One Reality

Augmenting the human experience through the combination of physical and digital worlds

PhD Thesis Manuscript by:

Joan Sol ROO

Supervised by: Martin HACHET
At: Université de Bordeaux / Inria – Potioc

Jury members of the PhD defence :

Nicolas ROUSSEL	*Président*
Eva HORNECKER	*Rapporteur*
Anatole LÉCUYER	*Rapporteur*
Jürgen STEIMLE	*Examinateur*
Nadine COUTURE	*Examinateur*

PhD defense: 15/12/2017

"Computing is not about computers anymore.

It's about living."

– Nicholas Negroponte

Abstract

In recent history, computational devices evolved from simple calculators to now pervasive artefacts, with which we share most aspects of our lives, and it is hard to imagine otherwise. Yet, this change of the role of computers was not accompanied by an equivalent redefinition of the interaction paradigm: we still mostly depend on screens, keyboards and mice. Even when these legacy interfaces have been proven efficient for traditional tasks, we agree with those who argue that these interfaces are not necessarily fitting for their new roles. Even more so, traditional interfaces preserve the separation between digital and physical realms, now counterparts of our reality.

During this PhD, we focused the dissolution of the separation between physical and digital, first by extending the reach of digital tools into the physical environment, followed by the creation of hybrid artefacts (physical-digital emulsions), to finally support the transition between different mixed realities, increasing immersion only when needed.

The final objective of this work is to augment the experience of reality. This comprises not only the support of the interaction with the external world, but also with the internal one. This thesis provides the reader contextual information along with required technical knowledge to be able to understand and build mixed reality systems. Once the theoretical and practical knowledge is provided, our contributions towards the overarching goal of merging physical and digital realms are presented. We hope this document will inspire and help others to work towards a world where the physical and digital, and humans and their environment are not opposites, but instead all counterparts of a unified reality.

Resumé

Alors que le numérique a longtemps été réservé à des usages experts, il fait aujourd'hui partie intégrante de notre quotidien, au point, qu'il devient difficile de considérer le monde physique dans lequel nous vivons indépendamment du monde numérique. Pourtant, malgré cette évolution, notre manière d'interagir avec le monde numérique a très peu évolué, et reste toujours principalement basé sur l'utilisation d'écrans, de claviers et de souris. Dans les nouveaux usages rendus possible par le numérique, ces interfaces peuvent se montrer inadaptées, et continuent à préserver la séparation entre le monde physique et le monde numérique.

Au cours de cette thèse, nous nous sommes concentrés à rendre cette frontière entre mondes physique et numérique plus subtil au point de la faire disparaître. Cela est rendu possible en étendant la portée des outils numériques dans le monde physique, puis en concevant des artefacts hybrides (des objets aux propriétés physique et numérique), et enfin en permettant les transitions dans une réalité mixte (physique-numérique), laissant le choix du niveau d'immersion à l'utilisateur en fonction de ses besoins.

L'objectif final de ce travail est d'augmenter l'expérience de la *réalité*. Cela comprend non seulement le support de l'interaction avec le monde extérieur, mais aussi avec notre monde intérieur. Cette thèse fournit aux lecteurs les informations contextuelles et les connaissances techniques requises pour pouvoir comprendre et concevoir des systèmes de réalité mixte. A partir de ces fondements, nos contributions, ayant pour but de fusionner le monde physique et le monde virtuel, sont présentées. Nous espérons que ce document inspirera et facilitera des travaux futurs ayant pour vision d'unifier le physique et le virtuel.

Acknowledgements

"I am the combined effort of everyone I've ever known."
(And even the ones I didn't ever met)
- Chuck Palahniuk, Invisible Monsters

This thesis was not a solo effort. Most of the ideas were the maturation of long discussions, with inspiration driven from many places. Without the help of many, this work would not have been possible.

To the people that was there for me

First, I must thank my family, which supported my decision to follow a research career, even if it meant being far away from them. Almost as important, I must thank my friends and colleagues at *Potioc*, most of which already left to continue their lives. Indeed, this thesis would not have been possible if it was not thanks to *Martin Hachet*, *Renaud Gervais*, *Jérémy Frey*, *Damien Clergeaud, Pierre-Antoine Cinquin* and *Jean Basset*. These pages are the results of countless hours of discussion with them, along with *Jérémy Laviole, Thibault Laine*, *Philippe Giraudeau*, *Léa Pillette*, *Anke Brock*, and *Julia Chatain*. They and others were there for me not only research-wise: they were my family for the past 3 years.

To the sources of inspiration and the scientific community

I must also thank many people that, even when I don't personally know them, they inspired me along the way. In no particular order: thanks to *Carl Sagan* for showing me the poetry in science; thanks to the fiction writers (*Philip K. Dick*, *William Gibson* and *Neal Stephenson,* among many others) for their views on the future, both on how it *could* and how it *shouldn't* be.

It goes without saying that this would not have been possible without the literature. I must thank the people that faced the hard problems and gave us the tools to take the next step, and the ones that lifted their heads past the current challenge, and were brave enough to challenge the current path of advance (too many to list by name here, but cited along the manuscript). To all the ones I can name, and to the ones their efforts but not their names make the history of human development, thank you.

To the founding and infrastructure

And all of this would not have been possible without the ISAR project (ANR-14- CE24-0013), financed by the French *Agence Nationale de la Recherche (ANR)*, in combination with the infrastructure provided by *Inria* and *Potioc.*

Contents

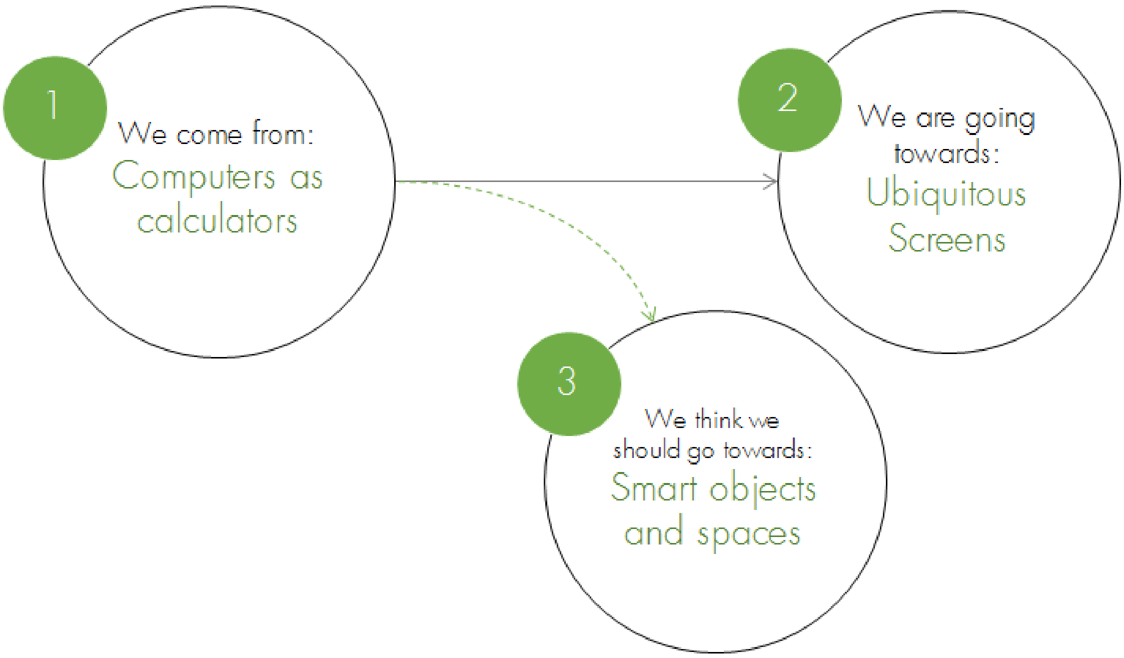
PART I

Context, Theory and Implementation

PART I: Context, Theory and Implementation

1. Introduction

"The visions we offer our children shape the future.
It matters what those visions are. Often they become self-fulfilling prophecies.
Dreams are maps."
— Carl Sagan, Pale Blue Dot: A Vision of the Human Future in Space



About this section:

This section takes a look at the role of the digital realm in our lives, which evolved faster than the interfaces we use to interact with the digital tools. In parallel to this fast evolution, researchers and fiction writers imagined how physical and digital worlds could work together. This section presents a quick overview of these visions, laying the grounds to the contributions presented later in this manuscript.

1.1. Technology evolved faster than interaction

The history of computing is a new one, with rapid changes. Not so long ago, computers were colossal room-sized machines, conceived as a way of automatizing calculation and data processing. Since then, they have become cheaper, smaller and interconnected, rendering them ubiquitously available, to the point where it is hard to imagine our daily life without them. This rapid evolution was, as natural evolution itself, mostly incremental. As a result, these systems and interfaces improved almost aimlessly, taking the next possible step instead of the ideal step. Yet, it is important to remember that incremental processes can inadvertently preserve obsolete constraints.

1.1.4. A short story about legacy interfaces

A classic example of legacy constraints are keyboards: the most common modern layouts were based on the ones from mechanical typewriters; yet these layouts were initially created to prevent mechanical locking, a problem no longer relevant for computer keyboards (and even more so in the case of tactile versions). Still, attempts to provide more adapted layouts have failed to become the standard.

Figure 1: The legacy of mechanical typewriters is still very present in the digital age, even in fully tactile screens

1.1.5. The legacy is not enough

Even when much have changed since their first conception, the influence of the first computers is still visible on the paradigms ruling contemporary computational activities, and how we interact with them. As a result, digital and physical realms are separated by a legacy interface: screens, mice and keyboards. We use a single kind of device (even if we have several of them), and we use it to work, get our news, socialize and relax. This interface is well suited to data entry and processing, but now takes many roles in a single form factor. This generic interface provide us a window, through which we need to jump in order to reach the digital realm.

In order to use digital media, we need to leave our body and environment behind. Even when our minds are able to perform such a task, it can be argued that this scenario is far from ideal: we live our lives in both the physical and digital worlds, yet they are isolated from each other. This leads to the question:

Could it be possible for the physical and digital realms to coexist and cooperate?

This question is not new, in fact, is one of the main overarching challenges for the field of Human-Computer Interaction (HCI). In this manuscript we present our efforts towards addressing this question, yet far from being the first to do so, this work builds on top of over 50 years of research.

1.2. The convergence of *Physical* and *Digital*

The desire to dissolve the separation of physical and digital is probably as old as the digital realm itself.

Back in 1965 when computers were still colossal devices, Ivan Sutherland described his view of the *ultimate display* [198]: a room where matter could be controlled by a computer. Soon after that, in 1968 Sutherland and his colleagues created the first 3D head mounted display, *The Sword of Damocles*, which supported primitive 3D rendering both superimposed to the physical world or in an immersive fashion, the latter called Virtual Reality (VR). Since then, much effort has been placed into novel ways to override the senses, and VR become a strongly researched topic for simulation, tele-presence, collaboration, and gaming, among many others.

In 1991, once the era of personal computers arrived, Mark Weiser described his vision of the next step: *ubiquitous computing*, where computation will be always available when needed [221,222]. Even when it is commonly argued that nowadays the ratio between users and devices is close to what Weiser anticipated, a cornerstone of ubiquitous computing was their *calm*: technology disappearing into the environment much like classic physical tools do [223]. For Weiser, ubiquitous computing is diametrically opposite to virtual reality: *"virtual reality focuses an enormous apparatus on simulating the world rather than on invisibly enhancing the* world that already exists" [221].

When facing this dichotomy (focusing on what exists versus simulating new worlds), Milgram instead considered both realms laying along continuum [135]. Along the spectrum, the physical world resides in one extreme, while completely digitally generated spaces reside on the opposite extreme. These realms can be combined on different degrees by means of sensory stimulation, creating a Mixed Reality (MR). For instance, Augmented Reality (AR) complements the physical space with digital information, while Augmented Virtuality (AV) places physical elements inside the virtual space.

Designing towards the collaboration of physical and digital, Tangible User Interfaces (TUIs) [182], which in its original sense meant interfaces that provide physical handles to digital information, and now refers more broadly to interfaces that follow the rules of interaction with the physical world (including other humans). The notion of Tangible User Interfaces was later extended with the vision of Radical Atoms [92]: computational matter that can change its physical properties as needed, an indivisible emulsion of digital and physical. Radical atoms are a suiting follow up step to the idea of ubiquitous computing: computation always available and indistinguishable from the physical world.

As physical and digital combine, the notion of "real" becomes diffuse. Eissele et al. [52] reminded us the philosophical distinction between the Greek words *"realitas"* (reality, what is constructed in the mind) and *"actualitas"* (actuality, the invariant truth), a meaning lost in translation. This is an especially relevant separation when studying mixed reality, which is performed by altering the information received by the senses. By using the Greek definition, when digital and physical work together, they are just counterparts of the same reality for the user. This is the definition used in this manuscript, and the reason why it is entitled "One Reality".

By contrasting the previous page with the present one, it is possible to see that there is a clear difference between what we use in our everyday lives and our field of study. Indeed, 50 years of research did not change the fact we still use computers and screens, albeit smaller and more powerful every day. We consider that in order to allow these approaches to become widespread it is necessary to: (1) explicitly explore their benefits for a more humane way of interaction, and (2) support their compatibility with the current approaches. These two aspects are explored on the following pages.

1.3. Designing humane futures

The questions of what it is and what it could be are closely related, as can be seen by the dialog between research and fiction. Historically, technological research focused on the feasibility and implementation of systems (mostly through an incremental approach), while fiction focused on the implications for society of technological advance (even without it being possible at the time). This trend has shifted towards a more holistic approach, where technique and human implications are considered together for the next generation of applications.

In research, fiction is not only a source of inspiration but also can help us question the implications beyond technology. For instance, virtuality has been considered countless times in philosophy and popular fiction, particularly around the topics of reality, perception, humanism and trans-humanism. An early example of this is *Star Trek's HoloDeck* (a materialization of the ultimate display), which led to the exploration of philosophical and sociological questions of VR [59]. Also inspired by *Star Trek* is the project questioning the value of unique objects in the era of digital replication [141]. Another recurrent topic in fiction is the fusion of humans and machines, as a way to explore *what makes us humans*. This includes pieces such as Shirow's "*Ghost in the shell"* [150] and Gibson's "*Neuromancer"* [69], now explicitly cited as source of inspiration by renowned researchers such as Rekimoto [107,108]. One of the most renowned authors exploring the human condition through possible futures is P.K. Dick. Besides his influence in mainstream media (movies such as *Minority report* [190] and *Blade runner* [176]), he was the source of inspiration for projects such as the *empathy box* [25] (taken from "*Do androids dream of electric sheep"*) [45].

On the general topic of the possible futures of physical-digital realities, David Rose presents his vision [168] much in the way fiction writers would do. He lists the futures of technology as falling in one of several trajectories: terminal World (screens), prosthetics/wearables (transhumanism), animism (agents), and enchanted objects (everyday objects with their features augmented). Rose invite us to not be mere observers of the technological advance, but to take action towards the future we want; to this end, he constantly refers to fiction as a source of insight: our dreams reflect our aspirations. According to Bret Victor, technology will continue evolving blindly on its original direction unless we direct it, and he argues that we must actively channel its advance so it is empowering, instead of limiting. For him, the implications go beyond interaction, as representations shape the way we think.

Beyond interaction itself, extending the scope of HCI from traditional tasks to our daily lives forces us to reconsider the roles of the involved elements. Humans become more than end-users, computers can take any shape and form, extending the interaction space. We should ask ourselves: *What values is our system supporting?* [181]. With a similar position, Sebastian Deterding argues that every design carries a set of values, either voluntarily or not, and we should conscientiously chose them [178]. More concretely, *positive computing* [30] argues in favour of addressing intrinsic human aspirations, as a way to use technology as a mean to help humanity quest towards its ultimate goal: happiness (not in its temporary sense, but as in *feeling fulfilled*).

Science is a tool to understand reality, but in the field of HCI we do more than observe, we create new tools that redefine how we perceive and interact with the world. The mental exercise of exploring the impact of potential technologies allow us to decide if we should or should not pursuit such possibilities. This is, at its roots, a very subjective process based on an opinion and personal drive, even if backed by previous research. Nevertheless, we cannot ignore the consequences of technological advance, and we should then embrace this fact and take responsibility of our designs. That being said, now is the right moment to present our approach.

1.4. Our Approach to Combine *Physical* and *Digital*

1.4.4. The Contribution

During this PhD, we focused on the dissolution of the separation between physical and digital, trying to address the following overarching question:

- *Could it be possible for the physical and digital realms to coexist and cooperate?*

In order to build towards addressing this question, this document addresses a set of sub-questions:

- *Can we use traditional digital interaction paradigms to interact with the physical world?*
- *How can we allow the progressive transition between physical and digital spaces?*
- *Can users understand and accurately operate when transitioning between representations?*
- *Can this hybrid approach be used for existing applications that already depend on the collaboration between physical and digital?*
- *Can we use digital technology to ground ourselves on the physical world?*
- *Could be use this approach to support remembering and storytelling?*

The first 3 questions are addressed in **PART II**, which focuses on interaction and fundamental research. As a first step, we extended the reach of digital tools into the physical environment allowing the creation of hybrid artefacts (physical-digital emulsions). Then, we present techniques that allow the transition between different mixed realities, increasing the immersion only when needed. Finally, we experimentally measured the user accuracy for tasks involving heterogeneous representations.

The last 3 questions are addressed in **PART III**, which focuses on applications. As a first example, we explored the asymmetric collaboration between physical and virtual locations. The final objective of this work is to augment the experience of reality, and that comprises not only the support of the interaction with the external world, but also with the internal one. For this reason, the final part of the manuscript presents our attempts to use technology as a medium to foster introspection (with the final goal of increase wellbeing and lead to a happier life), and as a support to share stories (both moments from the past, and fantastic ones).

When evaluation was relevant, it was performed with what we considered the appropriated methods and target population. As stated before, there is no single answer to how we can interact with systems. For this reason, the evaluation of the proposed contributions involve diverse methods, depending on the final functionality and the state of the prototypes. We focused on both qualitative and quantitative methods, and the evaluated metrics range from satisfaction to accuracy, including relaxation and mindfulness.

* * * * *

1.4.5. The Approach

In order to get closer to our objective, we explored different applications by creating prototypes, making our priority to allow existing tools (both traditional and digital) to seamlessly interact with each other. We opted for physical surfaces and materials already available in our everyday lives (pens, papers and desks) along virtual objects materialized through digital fabrication, in combination with digital tools.

From a technical standpoint, we opted to use displays and sensors, as they sit at the limit between physical and digital realms. All our prototypes involved spatial augmented reality (SAR) and tangible user interfaces (TUI), which by definition focus on what we are trying to accomplish. SAR and TUIs were used in combination with traditional computers, see-through devices and immersive displays such as head mounted displays, all complementary in nature. When the users' internal state was relevant for the application, we also used physiological computing (that is, the use of physiological activity such as breathing and heart rate to control the systems).

