virtual Reality, Archeology, and Cultural Heritage* 9.

Wang, Zheng, Kurt Bauernfeind, and Thomas Sugar (2003). “Omni-directional treadmill system”. In: *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on.* IEEE, pp. 367–373.

Ware, Colin and Steven Osborne (1990). “Exploration and virtual camera control in virtual three dimensional environments”. In: *ACM SIGGRAPH Computer Graphics*. Vol. 24. 2. ACM, pp. 175–183.

Welch, Greg, Gary Bishop, Leandra Vicci, Stephen Brumback, Kurtis Keller, and D’nardo Colucci (2001). “High-performance wide-area optical tracking: The hiball tracking system”. In: *presence: teleoperators and virtual environments* 10.1, pp. 1–21.

Welch, Robert B, Theodore T Blackmon, Andrew Liu, Barbara A Mellers, and Lawrence W Stark (1996). “The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence”. In: *Presence: Teleoperators & Virtual Environments* 5.3, pp. 263–273.

Whitton, Mary C (2003). “Making virtual environments compelling”. In: *Communications of the ACM* 46.7, pp. 40–47.

Whitton, Mary C, Joseph V Cohn, Jeff Feasel, Paul Zimmons, Sharif Razzaque, Sarah J Poulton, Brandi McLeod, and Frederick P Brooks (2005). “Comparing VE locomotion interfaces”. In: *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*. IEEE, pp. 123–130.

Williams, Betsy, Stephen Bailey, Gayathri Narasimham, Muqun Li, and Bobby Bodenheimer (2011). “Evaluation of walking in place on a wii balance board to explore a virtual environment”. In: *ACM Transactions on Applied Perception (TAP)* 8.3, p. 19.

Witmer, Bob G and Michael J Singer (1998). “Measuring presence in virtual environments: A presence questionnaire”. In: *Presence: Teleoperators and virtual environments* 7.3, pp. 225–240.

Wolf, Mark JP (2008). *The video game explosion: a history from PONG to Playstation and beyond*. ABC-CLIO.

Xu, Yan, Xiang Cao, Abigail Sellen, Ralf Herbrich, and Thore Graepel (2011). “Sociable killers: understanding social relationships in an online first-person shooter game”. In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, pp. 197–206.

Yee, Nick (2006a). “Motivations for play in online games”. In: *CyberPsychology & behavior* 9.6, pp. 772–775.

— (2006b). “The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments”. In: *Presence* 15.3, pp. 309–329.

— (2006c). “The psychology of MMORPGs: Emotional investment, motivations, relationship formation, and problematic usage”. In: *Avatars at work and play: Collaboration and interaction in shared virtual environments* 34, pp. 187–207.

You, Suya and Ulrich Neumann (2001). “Fusion of vision and gyro tracking for robust augmented reality registration”. In: *Virtual Reality, 2001. Proceedings. IEEE*. IEEE, pp. 71–78.

Zachmann, Gabriel and Alexander Rettig (2001). “Natural and robust interaction in virtual assembly simulation”. In: *Eighth ISPE Interna-*

tional Conference on Concurrent Engineering: Research and Applications (ISPE/CE2001). Vol. 1. Citeseer, pp. 425–434.

Zagal, José P, Jochen Rick, and Idris Hsi (2006). "Collaborative games: Lessons learned from board games". In: *Simulation & Gaming* 37.1, pp. 24–40.

Zheng, Ye, Matthew McCaleb, Courtney Strachan, and Betsy Williams (2012). "Exploring a virtual environment by walking in place using the Microsoft Kinect". In: *Proceedings of the ACM symposium on applied perception.* ACM, pp. 131–131.

Zhou, Feng, Henry Been-Lirn Duh, and Mark Billinghurst (2008). "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR". In: *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality.* IEEE Computer Society, pp. 193–202.

Zohar, Dov and Ido Erev (2006). "On the difficulty of promoting workers' safety behaviour: overcoming the underweighting of routine risks". In: *International Journal of Risk Assessment and Management* 7.2, pp. 122–136.

INSTITUTE
OF COMMUNICATION,
INFORMATION
AND PERCEPTION
TECHNOLOGIES

Sant'Anna
School of Advanced Studies – Pisa