Besides the understanding of the technological and interaction components, it is critical to have a deep understanding of the element at the centre of interest: *Humans*. More than brains taking rational decisions to direct our muscles, we are complex creatures designed through evolution to learn and interact with the world and each other. Because of this reason, we base our approach in Cognitive Science, with influences of design and a touch of artistic sensitivity. These related fields are by no means our area of expertise, but without them we would be doomed to fall once again at placing the computer at the centre of it all.

Even when these contributions bring both spaces together, they are by no means final systems. Instead, they serve as proofs-of-concept, trying to communicate our view: humans, objects and spaces and their digital counterparts are all part of the same, and we can steer the advance of technology towards supporting their collaboration in harmony. In the future, the used technologies might (and surely will) become obsolete, but we hope the knowledge presented in this document will still be relevant in our path towards a unified space containing where physical and digital coexist and collaborate instead of competing for the user's attention.

* * * * *

1.5. About this manuscript

Even when our overarching goal of this work is to augment the human experience, the process requires equal parts of theoretical and technical knowledge. Indeed, we face the difficult problem of supporting interaction outside of the traditional reach of computers. For this reason, this manuscript describes a bottom-up approach, without losing from sight the main goal. The document is structured as follows (Figure 2):

- **PART I** presents the context (of which this *Introduction* is part of), theoretical background and interaction possibilities enabled by mixed reality systems.
- **PART II** presents our approach to the creation of hybrid systems, starting with the combination of augmented objects and desktop computers, to then show a unified physical-digital space, and the evaluation of the human capabilities to create unified mental models when interacting with such spaces.
- **PART III** presents the implementation of specific applications based on the frameworks presented in PART II, including the support of asymmetric collaborative scenarios in the context of the Aerospace Industry, the creation of artefacts that support user's introspection, and the discussion of the potential of augmented objects as source of stories.
- **PART IV** provides an overview of the manuscript, and discusses what we learned along the way, and lays the ground for possible future works.

To ease the reading, additional information is placed in appendices, as follows:

- **APENDIX A** presents the required technical knowledge to take the theory into practice, including the behaviour of cameras and projectors, their calibration, to enable their usage. This appendix covers the technical details required to implement Chapter 2.
- **APENDIX B** presents an alternative reading order for this document, based on the actual process followed during this PhD. This appendix is provided as a way to show that research is not as straightforward as the stories we use to communicate our results.

Without further due, let's take a look at the theoretical bases on which this thesis was built upon.

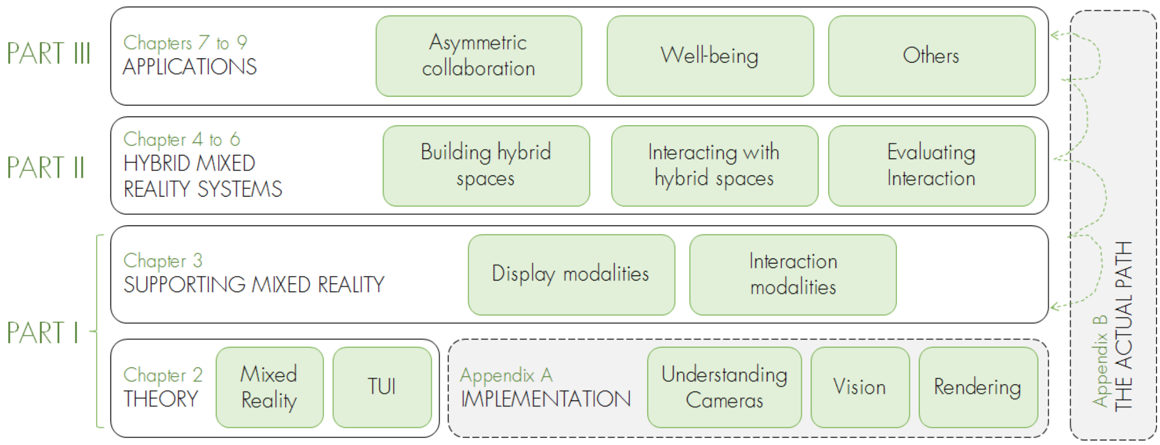


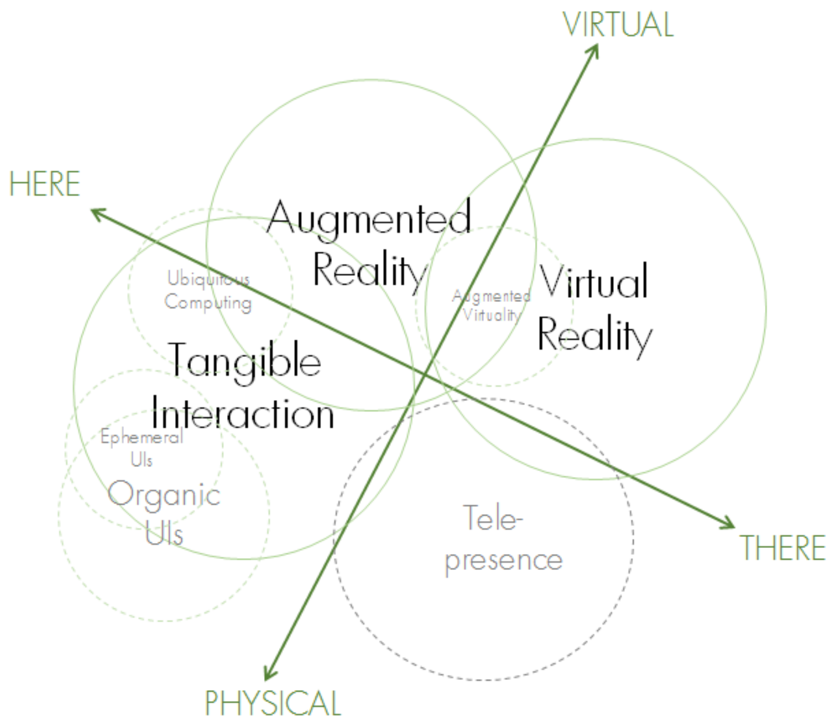
Figure 2: This manuscript is presented using a bottom-up approach, leaving the technical details and a more personal view of the process as appendices

PART I: Context, <u>Theory</u> and Implementation

2. Mixed Reality and Tangible User Interfaces

"It is time to realize [that Tangible interaction] leads away from HCI and into the realm of human interfaces in general. When this happens, the fields of discourse change."

-Kenneth Fishkin, A taxonomy for and analysis of tangible interfaces



About this section:

This Chapter covers the literature classifying mixed reality systems and tangible interaction, as a way to understand the different dimensions addressed by these systems. This will ease the understanding of the proposed solutions, since our objective is to support the usage of complementary approaches simultaneously.

2.1. Mixed Reality Systems

This section presents an overview of the literature classifying the existing techniques to combine physical and digital content. As explained in the Introduction, the use of the world "*real*" can be quite confusing in the context of mixed reality, and for this reason any mention in the literature of the *"real world"* is replaced with *"physical world".*

This section will show the different taxonomies and classification spaces that provide the four key aspects to understand a mixed reality system, namely:

- **Implementation:** *How* and -- most importantly -- *Where* the information is displayed?
- **Artificiality:** Up to which extent is this information digitally created?
- **Transportation:** Is the illusion trying to transport the user somewhere else?
- **Presence:** Ultimately, how well is the illusion working?

2.1.1. Ways of looking at Mixed Reality

The notion of combining physical and digital is not new. As the 1980s were the years of virtual reality (humans immersed inside the digital world), the 1990s presented a clear shift towards the physical world [135]. The most renowned representation of this shift in paradigm is the Milgram and Kishino's *Reality-Virtuality continuum* (which we will refer to as the *Physicality-Virtuallity continuum*). As a parts of a continuum, physical and digital can be then combined to different degrees (Figure 3).

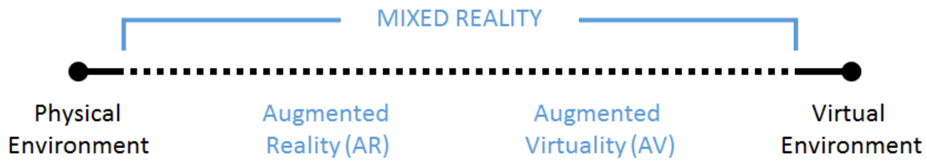


Figure 3: Milgram and Kishino (1994) [135]*: seeing the real (i.e., physical) world and virtual (i.e., digitally created) world not as opposites but instead as extremes of a continuum.*

In 1995 Rekimoto [162] studied the different ways computers can mediate our interaction with the environment, what he called HCI Styles (Figure 4). This simple description will in many ways guide this Section, as provides some valuable insights regarding spatial augmentation, tangibility and others. The vision of Rekimoto offers a richer view than the classic Reality-Virtuality continuum, which was at the time a way of considering the alternatives for optical combination of physical and digital information, mainly through the use of see-through devices.

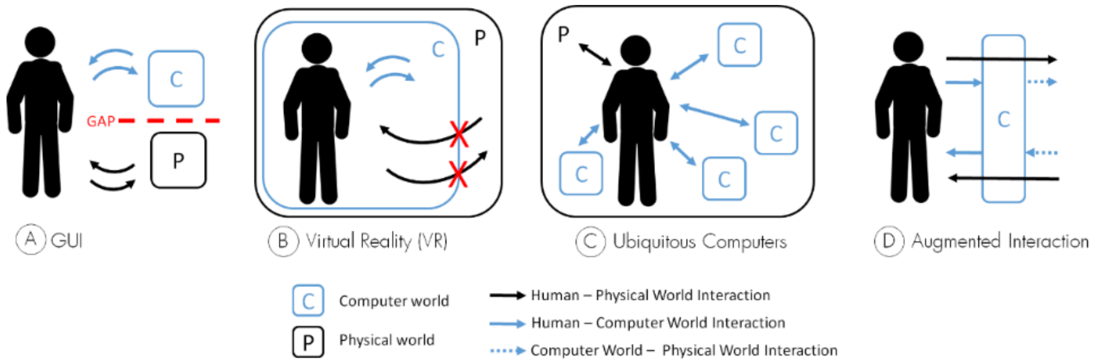


Figure 4: Rekimoto (1995) [162]*: Ways that computers mediate our interaction with the physical world, from left to right: Dissociated (GUI), immersed (VR), distributed (Ubiquitous Computing), and augmented (AR).*

2.1.2. Projector-based Spatial Augmentation

A different kind of augmentation not usually taken into account for classical classifications is spatial augmented reality: augmentation placed directly onto the environment, usually through the use of projectors. The Digital Desk [224] is the best known example of early augmentation using projectors. Even when projectors were used for augmentation since the early 1990s in the research context (and were used at the Disney Haunted Mansion as early as 1969 [137]), it was not usually explicitly considered as a way of augmentation until the end of the decade. Around 1998 several projects arose that explored spatial augmentation. Raskar presented The Office of the Future[158], which envisioned the extension of CAVEs [37] to everyday non-flat surfaces, and soon after was implemented using Spatial Augmented Reality (SAR) [159]. Initially, the focus of the interaction was on the digital information (and still required to wear stereo glasses), but then the focus moved towards the physical space [160]. Around the same time, Underkoffler and Ishii presented Urp [210], a tangible augmented workbench, and soon after Rekimoto presented augmented surfaces [163], where desktop workspaces were extended onto the environment using projection. Since then, SAR have been used countless times to support interaction with objects and the environment. With the advent of mobile projectors, it became possible to create handheld and controllable projector-based SAR [227,228].

2.1.3. Location and Technology

Among the many taxonomies and classifications of mixed reality systems, Bimber and Raskar [21] provides a technical view of displays and augmentation. His classification of augmentation studies the ***location*** where the augmentation takes place versus the ***technology*** involved, and provides a transversal view to Milgram's continuum (***degree*** of augmentation). The location is where the optical path is overridden using digital information (Figure 5), while the technology can involve projectors, opaque screens (video see-through), semi-transparent screens (optical see-through), or any other suiting technology. By looking at Figure 5, it is possible to see a strong resemblance with Rekimoto's HCI styles (Figure 4, previous page). Given the constant evolution of display technologies, our position is that, when designing the role and interaction of digital mediation, the both location and degree of digitality have a greater impact than the technology used.

As on other mixed reality taxonomies, here the focus is placed on the vision sense, as this is the most predominant sense and the easiest to mediate with technology. That said, the same taxonomies can also be used abstractly for other senses (for instance, audio and tactile displays).

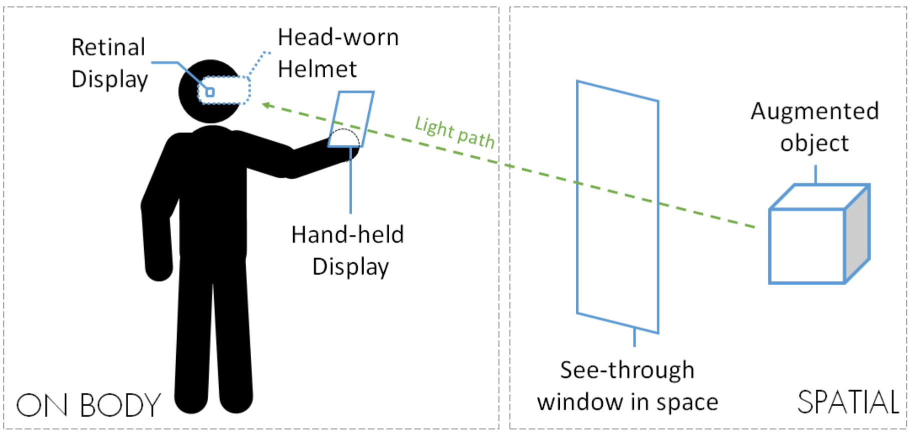


Figure 5: Raskar's (2006) [21]: location dimension of location vs technology. Note how the definition of Spatial Augmented Reality refers only to where the display is placed, independently of the technology used.

2.1.4. Artificiality and Transportation

The best way to understand the potential of hybrid MR systems is under Benford's *Artificiality* and *Transportation* dimensions [13]. Artificiality is closely related with the degree of virtuality [135], while transportation refers to the user(s) degree of immersion in either the local or a remote location (by "leaving your body behind").

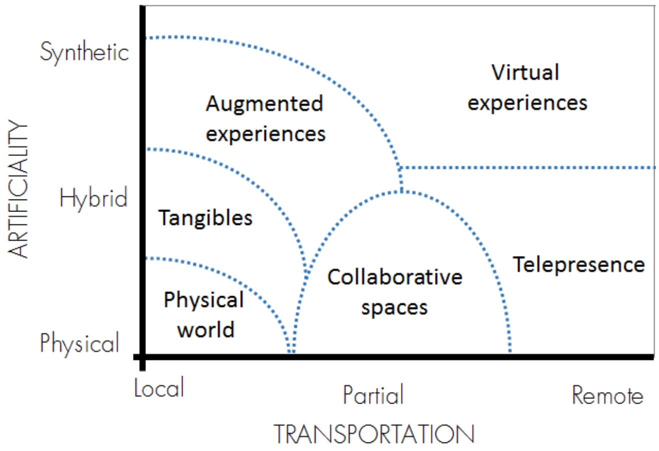


Figure 6: Benford's Artificiality and Transportation will be used when considering the space of interaction supported by hybrid mixed reality systems. The names for the regions are estimative, as the frontiers are fuzzy.

This way, a system can increase the amount of digital content while staying grounded at the user's location (augmented reality), override the senses with information belonging to a remote yet physical space (telepresence), transport the user to a digitally created space (virtual reality), or any intermediate combination.

This taxonomy is of great relevance for this dissertation, since it presents a space on which MR systems can cover not a single point, but areas instead. Through this lens, it is possible to see the impact of transitioning and complementation between MR modalities.

2.1.5. Presence and Object-Presence

Mixed and Virtual Reality applications are, at their basis, illusions: they create experiences by overriding physical information. The term ***presence*** [186] refers to the degree of *suspension of disbelief*: how much the users accept the illusions and stop questioning their veracity. In contrast with *immersion* (which refers to the amount of sensory overriding, mainly by means of hardware), presence is a subjective metric, and it is ultimately the metric we are trying to maximize. In the context of VR and tele-presence, the term *presence* refers to "*how much a given user feels there?*" (i.e., at the remote location), and has been extensively explored by researchers such as Mel Slater [186]. If there is no transportation (as in the case of AR), then the question of *"does the user feels there?"* seems ill stated. To address this issue, Stevens et al. [195] coined the term ***object-presence***, referring to how much a digital or remote element *"feels here"*.

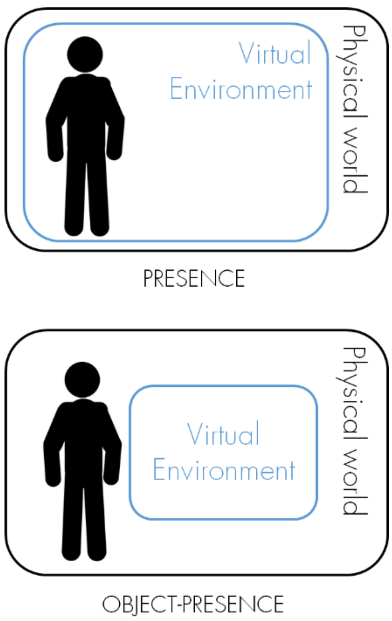


Figure 7: Presence versus Object-Presence, a distinction worth mentioning yet we will use the term presence for both cases, as it is hard to mark a clear separation of both scenarios in mixed reality systems.

Such a binary distinction between *object-presence* and *presence* can be problematic, as the previous classifications support different degrees of artificiality and transportation. For this reason, in this work we use the term presence as simply *"the degree of suspension of disbelief"*.

2.2. Physical Interaction

The interfaces presented in the previous section focus on the digital world, and the alternatives to bring that world closer to the users. To do so, most alternatives build around the limits of perception, in order to "trick" the users. This section describes an alternative approach: materializing the interfaces, in its broader sense (not just as *making it out of solid materials*, but instead *giving it physical properties and rendering it part of the physical world*). When referring to physical interaction, it is unavoidable to discuss Tangible User Interfaces (TUIs) [93]. TUIs were initially conceived as a way to provide physical handles to digital information. An early name for Tangible User Interfaces was Graspable User Interfaces [57], a very suiting name as they allow users to *grasp* digital content (both as "*holding*" and "*understanding*"). Beyond the fact TUIs allow users to manipulate digital information, they inspired us to take into account the physical properties of interaction: materials, spaces, movement and other people.

The first step towards understanding the benefits of tangible interaction is to consider Reality-Based Interaction (RBI), proposed by Jacob et al. [94]. RBI relies on 4 principles: naïve physics (predictable movements), body awareness (kinaesthetic sense), embodiment awareness and skills (spatial sense), and social awareness and skills. RBI aims to design interfaces based on how we already interact with the physical world, rendering these interfaces easier to predict and understand.

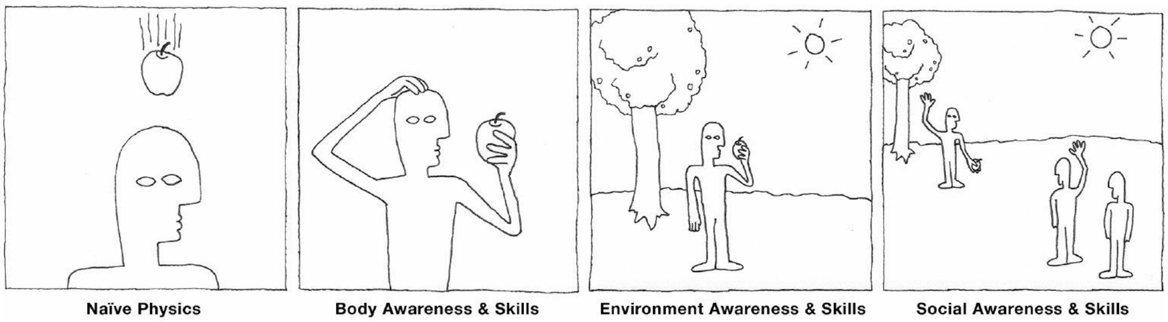


Figure 8: Reality-Based interaction uses 4 core principles to design more adapted interfaces.

2.2.1. Embodied, Extended and Distributed Cognition

One of the strongest arguments in favour of tangible interaction is that it reduces the requirement of abstraction and learning, as it provides more intuitive approaches. Here is where some controversy is created, as the use of "*natural*" and "*intuitive*" are both goals and frown upon by the HCI research community. The argument is that, *if humans are highly adaptable and are characterized by their remarkable capability to learn motor and cognitive tasks, why do we need "more intuitive" interfaces?*

Recent developments in cognition support the fact that the general approach of tangible interaction is better adapted, as they use already available, specific mental circuits (motor-spatial). As humans, we see objects based on how we can interact with them (i.e., their ***affordances***) [71], down to the neural pathways activated by the object shapes and their proximity with the user [35]. We are not just our brains consciously giving orders to our body, but a more subtle and complex system were our experiences are embedded in our body (as stated by ***embodied cognition***) [229]. In addition of being moulded by our environment, we are not separated from it. We use the external world as a support to think and store information as stated by ***extended cognition*** [199]. Even more, we are not isolated, but instead work in collaboration with other humans and artefacts to accomplish our goals, as expressed by ***distributed cognition*** [83]. We are then designed by evolution to be able to interact with objects, spaces and others; TUIs allow us to use this for our benefit. This is not simply relying on a fixed functionality, but instead taking advantage of the expressivity of the human as a whole.

2.2.2. Towards Technological Embodiment; Embodiment and Metaphor

The argument so far is that abstraction requires more cognitive resources. Taking the liberty to borrow from classical child development theory, Burner's [26] presents three levels that humans progressively obtain during their development: *enactive* (action-based), *iconic* (image-based) and *symbolic* (language based), each one with increasing abstraction. We can see similar stages in the evolution of HCI, just in the opposite order (facilitated by the technological advance). First interactions depended solely on complex languages (symbols), which turned progressively into images (icons), to lately gain a more physical-based approach (enactive-icons). The final stage of such a trend would be completely enactive artefacts, much in the line of what TUIs envision. A more formal approach to consider the evolution of technology is through Fishkin's taxonomy for TUIs [56], which takes into account two interface dimensions: *Embodiment* and *metaphor*.

- **Embodiment** is the physical distance between input and output. At one extreme input and output are distant from each other, and as they get closer to each other they trap the digital component between them, until it gets encapsulated.
- **Metaphor** is the cognitive distance, between the device and its meaning. At one extreme we have interfaces with no metaphoric connection, which neither look not are interacted in a way related with the digital counterpart. At the opposite extreme, the object does not represent the digital, the object and the digital are one and the same.

The interface definition used by Fishkin is a rather general one, involving any device that have interaction-loops following these three steps: the user performs an input action (1), the system senses the action (2) and changes its state in some form (3). For this reason, Fishkin's taxonomy presents a simple yet elegant lens to look at interfaces, since it can include traditional technologies, and also can contain other concurrent TUI taxonomies (e.g., *tools, tokens and containers* by Ullmer and Ishii [209]).

By looking at the progression of interaction devices over time, it is possible to see a tendency of new technologies towards higher degrees of both embodiment and metaphor. Much as was discussed before by analogy with Burners' stages, the main argument for the success of such a simplification is that the system were complex not because it was the ideal solution, but because it was the only mean possible. It could be argued that we are moving towards fully embodied and embedded artefacts. These new devices are called *Embodied User Interfaces* by Fishkin et al. [55].

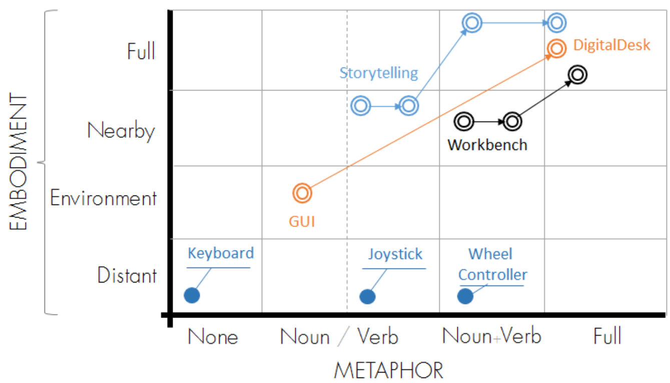


Figure 9: Fishkin's Embodiment and metaphor dimensions, as a way to explore the evolution of technology and workspaces

2.2.3. Embodied User Interfaces

Embodied User Interfaces are embodied digital artefacts framed in the physical world, where input and output are one and the same (full embodiment), and shape follows function, perhaps dynamically (full metaphor). These artefacts are not proxies supporting the communication with a computer: the computer and the device are one and the same. Notable examples of Embedded User Interfaces include: Organic User Interfaces comprised of non-flat displays [85], paper-based interfaces [1,86], wearables[75,153], and devices that can alter their physical properties (jamming-materials) [58].

2.2.4. Tangible Interaction: More than Objects

For Hornecker and Buur [87], tangible interaction is an umbrella term encompassing *"systems and interfaces relying on embodied interaction, tangible manipulation and physical representation of data, embeddedness in the real space and digitally augmenting digital spaces*". Using this definition it is possible to consider MR systems as a specific kind of tangible interaction, particularly in the case of spatial augmentation. They list 4 themes that define tangible interaction, which can work as lenses to better understand a given interface. These 4 themes are, in increasing level of abstraction:

- ***TANGIBLE MANIPULATION*** refers to the material representation that will be interacted with, and its relationship with the digital data.
- ***SPATIAL INTERACTION*** refers to the fact that interaction is framed in the physical space.
- ***EMBODIED FACILITATION*** points towards the fact that form defines how interaction happens, not only for objects but also spaces and groups of people.
- ***EXPRESSIVE REPRESENTATION*** focuses on the representation of information, and how it allows and supports its understanding.

On a coarse view, the ideas presented are reminiscent of RBI, while providing concrete tools to understand and analyse interfaces in a holistic way. This allow us to take a unified view at objects, spaces and users, from a cognitive perspective.

2.2.5. Ephemeral Interfaces

Much as other physical properties, the lifespan of the interface is also a factor to take into account. Doring et al. [46] studied the design space for ephemeral interfaces, that is, interfaces where at least one component is created to last a limited amount of time. This kind of interfaces have strong emotional impact: they not only happen at human speed (in contrast with pure digital applications), but also the timeframe for interaction is defined by their materiality. Ephemerals can involve light (as in SAR), bubbles [177], a candle [212], fog [204] (even your own breath [2]) or any other material.

Alternatively, interfaces can also be created by a temporary arrange of otherwise persistent objects, or simply artefacts that change shape in a non-permanent way [217]. The notion of ephemeral interfaces is counterintuitive at first, as in most cases they would be clearly impractical; yet, their power is not their efficiency. Even when they can be used for ad-hoc and improvised interfaces, their interest reside on their capability on reaching humans at many levels [46]: playful, artistic and emotional. As will be discussed later on (in **PART III**), artefacts that can reach users at a human level are of great interest for design in general, and positive computing in particular [30].

2.2.6. Specificity vs Flexibility

So far we only discussed the benefits of tangibility and physicality, which orbit around less abstraction leading to more intuitive interfaces. As a down side, the very feature that defines TUIs makes them limiting for cases where flexibility is required. Conversely, highly abstract systems such as desktop computers, while harder to learn, have proven to be flexible and empower the users for precision tasks. This can be seen when considering written communication and programming: even when both examples now include some degree of visual language, and extensive efforts have been placed into the creation of tangible counterparts, they are strongly built around traditional symbolic interfaces. Our position is that arguing in favour of one or the other alternative is futile: they are complementary in nature, the same way movement and language are complementary human features.

2.3. Possible Futures, Revisited

Revisiting a work mentioned in the Introduction, David Rose [168] lists the possible futures of technology as falling in one of several trajectories (Figure 10): Terminal World (screens), enchanted objects (objects augmented), animism (agents), and prosthetics/wearables (transhumanism). Each path can fulfil human aspirations on different ways.

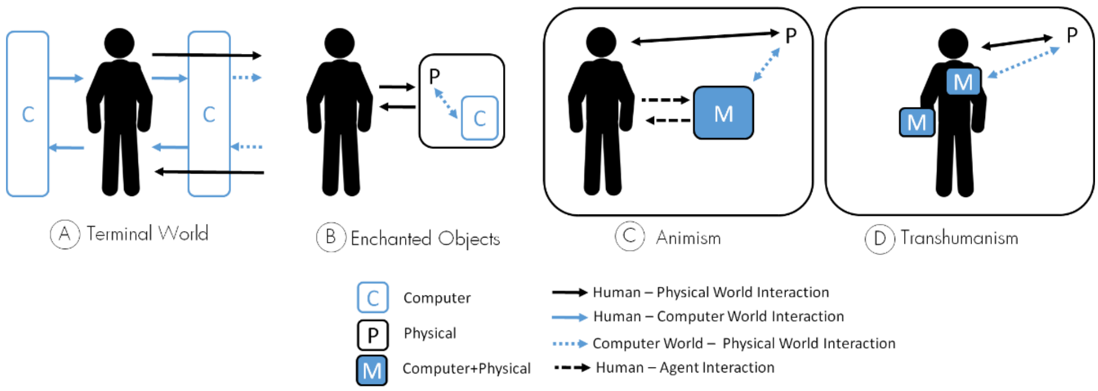


Figure 10: The possible futures by David Rose [168]*, illustrated here using Rekimoto's* [162] *iconography.*

Out of these possibilities, ***Terminal World*** as a world of screens is not much of a vision, but a reality nowadays: the digital content is ubiquitously available yet never physically here, every single screen competing to capture our attention. They address our implicit need to be informed and connected with others, at the cost of distracting us from the present moment and space. ***Transhumanism*** refers to augmentation of the human capabilities through the use of technology, as in the case of prosthetics and see-through displays. This alternative reflects our aspiration to overcome our body and mind limitations, at the risk of isolating us from each other: as users can customize their body and senses, their perception of reality can diverge.

At first sight, ***Animism*** and ***Enchanted Objects*** propose similar futures, yet they differ on the role of the artefacts themselves. ***Animism*** refers to agents and robots, artefacts that can interact with us, the environment and each other. These robots (a word that derives from the tem "*robota*" in Czech, which means *forced labour,* as in *slavery*) could help us tackle the difficult and menial everyday tasks, simplifying and improving our life quality. The challenges of designing such automaton (both virtual and mechanical) are numerous, and not only technical but also social. From the 4 visions presented by Rose, this one is perhaps the one farther away from its materialization, and as with transhumanism full of philosophical questions attached to it.

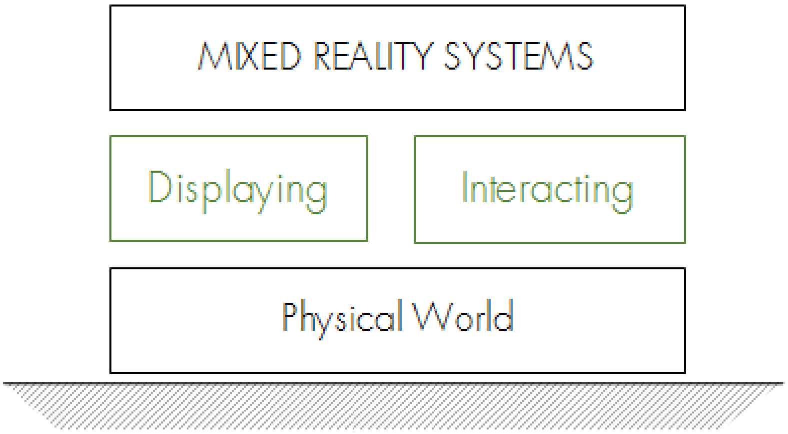
Finally*,* ***Enchanted Objects*** refer to giving special capabilities to everyday objects. This is not the same as turning specific devices into general ones (as giving calculator capabilities to a mobile phone), but instead using technology to improve the capabilities of a given device to perform its designed task (e.g., an umbrella that is aware of the weather forecast). This vision is in a way proposing hiding the digital power inside physical objects, giving us a future that looks more or less like the time before screens, yet where the computational power is still available. This vision is much in the line of what Weiser envisioned: *the era of calm computing*.

These four futures are at the same time complementary in nature as they could coexist with each other, while also competing to be the reigning mode physical and digital worlds combine. For us, enchanted objects (and their space counterparts) serve as the best way to augment our everyday lives, and when they are not enough, then the other possibilities can be used to complement them. Yet, to create these enchanted objects and spaces, we will use mixed reality as a medium.

3. Mixed Reality

A light introduction to the possible ways of constructing Mixed Reality Systems

"It's still magic even if you know how it's done."
— Terry Pratchett, A Hat Full of Sky



About this section:

This section introduces the technical knowledge required to be able to create mixed reality applications. This section focus solely on the vision-based augmentation, both from the computer perspective (based in computer vision and computer graphics), and the human perspective (the creation of visual illusions).

Even the figure gives an oversimplified representation of the problem at hand. I strongly recommend to check **Appendix A** for a more honest representation and a deeper (even when still introductory in character) explanation of the required knowledge to build MR systems.

3.1. Introduction to the topic

When creating mixed reality systems, there are several ways of combining the physical and digital information. This section briefly overviews diverse types of augmentation. This chapter excludes any technical details, and focuses only in displaying and interacting. For those readers that would like to understand the process to build MR systems from scratch, I invite them to read **Appendix A**.

3.2. Location and Technology: Intertwined

Looking back at the technology vs location classification by Raskar presented in the previous section, it is possible to see that there are roughly two categories of displays: A) augmented objects and B) windows, the latter located at any point between the user and the world.

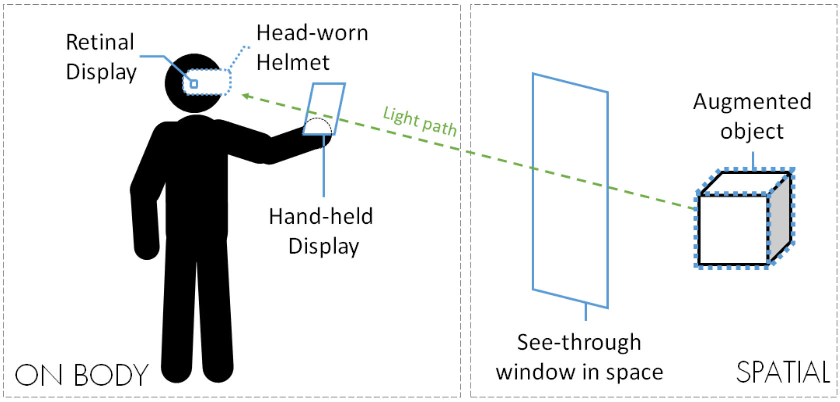


Figure 11: Raskar (2006) [21] *location dimension of location vs technology.*

Since the construction of mixed reality systems are technically challenging, the technology and the location where the illusion take place are usually studied together. The types of displays can be roughly divided in two: *see-through screens* and *augmented surfaces*.

See-through screens can provide high resolution information, potentially overriding the physical world, yet according to where they are located they can interfere with the interaction. They are implemented using either opaque screens in combination with a video feed, or using semi-transparent screens with a digital overlay. ***Hand-held devices*** are easy to control and nowadays widely available (phones, tablets), but the focus tend to move from the augmented element to the screen surface, and at least one hand is restricted to holding the device. ***Head mounted displays*** release the hands, and provide an immersive experience, but then the perceived reality of the user differs from the non-users around. ***Spatial windows*** have similar characteristics to handheld devices regarding the display capabilities, yet they free the hand for interaction. The displays are in most cases fixed, but can also be connected to an articulated arm [206,208].

Augmented surfaces: When the augmentation is place directly onto the environment, it provides a unified experience for all users. They can be implemented using several technologies. ***Projection*** allows the augmentation of the available surfaces without the need of object instrumentation, yet they are constrained by the physical geometry. ***On-object screens*** behave similarly to projection based augmentation, yet they do not suffer from occlusions at the cost of object instrumentation. ***Mid-air spatial displays*** (i.e., without a supporting surface) are currently under research, using either: a fine particle suspension as support[204], fast moving objects [96], or complex systems to release photons at arbitrary locations in mid-air [145].

When we mention displays, we refer to any arbitrary surface that can show information, disregarding the technology or the geometry involved. In practice, there are three ways of displaying:

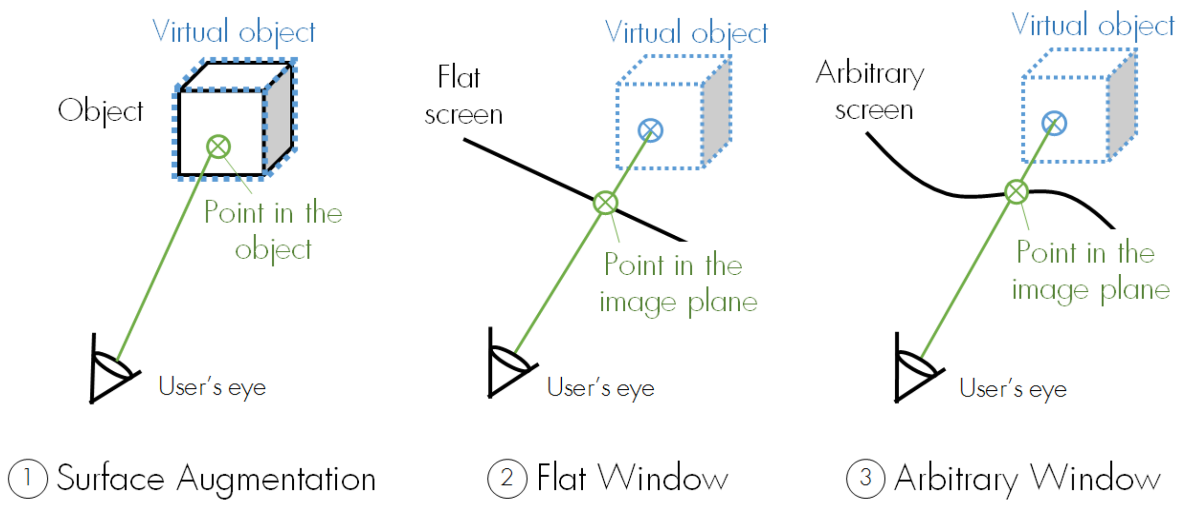


Figure 12: Raskar (2001) [157]*: Three general types of augmentation: surface augmentation, flat windows, and windows with arbitrary geometry. Note how surface augmentation is a special case of arbitrary window. Even when in his original work Raskar presented this for projector-SAR, it is also relevant for other augmentation techniques*

3.3. Surface Augmentation

The simplest spatial augmentation implies changing the appearance of physical surfaces using light. This technique requires a virtual counterpart placed at the location of the physical objects to augment.

Figure 13: Texture mapping (left) is the result of combining physical (right-top) and digital (right-bottom) counterparts with the same geometry. In this case it is performed using projectors, but it would also be possible to use screens or LEDs placed onto the augmented objects.

3.4. Windows

A window provides a view onto a scene through a frame. There are two cases: fixed-perspective and perspective-corrected windows.

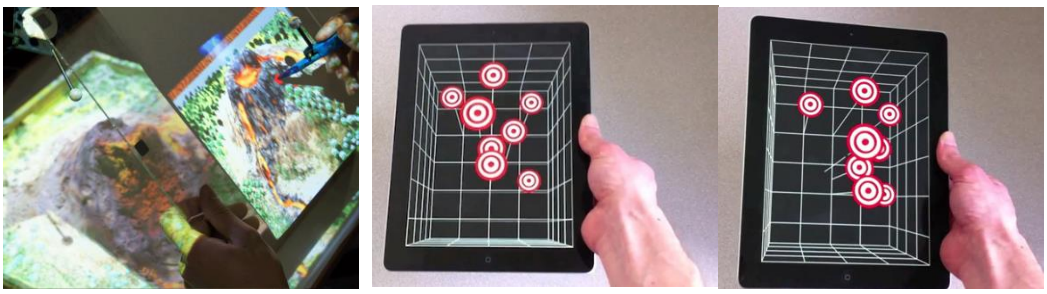


Figure 14: left: see-through augmented reality using a fixed-perspective. Right: Fish tank effect, even when the screen is flat, the perspective correction gives the illusion of depth [62].

3.5. More than Windows: Spaces

The idea behind Windows can be extended to arbitrary surfaces. A classic example of this are CAVEs, room size spaces, composed of juxtaposed flat or curved surfaces resulting on an immersive space around the users. Originally, the idea of Spatial Augmented Reality was based on the extension of CAVEs to everyday surfaces.

Figure 15: Spatial immersive augmentation: Square CAVE (left), dynamic perspective-corrected projection over an arbitrary environment (right, the box and the bluish hexagon are projected)

The easiest way to understand how arbitrary screens are achieved is by starting with CAVEs. Traditional CAVEs have a limited amount of flat walls, each of them behaving as a perspective-corrected window. A more complex surface can be seen as a juxtaposition of many smaller windows, reducing their size up to the point that each window involves a single light path.

Classical examples of this usually involve projection at room scale, yet this can be performed in both smaller[1] and bigger surfaces[2], using not only projection but any fitting technology.

[1] Projection mapping on a single grain of rice: https://vimeo.com/130165596
[2] Biggest recorded projection mapping surface is 20.000 m^2 https://vimeo.com/130039340

3.6. Projection-based Augmentation

When augmentation is performed using projection, it is necessary to know the relationship between the target surface, the projector and the user (Figure 16). For ***surface augmentation***, the result of rendering the virtual scene from the calibrated projector perspective will cause that the image pixels fall on the correct physical surface. The simplest way to implement ***windows*** is by using multiple-pass rendering: on a first pass the scene is rendered from the user's perspective, and on a second pass the user's view is applied as a texture over the supporting physical geometry (as in the case of surface augmentation). *To know more about this topic, please refer to Appendix A.*

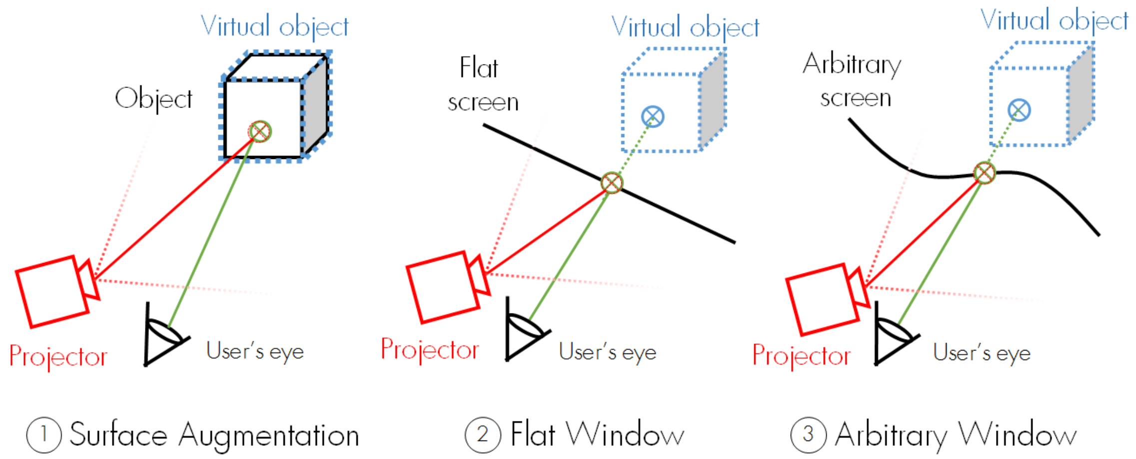


Figure 16: general types of augmentation implemented using projection. Note how surface augmentation does not require to explicitly know the head position.

Projection is a great tool, but it is not without limitations. The main issue with projection based augmentation is that, at its basis, uses a device not designed for arbitrary surfaces: the projector. Projectors are designed to be used over specific surfaces, which should be static, flat, and the right material to give optimal display quality. Under these conditions, projectors are great. The moment projection is used to create interactive applications, several issues arise:

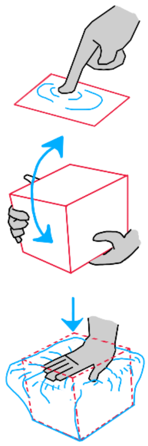
- **Calibration:** even when possible, the calibration of projectors is a tedious process; as they are blind in practice, they need to be calibrated using an external device
- **Occlusions:** in contrast with traditional screens, projectors send paths of light that can be intersected by users or other element in the environment.
- **Field of view:** projectors are able to illuminate surfaces only roughly in front of them.
- **Brightness:** Commercial projectors are sensitive to ambient light. If more than one projector is used to cover a volume, the brightness coverture will not be homogeneous, since there will be an uneven superposition between projectors.
- **Focus:** except for laser projectors, traditional LCD and DLP projectors have effective focus ranges. For static applications this is not an issue, but for interaction this becomes a problem.
- **Latency:** projectors involve mechanical components, and for this reason are slower than screens at the same price range. There are high-end projectors that do not suffer from this.
- **Overall high price:** The price of traditional screens depends, among other factors, on their size. Projectors image size can vary, while keeping the resolution and brightness constant. For small to medium surfaces, traditional screens are cheaper than projectors.

In some cases, additional hardware and software can remove or mitigate these problems; in other cases, the applications must be designed around the limitations.

3.7. Interaction Modalities

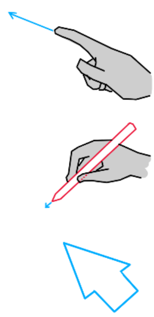
The interaction space is rich, and several alternatives are available, with complementary characteristics. Each of the factors involved in interaction has been extensively explored in the literature, so they are only briefly introduced here.

Manipulation (touch, handle and deformation): This input technique is the one most commonly used for TUIs. As previously explained, this is not reduced to simply manipulate solid bodies. The possibility to directly touch anywhere, and to do so on an expressive way, is an ongoing area of research using both on-object sensors [174]and vision [233]. Also, the interaction with non-rigid materials is currently being studied, with great advances regarding both the technology (tracking deformable objects with [237] and without [154] a rigid scan) and the application space (e.g., [1,133,205]). Augmented tools provide augmented behaviour. They can be either specific [133] ("workbench metaphor") or generic [217], and provide a range of abstraction according to the degree of embodiment. The main drawback of this technique when projection is involved is that shadows are harder to avoid when two objects are close together, but it has not prevented this technique from providing applications, such as Illuminating clay [154].



Touchless (Gaze, Body and Speech): Elepfandt et al. [53] studied the interaction space around the user, and concluded that gaze, gestures and speech are best suited for SAR (as direct manipulation causes shadows). Gestural interaction has become more popular since the release of commodity depth sensors, such as Microsoft Kinect [184]. Hand based interaction can be implemented using either sparse approaches or dense position estimation using kinetic models (Leap Motion and kinect2[185]). Such natural interaction could be incorporated easily on our everyday interaction with agents and artistic performances, but might fall short for intense or precise activities.

Pointing (direct and indirect, absolute and relative): Interacting on a higher level of abstraction can be done by controlling a cursor. Direct pointing is frequently used in VR [117] and have been used in SAR, either using tools [11,61] or fingers [165]. Relative pointing in space have been studied for flat [143] and arbitrary surfaces [68], proving to be a valid option to remote and/or precise tasks leveraging standard 2D/3D input devices. Flexible switching between absolute and relative pointing has also being studied, but only for flat surfaces [60]. Desktop-based augmented spaces such as "the office of the future" sand augmented surfaces [163] extended the "desktop metaphor" to the environment, while preserving the potential of the WIMP paradigm (windows, icons, menus, pointer).



The presented interaction techniques allow users to interact with augmentation at different levels of abstraction. The selected interaction technique will strongly depend on the task nature, and its location. One of the main concerns of our work is to allow users to switch between interaction techniques, thus allowing them to pick the most suiting technique for their needs.

3.8. Closing remarks

This Chapter presented a quick overview of the ways in which physical and digital information can be combined, and the different interaction modalities. When combined with the theoretical knowledge of Chapter 2, they allow us to create mixed reality systems. That is the focus of ***PART II***.

PART II

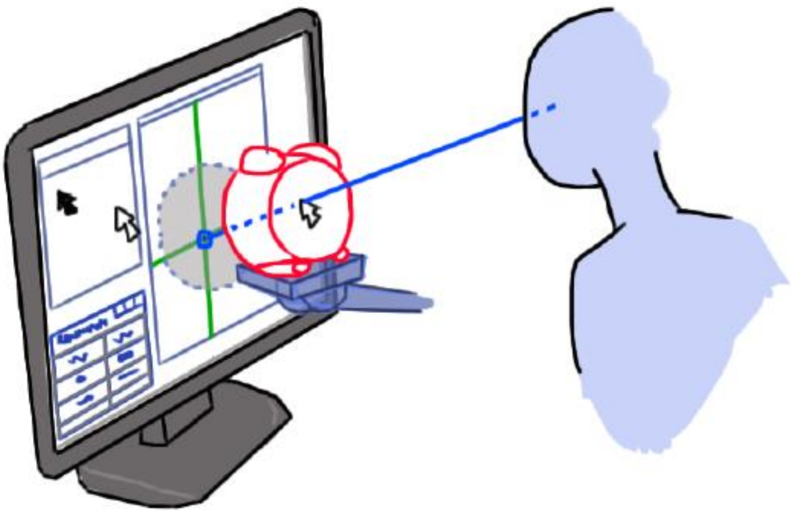
Supporting the interaction with objects, the space and others

PART II: *Supporting the interaction with objects, the space and others*

4. Tangible Viewports

Computer screens not as isolated windows, but as part of the working environment

"Reality is just a point of view"
–Philip K. Dick



About this section:

This chapter presents our first work on creating hybrid spaces. In this project we combined traditional desktop screens with tangible augmented objects, creating a single seamless space where a mouse cursor can jump onto objects placed in front of this screen, thus allowing to extend the reach of the powerful applications that live inside the computer, without redefining the interaction paradigm.

This project was published at ACM TEI'16 **[TEI16b]**, and was the first of several projects done in collaboration with ***Renaud Gervais*.**

4.1. Screens not as isolated spaces

Look at your work space right now. There is a high probability that the space is divided into two different areas: one for working digitally (computer) and one for working physically (pen and paper, books, building materials). This dichotomy has been present in our work environments for a long time, and a lot of effort of the TUI community has been directed towards a digitally enriched physical space. Compared to the traditional mouse-based paradigm of computers, tangible interaction has been shown to provide richer interaction experiences that are especially well suited for collaboration, situatedness and tangible thinking [182]. On the other hand, even when tangibility hold great promises for interaction, its use in real-world contexts remains rare, while we still use standard computers for the majority of our daily tasks involving digital information. The desktop computer is still a relevant tool to work with digital and physical matter, but we also think that its place on our desks should be rethought [158,163]. Instead of being considered as a self-contained platform that happens to be installed on a desk and its reach limited to the extent of its screen, it should be considered as a tool part of the whole toolset laid onto the desk, aware and capable of interacting with its surroundings.

We propose to leverage the potential of tangible interaction, while relying on the efficiency of standard desktop environments, in an integrated way. This is the objective of Tangible Viewports **[TEI16b]** (Figure 17), a screen-based tool enabling the use of tangible objects in a standard desktop-based workflow. For example, one can use a painting software to paint over the surface of the object as if it was part of the screen using the mouse cursor. From the viewpoint of the user, the object behaves just as a 3D model would when rendered in a viewport on the screen with the major exception that he can: (i) observe the object from a different viewpoint by moving the head, and (ii) reach out to grab the object with his hands and manipulate it freely.

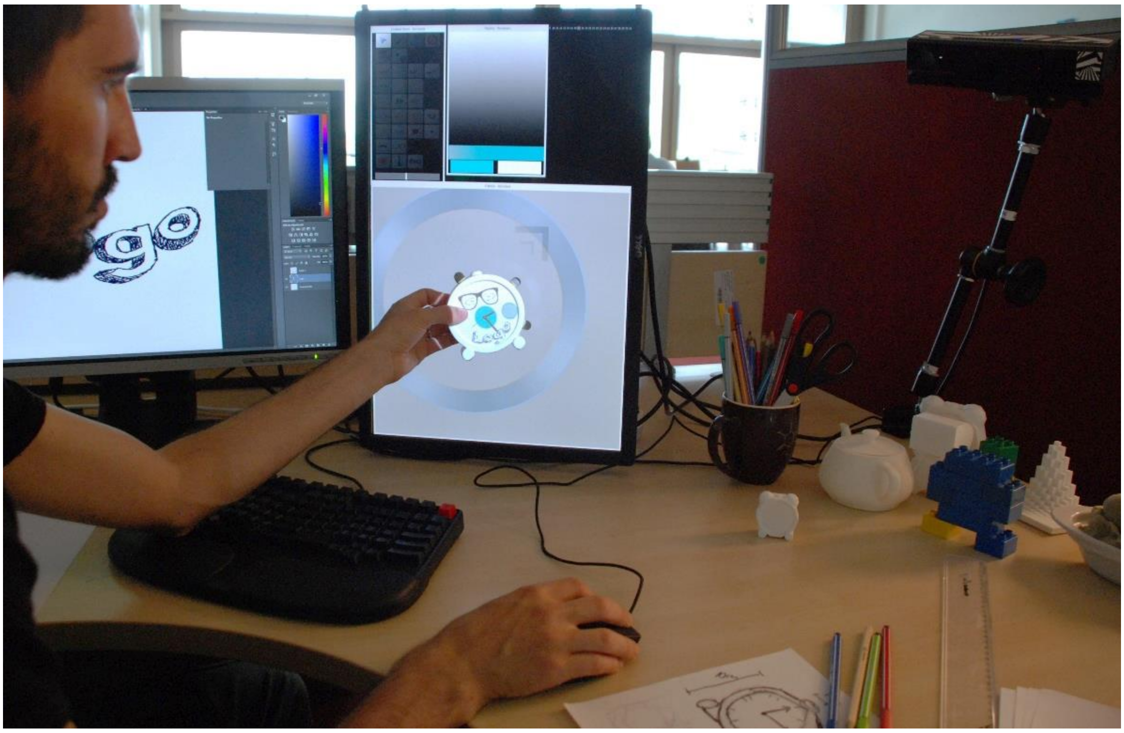


Figure 17: Tangible Viewports in action: a physical augmented object can be interacted with a mouse cursor, as long as it is in front of the computer screen. This renders the screen one more element of the working space, instead of being an isolated portal. At all times, the tangible object preserves both its tangible and augmented properties.

In this work, we emphasize the use of the desktop computer screen and its relation to augmented physical objects. This relation has been little explored as a complementary approach to tangible tools. Yet, we argue that it can be leveraged to create true hybrid applications that reduce the gap between highly flexible and expressive software, currently trapped inside a flat rectangular screen, and the intuitiveness and graspable nature of our environment.

The main contributions of this work are 1) , an on-screen window in which the mouse cursor interacts with real objects as if they were virtual 3D models, 2) a proof-of-concept prototype of an integrated workspace that combines augmented physical objects and native applications, and 3) the exploration of the interaction space of this hybrid work environment.

4.2. Specific Related Work

This project is related with several overlapping research areas, which explore the interaction around augmented objects and spaces.

Augmented and smart spaces are systems that use see-through augmented reality or projectors, often in an office environment, to enhance the workspace. The DigitalDesk [224] allowed the augmented interaction with paper over a table-top setup, using a projector to support the augmentation. The Office of the Future [158] envisioned a hybrid workspace that would combine the physical environment with a spatially augmented display system in order to create a continuous mixed-reality space. Similarly, Augmented Surfaces [163] is a system that creates interactive surfaces on a table, wall and laptop using projectors. Users could use their mouse cursor to drag information between the different surfaces. Moving towards desktop systems, Kane et al. [102] present a hybrid laptop-tabletop system that uses two pico-projectors mounted to a laptop computer to add interactive areas on the table around the device. The system is able to detect tangible objects on the table but does not augment them in any way. HoloDesk [82] is a situated see-through display where virtual and tangible objects can be manipulated directly with the users' hands, but does not integrate any traditional computer-related tasks. Closest to our work is the Skin system created by Saakes [171]. It consists of a workspace aimed at designers interested in materials for tangible objects. It uses a naive projection setup – i.e. no tracking and no automatic "mapping" of the textures on the objects – where textures, previously captured using a camera, are reprojected on tangible objects. We are instead interested in reducing the gap between desktop-based tools and the use of tangible objects, which also make the exploration of dynamic mediums possible – e.g. animations and programmed behaviours.

From an interaction point of view, the work of Lee et al. [124] is closer to ours. They present a see-through desktop environment that supports transitioning from 2D and spatial 3D interactions easily. The system allows users to see the content of the screen and their hands behind it at the same time, only focusing on virtual elements. Also close to Tangible Viewports is the work of Akaoka et al. [1], a platform for designing interactive augmented objects, using either natural interaction or a standard desktop computer. In contrast with these projects, we are instead interested in bringing interaction with physical objects to the traditional desktop workspace.

Regarding the use of a cursor, pointing in SAR [68] is related to pointing in other contexts, namely multi-display environments (MDE) and stereoscopic displays. It is possible to compare SAR to MDE in that the tangible objects that are being augmented act like multiple continuous (and potentially small) displays. Works most related to Tangible Viewports include Ubiquitous Cursor [234], which uses the geometry of the room to display the cursor when transitioning from two different screens, and Perspective Cursor [142,143], which takes into account the user's position in relation to the screen in order to give the illusion of a coherent movement from the user's viewpoint. We directly use the principles of Perspective Cursor in our work, the main difference being that we use the cursor on more complex 3D shapes instead of being limited to planar displays. Pointing on a stereoscopic display has been studied by Teather and Stuerzlinger [201] where they evaluated different cursor types in what is effectively a “2.5D”, or projected pointing task. Again, we use a similar pointing technique but use real objects instead of virtual ones. Beyond the benefit of tangibility, using SAR also does not require the users to wear any glasses or cause discomfort due to the vergence-accommodation conflict as it is the case when using stereoscopic technology.

This work contributes to the state of the art by leveraging the flexibility and power of desktop-based tools while still benefiting from tangible interaction, in a seamless manner.

4.3. Creating a Seamless Hybrid Space

The general idea of our system is to embed physical objects within the standard desktop paradigm. In our approach, we consider the screen and chosen physical objects on the desk as spatial canvases where digital information can be displayed. This design differs from other approaches (e.g.[163]) that extend the reach of the cursor to the environment, while we bring the physical objects within reach of the screen cursor.

The transition from a 2D cursor located inside the screen to a cursor following the 3D geometry of the surrounding physical environment requires a change in paradigm for the users. Instead, we embed the object in front of the screen to reduce this change of context, as illustrated in Figure 18. This design choice is supported by studies that have shown the very low performance drop for focal depth changes compared to angular movements [32,200]. In our current system configuration, normal use causes shallow depth of scene (less than 50cm – a working space created by a typically recommended distance between the user and the screen) and users are not required to rotate the head position.

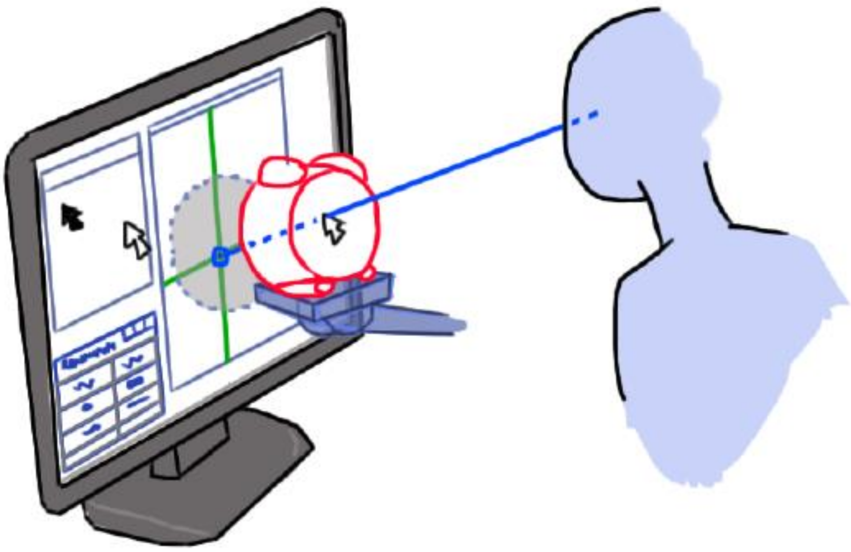


Figure 18: The user can interact with the augmented object when located in front of the screen, emulating the behaviour of a 3D viewport.

4.3.1. Spatial Augmented Reality Setup

Our SAR setup is comprised of an augmented desktop environment and physical objects that can be brought in front of the screen. The objects can be manipulated freely by the user, or they can be placed on a support for convenience. Figure 19 illustrates the setup. The projector handling the augmentation is located behind the user, and oriented so that its vertical field of view would span from the edge of the desk up to the top of the screen. It only emits light towards the physical object, so it does not perturb the visualization of the screen.

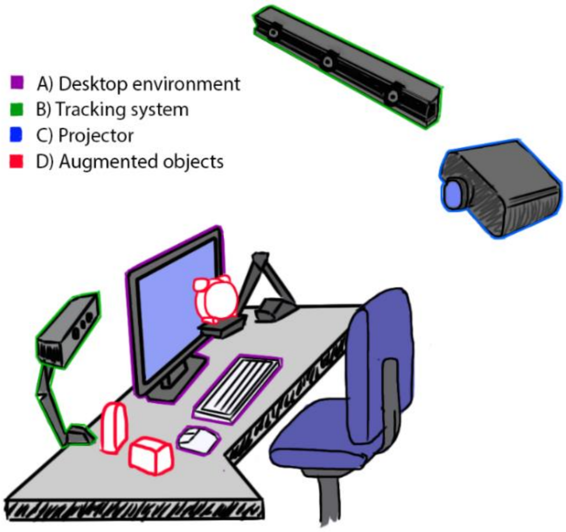


Figure 19: The SAR installation: A) The desktop environment using a standard screen and input devices, B) 6DoF tracking system (OptiTrack Trio and Microsoft Kinect v2), C) Projector and D) Physical objects that are being augmented.

The main program handling the whole installation is written with the creative coding toolkit ***vvvv*** [238] and rendered using DirectX. The video projector is a LG PF80G with a resolution of 1920 × 1080 pixels calibrated using OpenCV's camera calibration functions. The tracking of the objects is achieved using an OptiTrack Trio camera, which tracks small reflective markers. The tracking runs at 120 FPS with an overall latency of 8.3 ms and a precision of 0.8 mm. It is important that the OptiTrack system shares an overlapping field-of-view with both the Kinect and the projector. The whole implementation runs at 30 FPS (not optimized) using a 3.6 GHz Core i7 PC with Windows 8 equipped with two GeForce GTX690 graphic boards.

4.3.2. Implementation of Augmented Object

To ease the implementation of the augmented objects, we used 3D printed objects created using a MakerBot Replicator 2 in white PLA plastic with a precision of ± 0.2 mm. alternatively, it would be possible to use already existing or sculpted objects given that they would require 3D scanning before (using for instance KinectFusion [144]).

Our initial version of Tangible Viewports had the infrared markers attached to a magnetic base that could be connected to a Manfrotto articulated arm. This allows the user to comfortably position the object in 6 Degrees of Freedom (DoF) in front of the screen. We later put the markers on the objects themselves so that they could be handled more easily. Magnets were then glued directly under the objects, enabling them to still be connected to the articulated arm for longer working sessions.

4.3.3. Perspective Cursor

Our system relies on the capability to create the illusion that a physical object is entirely part of the screen space when located in front of it. In order to do so, we ensure that the cursor movements inside the working area occur in a continuous way, independently of where this cursor is displayed (screen or tangible viewport). The user thus perceive the visual space as a whole.

A window dedicated to the interaction with the object is created on the screen and its position is retrieved by using the Windows API. The screen is also tracked in world space by the OptiTrack system.

Knowing the 2D cursor position in the viewport space (i.e., a plane in space with known dimensions) allows us to infer its position in world coordinates. A virtual camera is created to reproduce the user's view of the window (and whichever augmented object located in front of it). The user's head position is obtained by Kinect v2 skeleton tracking. As soon as a physical object starts occluding the screen's cursor for the observer, a 3D cursor appears at the correct location on the object, as illustrated in Figure 20.

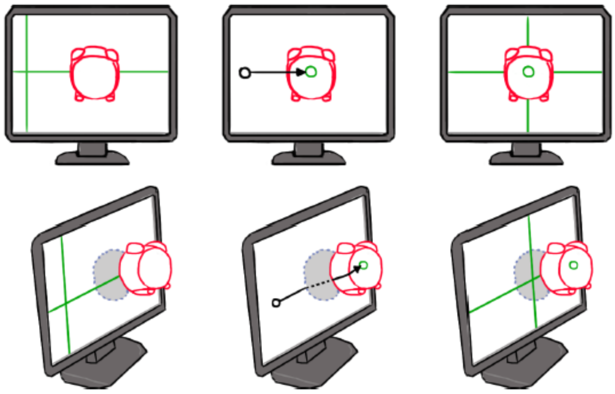


Figure 20: (Top Row) From the point of view of the user, the cursor behaves as if the object was part of a 3D viewport. (Bottom row) Side view showing the actual behaviour of the mouse cursor, "jumping" from the screen onto the object when being occluded by the object from the user's viewpoint.

This is done by ray-casting in world-space over the virtual scene from the user's viewpoint to the screen's cursor position (Figure 18). We thus obtain the 3D position and orientation on the first element on the line of sight of the user. The resulting transformation is then applied to the 3D cursor, which is displayed as a small disk aligned with the local surface's normal. This cursor is rendered as part of the virtual scene and reprojected onto the augmented object. On the screen, a horizontal and a vertical line passing through the cursor position are displayed for enforcing the link between the tangible viewport and the screen.

In the end, this technique is fully transparent to the users. Users work with Tangible Viewports as they would do with any standard application. It is also to be noted that the head position of the user only impacts the behaviour of the cursor; the cursor's appearance and the augmentations on the object are completely viewpoint independent. This is especially important for collaborative settings.

4.3.4. Direct Touch and Gestures

Beyond cursor interaction in front of the screen, direct touch on the objects is also supported. This is achieved by attaching a small reflective marker to a ring on the user's finger or on a tool (e.g. pen) so that it is detected by the OptiTrack system. We also tested the use of the Leap Motion in order to avoid instrumenting the finger of the user. However, the Leap is unable to detect direct touch and is better suited for fine gestures near the object. For coarse gestures, the hand tracking of the Kinect API is sufficient.

* * * * *

4.4. Interaction Space

In our hybrid workspace, interaction can either take place on the screen, on the augmented object, or on both display supports at the same time. In the following sections, we explore the interaction space by describing examples of techniques that we developed for each of these categories (Table 1).

		MODALITY		
		Mouse and Keyboard	**Hybrid**	**Touch and Gesture**
LOCATION	**Screen**	Widgets, native applications, programming	–	Touch screen based interfaces*
	Hybrid	Drag and drop, hybrid widgets	Pick and drop, object annotations, data visualization	Gestural control of virtual version
	Object	Pointing on objects	Bimanual interaction	Navigation, tangible design

* Out of the scope of this work

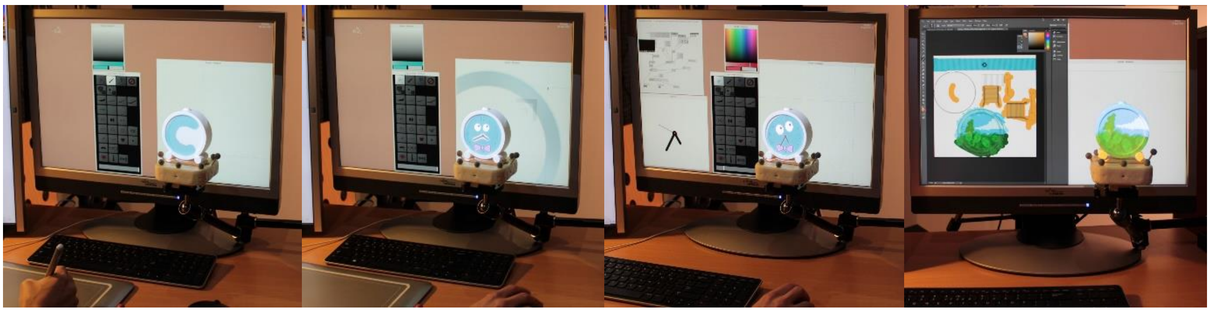
Table 1: Interaction possibilities supported by combining screens and objects

4.4.1. Interacting with the Screen

Because our objective was to conceive a system that benefits from the advantages of standard desktops, all the usual techniques designed for such environments can directly be used (Figure 21).

Widgets: We have designed a custom application based on such standard widgets for modifying the appearance of an augmented object. For example, selecting the background colour of an augmented object can be done directly by way of a colour palette. This application served as a basis for the evaluation of the system that presented later in this Chapter.

Native Applications: It is also possible to use native professional applications. As an example, we linked the output of Adobe Photoshop, a software that is ubiquitously used in the design and artistic industries, to our system. Hence, we leverage the skills that professionals already acquired with these tools. The most straightforward use is UV painting which consists of adding graphics on a 3D model. It is a task that can be done either in a 2D painting environments using a UV layout or directly on a 3D view of the object. Both can be achieved using Photoshop. We retrieve the texture that is being painted in real-time and update the augmented object accordingly. Every time an operation is performed on the design, the physical object's appearance also gets updated. This can be especially useful in object design, where the final result is not a 3D render but an actual object.



(A) Direct painting (B) Virtual elements (C) Creative coding (D) UV painting using Adobe Photoshop

Figure 21: (Different features to modify the appearance and behaviour of the physical object

Programming: In addition to the connection of existing tools, we also included Creative Coding capabilities. In practice, creative coding is often comprised of programming toolkits that are focused on visual results and short feedback loops. For these reasons, it is often used for prototyping. Examples of such toolkits include Processing, *OpenFrameworks* and *vvvv*. These approaches make it possible to envision a near future where physical objects are comprised of thin and flexible screens [84]sand with which users could tinker with their augmented content (dynamic appearances and behaviours). Programming is an activity that is almost exclusively conveyed on standard computers. It is possible, then, to create a program and visualize its execution in real-time on a tangible object. As an example, we created a simple program where the appearance of a clock evolves with the time. The results of this program can be visualized directly on an augmented physical clock (Figure 21).

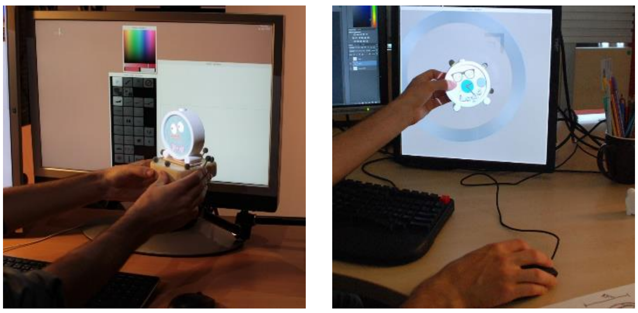
4.4.2. Interacting with the Physical Object

This section presents the interaction techniques we have implemented to support the use of physical objects: direct interaction, pointing on object using the tangible viewport window and bimanual interaction.

Direct Interaction: Working with physical objects has the benefit of enabling manipulation directly with our bare hands. No 2D to 3D mapping operations are required to create a desired point of view as is required in desktop 3D applications. Also, since the augmentation occurs on the surface of the object, changing the viewpoint can simply be achieved my moving the head. The user can thus observe the object in a natural way, which radically differs from what he or she is used to do with a virtual version of models displayed on flat screens. Also, direct touch can be used whenever precision or specific tools are not required. For example, when creating interactive objects, one can use interactors or trigger behaviours directly, similar to [144].

Pointing on Objects: In addition to direct manipulation of the tangible objects, our system supports cursor-based indirect interaction for completing interaction tasks directly onto the physical objects. These tasks can be pointing, drawing, selecting or moving virtual elements. Compared to an approach where the user would interact directly on the physical object, indirect interaction offers several complementary advantages. It does not require specific input devices, it is fast and accurate, less prone to fatigue, and it integrates within the desktop workflow.

Bimanual Interaction: Handling the physical object and using the mouse can be achieved at the same time following a bimanual interaction approach [10]. The hand holding the object plays the role of reference frame and assists the dominant hand which is dedicated to fine mouse movements. This approach leverages the precision and stability of 2D pointing and the easiness of 6 degree of freedom manipulations of 3D objects.



(A) Direct interaction *(B) Bimanual operation*

Figure 22: Different features to modify the appearance and behaviour of the physical object

4.4.3. Hybrid Screen/Object Interaction

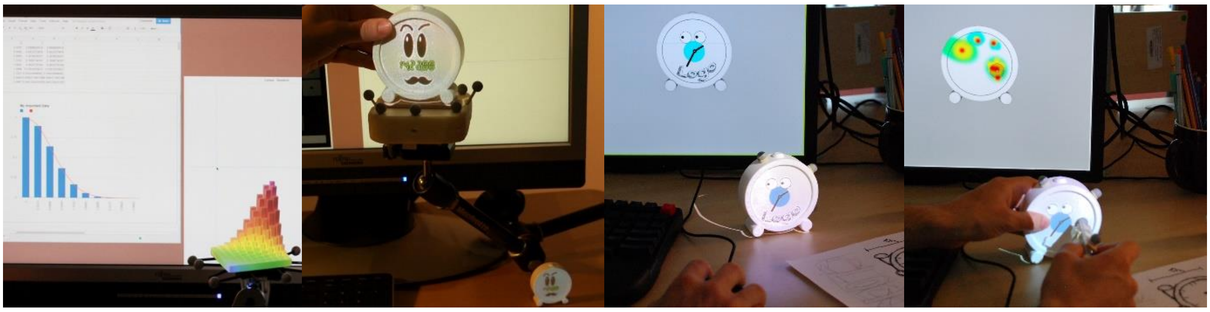
Both the physical objects and the screen are part of the same working space. Consequently, it is possible to directly link operations on the screen with actions on the physical objects. The converse is also true. We present application examples that use both object and screen simultaneously.

Drag and Drop: Since the viewport creates a seamless continuum between the screen and the object, drag and drop operations can be used with the mouse cursor. This operation would not be possible using touch and would have to be replaced by pick and drop.

Hybrid Widgets: The standard approach for applying transformations (e.g. scaling and rotation) to visual elements displayed on a screen is to use widgets centred on these elements. The problem with standard SAR setups is that, although technically possible [17], it is very difficult to create the illusion of floating visual elements around the object as soon as no material can support the display. We designed hybrid widgets that are operated on screen. We reproject the position of the selected element on the screen based on the user's viewpoint and we place 2D widgets centred on this location. When moving the physical object, the position of the widgets is updated accordingly on the screen. These transformation widgets that allow the rotation and scaling of the selected element are illustrated in Figure 23. They are relatively big and they do not touch directly the physical objects. This design choice has been made to avoid problems of eye accommodation between the depth of the object and the depth of the screen. Hence, after selecting an object to modify, users can quickly grab and manipulate the widgets, without eye fatigue.

Object Annotation: Another opportunity offered by the fact that a screen stands behind the physical object concern the display and the entry of text. Indeed, these operations may be difficult to complete in many traditional SAR setups. In our case, it is easy to annotate a physical object by selecting an anchor point (either with the mouse or direct touch) and typing a related note being displayed on the screen, with the keyboard. Inversely, one can select a note on the screen, and see the corresponding area directly on the physical object. This creates a text box positioned in an empty zone of the screen which is linked to the projected position on the screen of the anchor point.

Physical Data Visualization: Beyond annotations that can be beneficial for many fields (e.g. inspection of manufactured objects), we have explored the use of a hybrid approach in the scope of data visualization. Data visualization (and especially 3D data visualization) has been shown to gain from a physical representation [95]. Using the tangible viewport window, it is possible to add interactivity to physical visualization. In particular, to query more information on some aspects of the visualization, one can just point at the area of interest to see related data on the screen, or she or he can select an entry on the screen to see the corresponding elements on the physical visualization (Figure 23-A).



(A) Linked data representations *(B) Synchronized objects* *(C) Synchronized views* *(D) Manipulation statistics*

Figure 23: Different synchronization modes between virtual (on-screen) and augmented object

4.4.4. Synchronized Views

We also explored the synchronization between a virtual version of an object displayed on screen and a physical one. When the tangible object is not in front of the screen, the tangible viewport window displays a virtual version of the augmented object (Figure 23-C). Modifying the virtual version updates the tangible version in real-time, and vice versa.

Being able to have two representations of an augmented object, one on screen and one physical opens possibilities, particularly in the context of collaboration. For example, it would be possible to expose the view of a user handling the physical object or providing advanced visualizations such as a heat map of touched areas (Figure 23-D). Also, multiple users can have their own duplicated augmented object (Figure 23-C). These users can be working either locally or remotely.

The synchronization between real and virtual can be paused, for example using a gesture (e.g. pulling the object rapidly away from the window), to compare multiple versions. The inverse action (e.g., bringing back the physical object in front of the window) could then merge the two versions on the physical canvas.

4.5.4 Illustrative scenario

To illustrate the use of Tangible Viewports, we describe here an object design scenario where an artist is experimenting with new visual design ideas for a product[3].

She can start by sketching first ideas on a paper, and then use a modelling tool to create a 3D sketch. Equipped with a 3D printer, she can print one (or many) physical objects to have in front of her. She can first directly paint on the object using the mouse cursor. Then, she can use a digital painting application such as Adobe Photoshop or a vector graphics editor like Inkscape to draft a logo on her computer. Then, using the mouse, seamlessly drag the logo from the editor directly to the physical prototype she just printed. The prototype can be physically manipulated to review the appearance. Modification to the design on the desktop computer will be automatically reflected in real-time onto the object. She can scale and rotate the logo directly on the physical object to see directly the impact of her modifications. This way, the feedback loop between the design activities (which require specialized software) and the validation of the effect it has in physical form can be greatly reduced. If required, new versions of physical objects can iteratively be 3D-printed, as we currently do with 2D printers when working on 2D documents. By making the interaction with the physical objects coherent with the traditional way of manipulating 3D information on a desktop computer, it is possible to leverage the experience of users with their professional tools, while at the same time adding the richness of tangibility and physical visualization.

The link to the desktop environment can also foster the use of a more dynamic medium, using animation or programming, directly on real objects. In this scenario, we can also imagine one or several collaborators participating to the design choices. These collaborators can directly observe and manipulate the augmented object, and ask the main designer to update the design in real time. This kind of social collaboration is harder to obtain with traditional design tools.

[3] For a visual version of the scenario, the reader can refer to the video teaser: https://vimeo.com/142358002

4.6. 5 User Feedback and Discussion

We conducted an exploratory study where we asked participants to manipulate a preliminary version of the system, as well as a non-tangible version of the tool. The objective of this study was to assess how physical objects integrate within a standard screen space. We have designed a simple custom creation tool (see Figure 21 and Figure 22) for this purpose. Fourteen participants (9 males and 5 females, mean age 25.6±3.7) took part in this study. Half of the participants started the experiment with Tangible Viewport, then they moved to the non-tangible one, and half did the opposite. In both cases, participants were introduced to the main features of the tool, and the experimenter explained what was expected from them. Participants were asked to create a personal visual design of a clock. The only difference between the two versions of the tool is that, in the tangible version, the results of the creation was directly displayed on a 3D printed clock, whereas the virtual representation of the model was used in the standard viewport version. For changing the view on the object, subjects had to either manipulate the object and/or move the head “naturally” with Tangible Viewport, whereas they were using a trackball metaphor operated with the mouse middle-button in the standard viewport version, as commonly done in standard desktop 3D tools.

Subjects were asked to follow a tutorial for customizing their clock (Figure 6), which included: 1) choosing a background colour and painting the front face, 2) adding virtual elements and resize/rotate them, and 3) making a drawing on the side and back of the object. This scenario was designed to ensure that the main features of the tool were used under different conditions. For example, Step 3 tests the ability of the participants to draw freely on curved surfaces.

After the experiment, participants were asked to answer two questionnaires: the User Experience Questionnaire [122] and a custom questionnaire aiming at obtaining user feedback about the usability of the tested systems (5 points Likert scale) and their preferences between the two. Both questionnaires showed no significant difference between the two versions of the system. They were also invited to leave comments and feedback about what they liked and disliked about each version of the system. Overall, the majority of the participants preferred manipulating the tangible version (12 out of 14) and were more satisfied with the final result (11/14). No participant mentioned difficulties moving from the screen to the object. These results seem to indicate that the tangible viewport metaphor works well, and it is comfortable to use.

Regarding the comments, among the most appreciated features spontaneously cited by the subjects is the ability to work with a real object (9/14) and to have a physical view on the final product (6/14). For example S1 liked that “you can see the real object with the elements you draw. That way, you can observe the final product before it is produced”. S9 mentioned that “The creation feels much less virtual” and that “going from the screen to the object is fun”. A few participants also insisted that they liked to be able to manipulate the object with their hands (5/14), while others found the magnetic base uncomfortable (5/14) – which is why we later replaced the base and attached the markers directly onto the objects. Complaints were made (5/14) regarding the fact that the editor lacked important features such as “undo” and “magic wand”. This highlights that the interaction between the screen and the object was working well enough that the main topic was the painting features.

Regarding the technical solution, several participants (6/14) mentioned that the augmentation calibration was not precise enough, which could be improved by using more advanced known solutions such as the one used by Jones et al.[98]. They also explicitly mentioned some delays and robustness issues on the head tracking (5/14). The second iteration of the system corrected these issues by replacing face tracking by skeleton based head tracking and better Kinect positioning. Regarding the cursor, some participants (4/14) did not like the fact that changing the head position was moving the cursor on the object, a side effect of using perspective cursor. This issue could be addressed with a system that would prevent the cursor on the object to move when the head position of the user changes and instead correct the on-screen cursor's position when it reaches the edge of the object's silhouette, from the user's point of view. Such alternatives will be studied on the future.

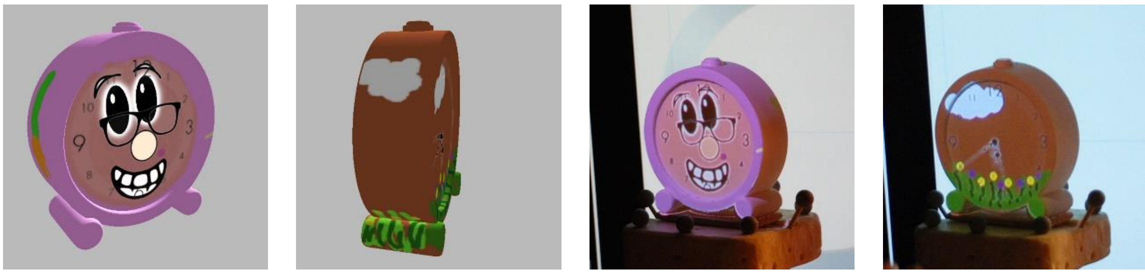


Figure 24: Examples of participants' creations using Tangible Viewports.

4.7. Conclusion

This Chapter we introduced Tangible Viewports and we described an effective implementation of this concept. A preliminary study showed that the overall usability of this system is good. This concept does not aim at replacing existing systems. Indeed, we have shown that, from a technical and user point of view, the seamless integration of physical and virtual tools is not just feasible but enriches both. The resulting system allows users to choose the interaction modality that better suits the task at hand, instead of being constrained by isolated counterparts.

One of the current limitations of a tangible approach is the rigidity of the physical elements, which cannot (yet) be reshaped in real time. Our vision is that 3D printing will become as efficient as 2D printing in a near future. Alternatives to our current approach involve free-form sculpting[131]power, to perform iterative modifications of the printed geometry [202] or the usage of jamming materials [58], which in term could be complemented with semi-transparent screens [220]. In combination, the literature plus the proposed approach could allow users to rapidly iterate designs with different degrees of physical components and digital prototyping information.

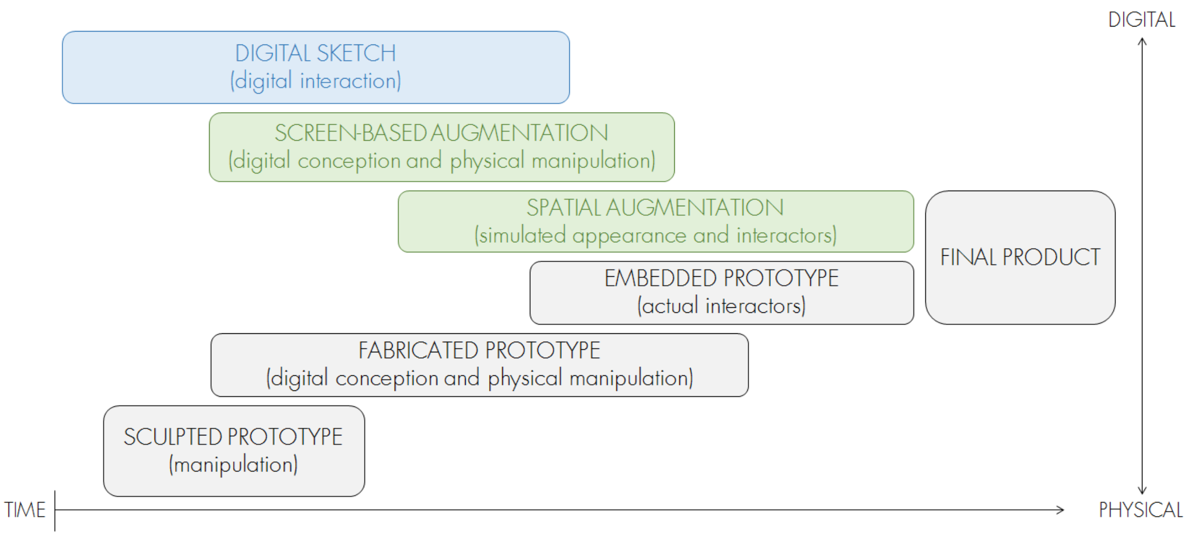


Figure 25: The convergence between physical and digital versions towards a final product.

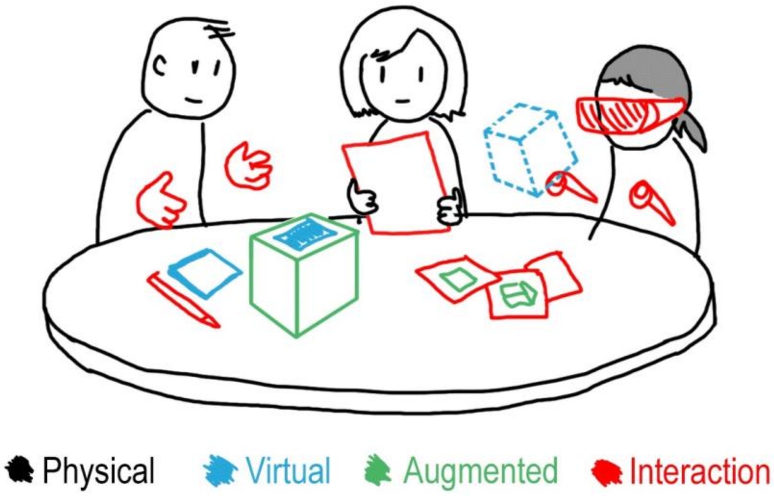
PART II: Supporting the interaction with objects, the space and others

5. One Reality

Navigating across the physical-virtual continuum

"One of the things our grandchildren will find quaintest about us is that we distinguish the digital from the real, the virtual from the real. In the future, that will become literally impossible."

- William Gibson



About this section:

This chapter presents the works on the combination of multiple Mixed Reality technologies in order to support the navigation across the physical-virtual continuum. The conceptual framework and the implementation details allow us to implement the remaining contributions presented in this manuscript.

The works described were first published as a work-in-progress at IEEE 3DUI'17 [**3DUI17**] (best tech-note award), and soon after at ACM UIST'17 [**UIST17**].

5.1. Introduction

So far this manuscript presented ways to combine physical and digital, reaching towards systems that benefit from both. We discussed the fact that conceptual modalities of mixed reality need to be backed with actual implementation, and these implementations are not without limitations. Our approach is not to settle with a given technology, but instead to find ways to support multiple technologies in unified ecosystems. A first step towards that goal was presented in the previous Chapter with ***Tangible Viewports***. We wanted to go further, by supporting all existing interaction modalities in a single unified space; as a result, we created One Reality, presented in this Chapter.

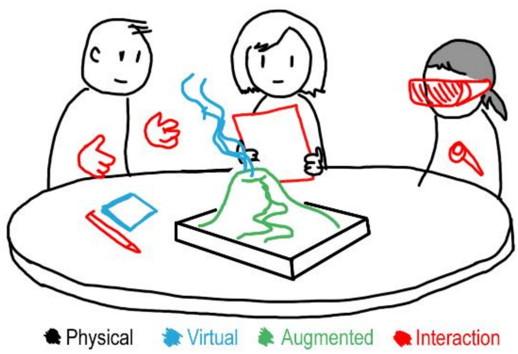
5.1.1. One Reality Contribution

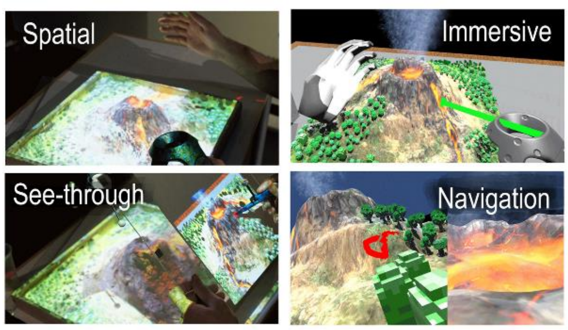
We propose a conceptual framework and its implementation where the physical world that stands in front of the users can be progressively augmented and distorted, with the final goal of augmenting their perception. By providing augmented experiences, anchored in the physical reality, we seek for a hybrid space where users benefit from both the force of the physical sensing, and the flexibility of digital interaction. At each conceptual step, we also associate a technical implementation, based on off the shelf technology.

Our contributions in this work are: 1) the design of a conceptual framework, where the user can smoothly travel between the physical and the virtual worlds, 2) the implementation of this framework on a unified system, and 3) the exploration of the interaction possibilities. The result builds towards a unified way to look at the intersection between physical and digital realms, combined on a single reality for the user.

5.1.2. Examples

To better understand this concept, let's consider two concrete scenarios our approach can address: collaborative learning and hardware maintenance.





Collaborative learning: Imagine a group of students working around an augmented volcano mock-up. They can discuss and observe the simulated behaviour, while physically touching and moving around and maybe dissecting the mock-up. Some information can be difficult to understand from an egocentric perspective, so they can travel inside the mock-up for more information, using VR. In this case, they can "jump" into the physical model, follow the tubes, and experience the volcano from a different point of view. While immersed, they can still discuss with her classmates, their bodies visible as giants around the mock-up. This simple example, inspired form TV shows like "Cosmos"[172], "the Magic School bus [34] or "Once upon the time... Life" [12], can be generalized to any pedagogical content, where the changing of viewpoint of the physical environment may improve the learning and understanding of the studied phenomena. Using the shared physical space as starting point fosters discussion.

Maintenance task for a car engine: the engineer is in a process where she has to work physically with the object standing in front of her. Thanks to situated projection, she can benefit from digital support that will guide her during the process (e.g. highlight a given piece) as done in [236]. Now, the engineer needs to observe the engine from the back, or have a closer view of a specific part. Because she cannot manipulate or move around the physical engine, she decides to change her point of view virtually. To do so, she puts on an HMD, and virtually navigates in and around the object of interest. She can also observe a virtual engineer performing the required task, and take the viewpoint of this expert. Numerical simulation can be launched too, in order to observe for instance the flow in and around the engine. These situated interaction, which would not be possible with standard SAR approaches, may help the engineer to better understand the physical engine she is working on.

Before exploring in detail the different levels present in our framework and their implementation, the next section briefly reviews the previous work that enabled their conception.

5.2. Specific Related Work

The ideas presented in this work build upon the vision that the digital realm can be integrated into the physical one (extensively discussed so far), and the different technologies that were used in the past to explore this possibility.

In order to take advantage of the complementary characteristics of mixed reality technologies, hybrid systems have been studied. For instance, see-through displays and SAR have been combined, notably in order to complement the HMD's high resolution with the projectors' scalable FOV (field of view) [16] Transitioning between see-through AR and VR have been also explored by Kiyokawa et al. [112]. In the context of multi-display environments, the combination of screens and projection has been studied, both with [28] and without see-through technologies **[TEI16b]**. Dedual et al. [40] proposed a system that combines interactive tabletops with video see-through AR. *Smarter Objects* [81] and *exTouch* [106] use video ST-AR to control embedded systems; even when the physical artefact was the focus of attention, no spatial augmentation was presented, except the electronic behaviour itself.

The best way to understand the potential of hybrid MR systems is under Benford's Artificiality and Transportation[13]. Benford’s taxonomy is of great use because MR systems can cover not only a single point of the artificiality-transportation space, but also areas (Figure 26). MagicBook [20] is a physical book that supports different degrees of artificiality and allowing the transition from an egocentric viewpoint (no transportation) to a location inside the book. Similarly, Kiyokawa extensively studied collaborative MR systems [112–114], were the focus could alternate between the local scene to a remote location, and this remote scene could be obtained through scanning, virtually created, or a combination of both. Rekimoto studied dynamically creating information locally [161], and more recently explored the potential of taking different perspectives of the physical world [105,107,109,115] and asymmetric perspectives of a virtual scene [89].

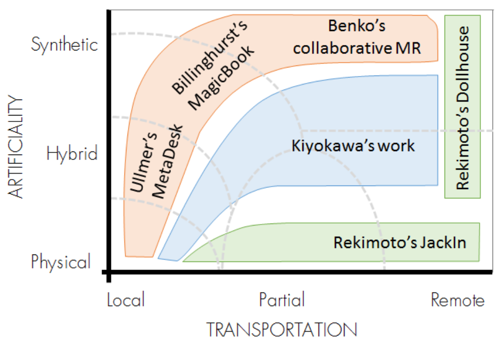


Figure 26: Some examples of hybrid Mixed Reality systems as seen through the lens of Benford’s taxonomy (artificiality vs transportation)

To summarize, the research involving the combination of physical and digital is vast, either focusing on a specific technology or combining complementary alternatives. In this line of research, we propose a smooth transition through progressive immersion, as described in the next section.

5.3. Conceptual Framework, Implementation and Interaction

This work focuses on reducing the gap between physical and digital, without limiting the interaction to neither of them. We propose to provide the users the liberty to increment the degree of immersion when needed, without losing contact with the physical world.

In order to explore this incremental immersion, we created a hybrid working space that supports multiple mixed reality modalities simultaneously, and enables the user to free transition between them (Figure 27). In it, the environment and the users are scanned and tracked in real time, enabling their augmentation, movement logging and digital reproduction. To interact with the system, physical tools are rendered available even in virtual spaces, and vice versa.

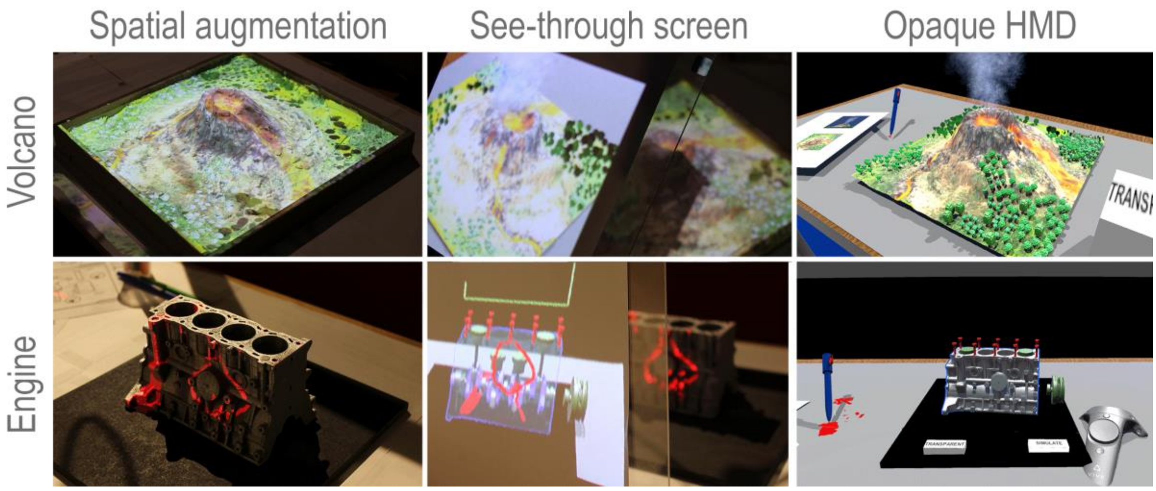


Figure 27: Example scenes: volcano mock-up made out of sand (top), 3D printed Toyota engine (bottom). Each scene can be interacted with different display technologies: spatial augmentation (left), see-through displays (middle), and opaque head mounted displays (right).

When designing the progressive transition between the physical and digital realms, we considered 6 incremental levels where digital characteristics are progressively included (Figure 28). The following subsections describe, in an incremental way, the features of each of the levels in combination with their implementation and interaction consequences.

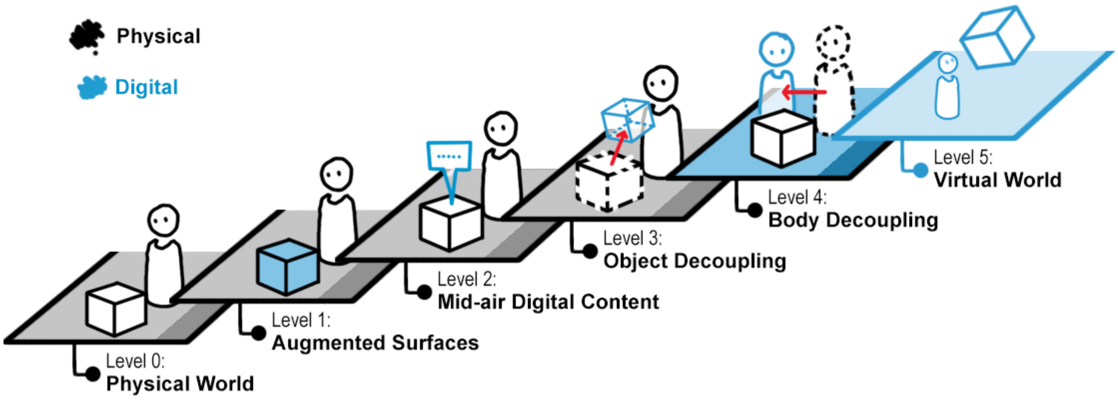
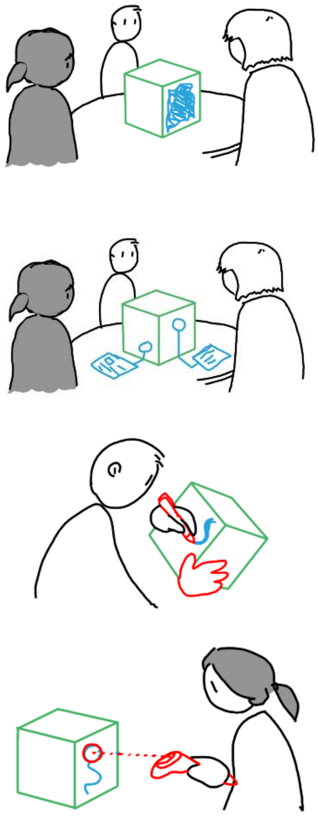


Figure 28: Conceptual framework with 4 levels of incremental augmentation between Physical and Virtual worlds.

5.3.1. Level Zero: Physical World

The starting point of interaction is the physical world, where participants can interact with each other and with physical artefacts. We consider this level is comprised of both natural elements and technological devices that alter physical properties in a non-cosmetic and persistent way (i.e., the modifications stay even when the artefact is not working, for example mechanical artefacts or pens). In the physical world, objects can be manipulated when they are not too heavy and not too big, and people can move their heads and bodies to change their point of view, as long as nothing is blocking them.

5.3.2. Level One: Augmented Surfaces



At the first level, digital information can be placed at the surface of physical objects in the user's environment, as in the classical work of Rekimoto et al. [163]. This can be used to display complementary information (e.g., show text, change the object perceived texture or show annotations on its surface), or arbitrary information (e.g., using the surface as a screen to render a digital 3D scene).

We implemented the augmentation using Spatial Augmented Reality, which allows to place the digital information directly onto physical surfaces. When the surface of a given object is not suited to support the digital information, the content is then leaked to adjacent surfaces, such as tables or walls. With SAR, digital information is equally available for all the users and viewpoint-independent, as it is the case with physical information. By tracking objects of interest in the scene, the system supports various interaction modalities (Figure 29). It is possible to directly manipulate the augmented objects, and to use both direct and indirect interaction, using pens or pointers. Given the physically nature of the display supports, non-augmented tools can be used naturally.

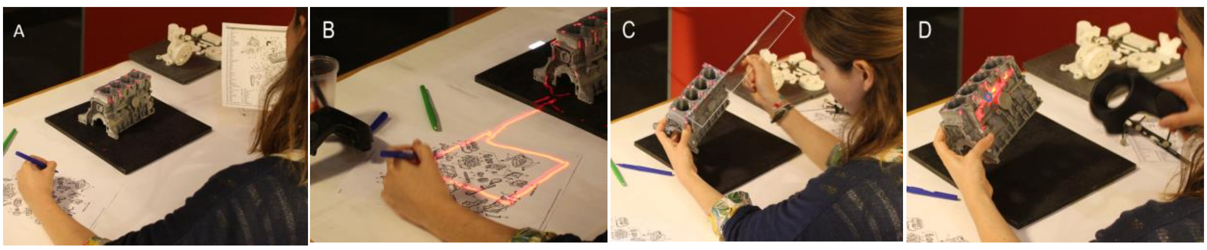
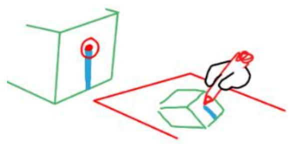
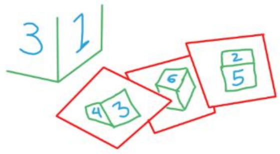


Figure 29: Using spatial augmentation in combination with object tracking enables: to use augmented pens to draw directly onto artefacts (A) and their surroundings (B), to manipulate the augmented artefacts to use in combination with traditional tools such as rulers (C), or the use of indirect interaction (D).

5.3.3. Paper-based interfaces

A widely distributed support of information is paper. When paper is tracked, any digital information displayed on it will follow accordingly. The content of augmented paper can be created using augmented pens; it is also possible to use augmented paper as support for digitally created content -- a paper window [1] --, bringing computer screens closer to the physical environment[123]. An additional benefit of using projection is that normal paper can be placed over the augmented counterpart, much like tracing paper. Among the possible digital content, paper windows can display a 3D render of the augmented scene, which can be interacted via ray-casting like traditional displays. As a result, users can interact with their surroundings from a given point of view (e.g., a colleague's viewpoint) (Figure 30). We call these windows *"interactive pictures".*





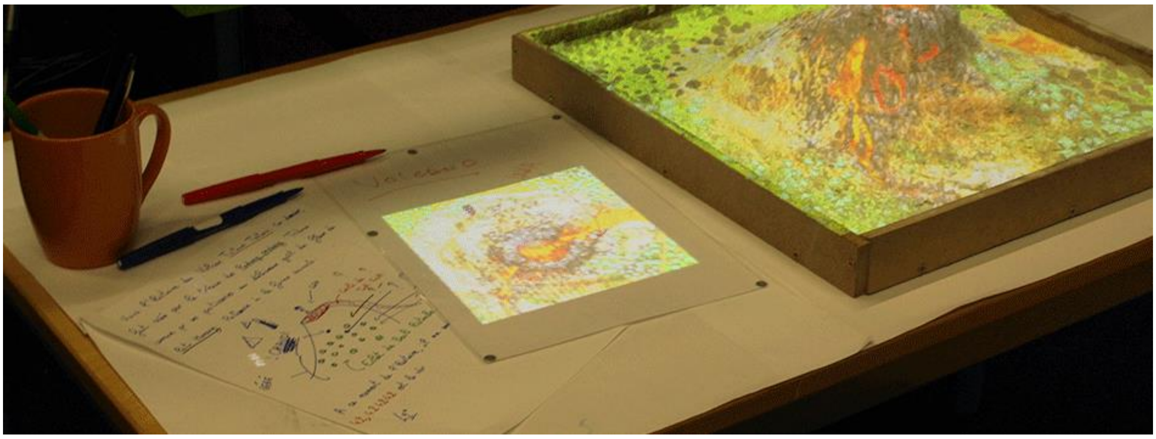
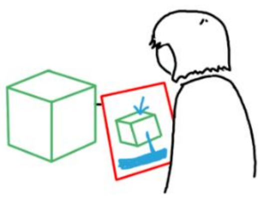


Figure 30: Interactive pictures can display and support interaction with the augmented scene from a different point of view. They can be used in combination with traditional pen and paper.

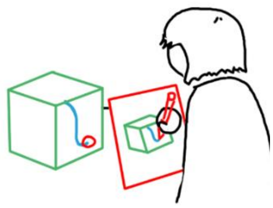
5.3.4. Level Two: Mid-Air Digital Content

While surface augmentation can give dynamism and interactivity to passive objects, it is unable to directly display content in mid-air. In order to provide support for this it is possible to use see-through devices to create the illusion of floating elements.

We prototyped hand-held see-through displays using the interactive pictures presented in the previous level, by attaching the camera position to the paper position (Figure 31) The user can then interact with the digital elements in their field of view using a tracked pen and ray-casting. Compared to head-worn displays (e.g. Hololens), such an approach has several advantages: First, users do not need to wear specific equipment. Then, the proximity to the physical world is strengthened; it is possible to switch very quickly and easily between the two visual modalities.



Both the physical-augmented object (level 1) and the through-the-lens object (level 2) can also be observed at the same time. Finally, multiple observer can use the same through-the-lens image. This reinforce collaborative and social interaction. The main drawbacks are that, as with other hand-held technologies, they require the device to be held, and they have a limited field of view. For the cases where comfort is relevant or both hands are needed (e.g., long or precise tasks) it is possible to detach the viewpoint and use indirect interaction, as seen in the previous level.



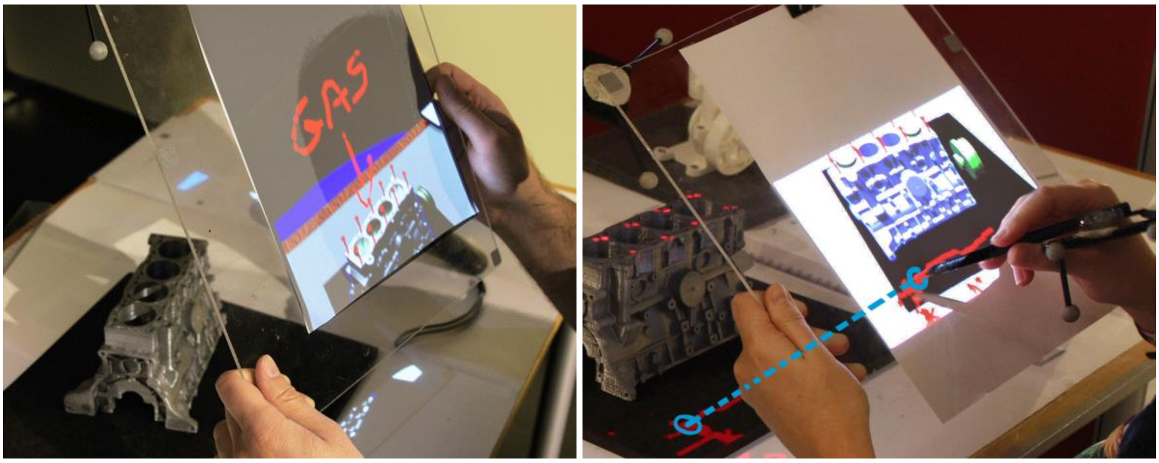
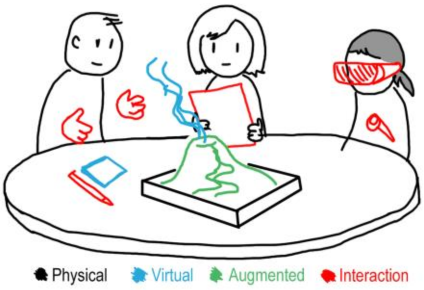


Figure 31: See-through displays, implemented using interactive pictures, allow to create the illusion of mid-air information (red text and arrow) floating over the engine (left), while also supporting indirect interaction with the scene via tracked pens and ray-casting (right).

Back to the Examples

To better understand the first three levels, let's explore the examples in more detail.

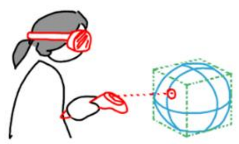
The students around the volcano mock-up (level 0) can observe superficial simulation (level 1) of its activity and in-place information, while also being able to move around and touch the mock-up, sharing a unified experience. A see-through screen can help the students to visualize mid-air information (level 2) such as steam and lava coming from the top of the volcano, or its interaction with virtual trees. They can use the screen to directly take notes of what they see through it, or place different views of the volcano side-by-side to discuss and make their own drawings.



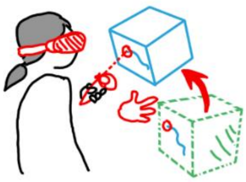
In the case of the engineer performing maintenance tasks on the engine (level 0), the augmentation can be bi-directional. She can both use previously stored information to guide the process, or create new annotations, creating a tutorial on how to perform the task (level 1). She can also use the see-through approach (level 2) to visualize mid-air annotations, measurements and spatial instructions, such as "screw here, then drill there". Once again, this can be bidirectional, either following or creating instructions through graphical annotations or by example.

5.3.5. Level Three: Object Decoupling

So far, level 1 and 2 aimed at keeping the augmentation as close as possible to the physical world. Starting at level three, we propose to soften the physical constraints in order to give more flexibility to the users. This can be useful when trying to understand complex processes (simulation), when needing to perform tasks not normally possible, such as lifting a heavy object (overcoming physical constraints), or trying to observe actions performed in the past (replay).



To this end, beyond the see-through displays we described in the previous section, we introduced a fully immersive modality that reproduces the physical world standing in front of the users, and that is perceived through a VR Head-Mounted display. Both worlds (physical and virtual) are mapped one-to-one. This means that the users can physically touch the objects they are observing virtually (Figure 32). This gives the immersed user a strong anchorage with the physical world. To increase the immersion, we use hand tracking based on Leap Motion, which provide not only feedback but also the possibility to directly interact with purely digital content.



At this level, the users can now free themselves from strict mapping and perform actions that would not be possible in reality (e.g. manipulate a heavy or fragile object). To do so, they can use virtual controllers. In addition to the features we already described, the users have this new ability consisting in distorting the reality.

Awareness considerations

When physical and digital are decoupled, it is necessary to take additional considerations to help the user to prevent accidentally hitting physical objects. For this reason, when users gets near a physical object (either with their hands or tools) we display its wire-frame or bounding box. It is also important to know where the users are (for themselves, and for others). For this, we display user avatars. The representation depends on the modality: augmented surfaces display the position and orientation in an iconic way (arrow), while see-through and HMD-VR show a 3D reconstruction of the users, enabling them to see themselves and others. Currently the 3D reconstruction is only based on the Microsoft Kinect point-cloud, yet it could be possible to extend it to rigged meshes [183], animated using Kinect skeleton tracking.

Finally, what the helmet shows can be displayed onto an interactive picture. This way, non-immersed users can see what the immersed user does. Given the behaviour of interactive pictures, users are also able to *interact* with the scene in front of the immersed user, indicating for example a point of interest, or moving elements through the window. Similar results can be obtained for non-immersed users using the estimation of head's position and orientation using skeleton tracking.

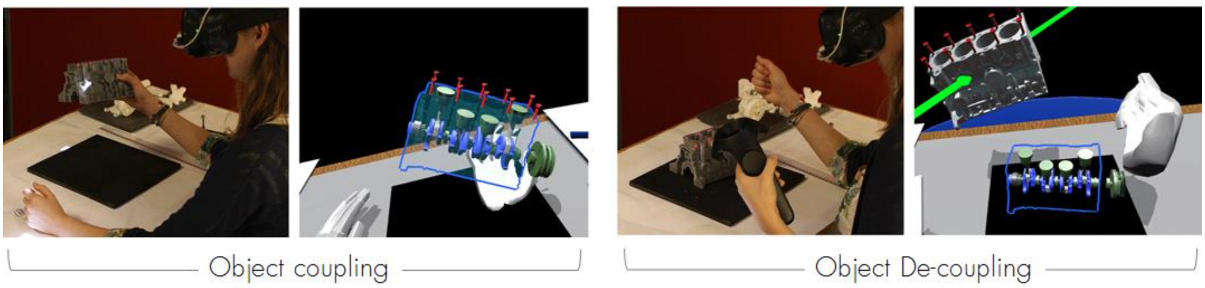


Figure 32: While immersed, the physical and digital counterparts of augmented objects can be mapped one-to-one (left). This coupling can be relaxed for additional flexibility (right).

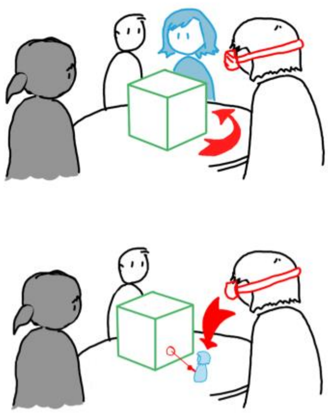
Back to the Examples

At this level, the students can virtually modify the geometry of the objects. For example, they can grab the volcano standing in front of them, and stretch it to observe what would happen if the volcano was higher. They can also extend the frontiers of the current physical mock-up by using 3D terrains around the volcano, or by visualizing internal structure above it. They can also copy and paste the volcano to experiment interactions between multi-objects.

In the case of the engineer, she can observe the engine in isolation by wearing an HMD, removing occluding elements or only showing their wire-frame. She can also virtually move a piece mid-air, in order to study it and annotate it. The annotations are reflected onto the physical engine in real time, since both views are synchronized. Finally, she can simulate the proper working behaviour of the engine, to clarify the function of the individual pieces.

This level gives users the illusion of changes in the physical world, while they keep their body as frame of reference: for a given user, the physical landmarks (objects and other users) stay still relatively to him or herself. The next level explores the separation between physical and digital body.

5.3.6. Level Four: Body Decoupling



The fourth level explores the immersive virtual navigation of the physical space. The users can then change their perceived position, orientation and *scale* in space (Figure 33). For this, HMD-VR helmets are ideal, since the users are presented with a 3D scene overriding their senses, with an arbitrary point of view. In order to control the navigation, we use a teleportation-based interaction using a wand controller, where the resulting scale depends on the target surface (and in our case is configured manually).Besides being able to teleport, users can also rotate either around a point, or on their current location. Since it is possible to ray-cast through the interactive pictures, users can pick the target position *through* a window, both before and during the immersive session.

The navigation is then not restricted to the current table and their surroundings, but could be used for tele-presence. Indeed, it is possible to transition to distant places (e.g., a co-workers office, or the surface of Mars), one step at the time (i.e., as in Google Street View[5]), or jumping there through an interactive picture.

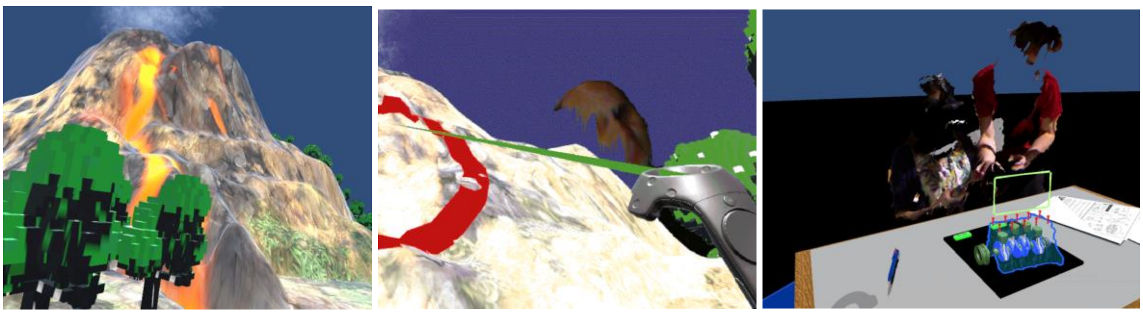


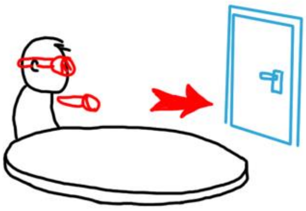
Figure 33: Users can navigate the augmented scene virtually. Left: the user jumped inside the scene and experience the volcano from an egocentric point of view. Right: the engineer rotated to her right, to see herself and her colleague.

Back to the Examples

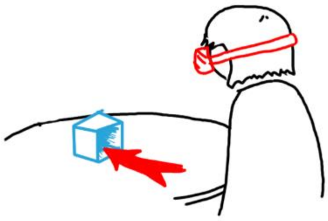
Interacting with the mock-up is enriching, but in the end it is a representation of something bigger. If the students want to experience or study the surface of the volcano, they can do so by wearing an HMD-VRs and teleporting on its surface (Figure 33). The virtual model can then be navigated by jumping from place to place, experiencing the represented scale. While on the surface, the students can see their schoolmates and themselves as giants around the table, allowing to asymmetrically collaborate, not just speaking but also using the aforementioned tools. They can use both the volcano surface and the interactive pictures to communicate with each other.

If the engineer needs help to repair the engine, she can observe another engineer performing the task, either using a recording or tele-presence. She can see her colleague working from his or her point of view, or virtually move around for a better view.

5.3.7. Level Five: Virtual World



Once the digital environment is framed in relationship with the physical one, it is possible to then transition to purely virtual spaces (Level 5). These spaces can be located at different positions and scales in relationship with the physical space, reminiscent of the iconic IBM's "Powers of Ten" [50].



These virtual spaces can have physical anchors. Framing the digital spaces in relationship with the physical ones can ease the transition, and make digital spaces less abstract. Imagine if applications were objects that required to be handled or observed to be used. This enables to spatially distribute digital spaces. The transitions can also be based on semantic relationships instead of spatial ones: using for instance object as triggers. This can create a truly hybrid space, that not superimposes but also interconnects both digital and physical worlds.

The potential of such possibility is further explored in **Chapter 7** through a concrete scenario: asymmetric collaboration. Additional applications that could benefit from this possibility are presented in **Chapter 9**.

5.4. Overview

In this section we described the 6 levels of our conceptual framework, accompanied by the implementation of the 4 central levels that provide augmentation. This framework was conceived to incrementally provide digital tools to interact with the physical world, up to the point where purely digital applications can be framed within the physical space.

The selection of a given modality involves a trade-off (Table 2). As digitality increases, so does the flexibility of interaction and simulation, while the connection with the physical world gets thinner by the instrumentation. It is possible to see the overlapping with the related work presented earlier in this Chapter (***Specific Related Work*** at page 55), and such projects can both be looked at through the lens of our conceptual framework, and integrated to equivalent ecosystems. We took awareness considerations to keep the user anchored in the physical world as much as possible, yet we argue that immersion should not be increased because it is possible, but only when it is needed. The result moves towards a space where physical and digital do not compete with each other for the users attention, but instead complement in the construction of a unified experience.

PROPERTY	LEVEL			
	1. Augmented Surfaces	2. Mid-Air Digital Content	3. Object Decoupling	4. Body Decoupling
Location of the display	On the surface of objects (main object, table, see-through screens)		Head Mounted Display	
Nature of the display	Object appearance Additional info (pics, text)	Virtual information and objects located in 3D space		
Physical-Digital relationship	One-to-one mapping		Scene and POV coherency	Independent
Main interaction	- Natural interaction (Object manipulation, pen and paper) - Digital tools (virtual ray)	- Natural manipulation of the screen - Window-based interaction (ray casting)	- Natural manipulation of the co-located objects - Dedicated interaction when detaching representations	VR interaction techniques
Main POV	Natural POV		Natural POV (simulated)	VR navigation

Table 2: Each of the augmentation levels and the preferred properties. Note how higher levels contain the features of lower levels, yet we recommend to not increase the immersion unless needed.

5.5. System Implementation

The system runs using two Alienware desktop computers, one worked as a server (Intel i7-3820, 8GiB RAM with an NVIDIA GTX 660 Ti) the second one as a client (Intel i7-3820, 24GiB RAM with dual NVIDIA GTX 690s). The use of multiple computers was necessary given the USB BUS bandwidth required by the sensors. The client performs dense acquisition (Kinect) and rendering (HTC Vive, projectors), while the server handles sparse tracking (Optitrack, Leap Motion) and stream it using UDP and OSC. We selected this distribution in order to minimize the intranet usage, yet dense acquisition and rendering could be distributed among several computers.

The whole setup was organized around a rectangular table of 130cm by 80 cm, which was used as the reference for calibration. All the sensors and projectors were placed on an overhead platform (Figure 34). Regarding the displays, we used: 1) an HTC Vive as HMD, and 2) three off-the-shelf projectors, providing 360^0 augmentation. The projectors are: an LG PF80G, an LG PF1500 and an Asus B1M; this selection was based on the available hardware, and the setup could be greatly improved by using 3 projectors of the same model, preferably with higher brightness and frame-rate. The tracking uses 4 Optitrack Flex cameras covering the volume over and around the table. A single Microsoft Kinect v1 provides a partial point-cloud of the table contents and the users, including their skeleton tracking. Given that we used HTC Vive, the tracking lighthouses are also placed on the overhead platform. For hand tracking we attached a Leap Motion to the helmet.

Besides the USB bandwidth, the main issue of this setup is the infrared interference between sensors: all of them -- except the Kinect -- use the same wavelength. To address this, we synchronized the Optitrack cameras with the lighthouses based on official documentation[4]; Leap Motion emitters can also create interference, but the impact was diminished with Optitrack parametrization. The use of additional Kinects would greatly improve the point-cloud coverage, yet additional considerations are required to prevent interference [27]. Regarding Microsoft Kinect v2, we were unable to include it in the system since it uses the same wavelength as the other sensors, leading to interference that could not be mitigated.

[4] http://wiki.optitrack.com/index.php?title=Sync_Configuration_with_an_HTC_Vive_System

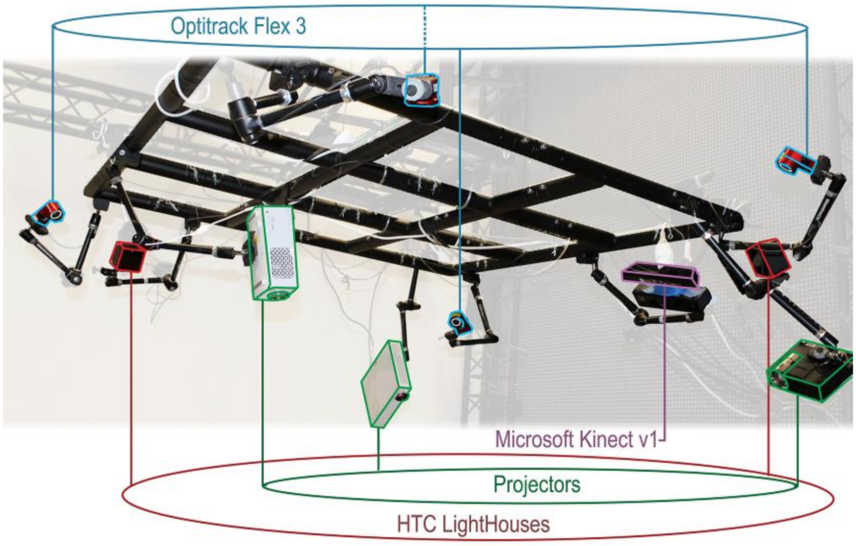


Figure 34: The setup supports surrounding tracking and augmentation. The tracking is performed using 4 Optitrack Flex 3 infrared cameras (sparse information), a Microsoft Kinect v1 (dense information point-cloud), and 2 HTC lighthouses (off the shelf HTC infrared emitters, used to track the HTC components). The augmentation is performed using three overhead projectors.

The alignment of all the subsystems was performed in stages. Both Optitrack and Kinect were calibrated offline, computing the alignment between coordinate systems (3D to 3D calibration, translation and rotation), using the table as frame of reference. The projectors were calibrated using OpenCV camera calibration (2D to 3D calibration, intrinsic and extrinsic estimation), by matching reference pixels with their 3D position, by using the already calibrated Optitrack. The HTC Vive was calibrated online, by placing the controller (most reliable transform) on a previously determined place in the table, and computing the transform between coordinate systems (3D to 3D calibration). Finally, the Leap Motion was manually calibrated, by finding the offset with HMD centre.

5.6. Conclusion and directions to future work

In this Chapter we presented a hybrid mixed reality conceptual framework, providing incremental augmentation and instrumentation. The framework was implemented using a combination of multiple display technologies. The first one is spatial augmentation, always available augmentation where only the surfaces appearance is modified. See-through devices allow to display mid-air information, and partially override physical information. Finally, immersive displays provide a virtual replica of the physical scene, taking advantage of the freedom of virtual spaces without losing connection with the environment, if desired.

Even when the different technologies coexist simultaneously in our system, the interaction was designed to perform the tasks while keeping the connection with the physical space and other users; when a task is not possible, then the amount of digital support is increased, in combination with awareness considerations to help the users keep the link with their environment. The resulting system focuses on the smooth transition from a purely physical to a purely digital experience.

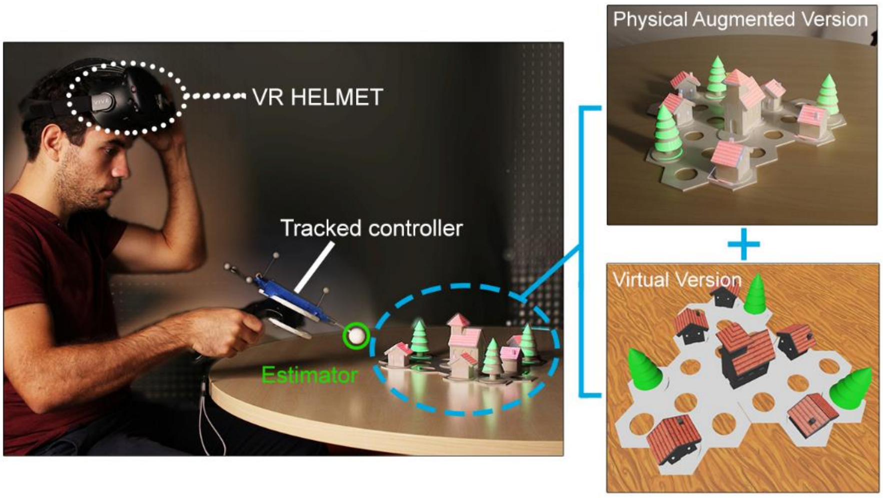
An underlying question driving this work is "where are these virtual spaces for the mind?" Before being able to assert that these systems are viable, it is necessary to know if users are able to use the heterogeneous information; that is the objective of the next Chapter.

PART II: Supporting the interaction with objects, the space and others

6. Position Estimation in Mixed Reality

Transferring information between *display modalities* and perspectives

"We wish to pursue the truth no matter where it leads — but to find the truth, we need imagination and scepticism both."
– Carl Sagan, Cosmos A personal voyage



About this section:

This chapter presents the evaluation of the users' capabilities to combine information obtained from heterogeneous displays (namely, projection based Spatial Augmented Reality and immersive Virtual Reality HMDs), with both egocentric and exocentric viewpoints. The task involved position estimation around an augmented mock-up and its virtual counterpart. The results show very robust estimation capabilities by the participants, and indications that the combination of technologies is not only possible, but also it improves the performance regarding pure virtual conditions.

The presented results are currently under revision for the upcoming ACM CHI'18 [**CHI18**]. This project was made possible thanks to the help of ***Jean Basset*** and ***Pierre-Antoine Cinquin***.

6.1. Introduction

In the previous Chapters we discussed the potential of hybrid Mixed Reality systems. As technology becomes mature enough to allow such hybrid MR systems become more common, it is necessary to better understand the ability of humans to interact with them. Indeed, once the technical issues are addressed, the success of a hybrid MR system will depend on the users' capability to create a unified mental model based on these heterogeneous representations.

This Chapter presents our efforts to answer the question:

Are users able to correctly complement digital and physical information in hybrid mixed reality systems?

To this end, we designed an experiment that requires participants to transport information between spatially augmented and virtual views of a single scene. This allows us to build knowledge to help the development of future hybrid systems that jointly stand on both the forces of VR and AR technologies.

The contributions of the presented work are:

1. The design of the experimental protocol based on Cognitive Science methodology,
2. A first user study focusing on *egocentric* (i.e., the natural point of view) position estimation task in SAR, VR and their combination, and
3. A second user study focusing on mixed egocentric and *exocentric* (i.e., from outside of the user's body) position estimation in MR and VR.

6.2. Specific Related Work

This work focuses on the empirical evaluation of participants' performance when using MR systems. As such, it is based on the literature on MR systems explored so far, and differs from (1) previous evaluations of MR systems by (2) driving inspiration from experimental approaches in Cognitive Science and VR.

6.2.1. Previous Evaluations of Mixed Reality Systems

When looking at the MR literature, the users' accuracy is rarely a studied factor[48], even if it is critical for the system's success. Most evaluations focus on either the quality of the subjective experiences generated [19,20], their impact on learning [49], or the communication between users [14,112,140].

To our knowledge, only a handful of evaluations explore the performance on perception tasks, perhaps given the difficulties of building MR experimental protocols [8]). The existing evaluations focus on depth estimation, asking the participants: if a virtual object is in front or behind a wall [65], asking to estimate the distance to a target either verbally or with an object (error in meters) [99,130], or the depth of a virtual floating object (error in cm) [15]. In all cases where accuracy was measured, the targets were visible when the estimations were made, thus evaluating perception and not explicitly evaluating the mental representation (i.e., *could the users operate once the information is not visually available?).*

* * * * *

6.2.2. Understanding Spaces

In Chapter 5 we show how mixed reality allow users to observe a single scene from both external and internal views of a single space, much in the line of Leigh *Mortals and Deities* presented at CALVIN [127], and Stoakley's World-in-Miniature[196]. It would be then of great interest to know more about the users' performance when combining the information of these heterogeneous views, and perhaps know more of the underlying mental processes. For this, we can drive inspiration from the approaches used in Cognitive Science and VR.

Wang and Simons [218] studied the capability of participants to perceive changes on a physical scene, either by rotating the scene or allowing the participants to change their viewpoint (the latter being more accurate). M.A. Amorim [4] evaluated the capability of participants to orient themselves in relation with an object, and vice versa, while in VR. Steinicke et al. [193] showed that using a virtual replica of the physical space as transitional environment can reduce depth compression in VR, evaluated through blind walking.

In order to evaluate the construction of a correct spatial model, these experiments share a three step structure:

1. The participants are provided with information,
2. The information is quickly removed while the scene is not visible for the participants,
3. The participants are asked to operate based on what they were shown.

This same structure is the one selected for our studies, where the task is completed based on both the physical and digital worlds.

6.3. Study Design

We designed a Mixed Reality experimental protocol that requires participants to estimate the location of a previously presented target. This experiment was used on two different studies: the first study involves an egocentric task (i.e., from the participant's perspective), while the second study combines both egocentric and exocentric views. This section describes the details of the environment where the task takes place, and the details regarding the system implementation and calibration.

6.3.1. Scene and Task

We tested the interaction with an augmented physical mock-up and its 3D counterpart (both the mock-up and the evaluation setup are presented on the next page, Figure 35 and Figure 36 respectively). For this, we used the 3D printed mock-up of a small town with 3 types of landmarks (5 houses, 3 trees and a church) over laser-cutted hexagonal bases (18cm diameter, 6 cm diameter per cell); the landmarks were distributed on a non-symmetrical layout, and the mock-up was placed at the centre of a circular table (135cm diameter). The SAR version provided basic texture mapping. The virtual version reproduced not only the augmented mock-up and table, but also the room where the experience took place.

Targets and Position Estimation: Both studies involved a position estimation task. Participants were sitting facing the mock-up, and were iteratively shown a spherical target (3cm of diameter) in a location either inside or around the mock-up at one of three possible heights (3, 6 or 9cm). After the target was hidden, the participants were asked to place an estimation using a sphere of the same diameter (3cm) attached to a wand controller (Figure 37), and to confirm the estimation by pressing controller's trigger (using the soft "hair-trigger" trigger mode to mitigate the movement caused by clicking).

Figure 35: SAR (left) and VR (right) versions of the mock-up and the targets (3, 6, and 9cm in height)

Homogenization of Conditions: The researcher conducting the study was sitting next to the participant (90 degrees to the participant's right, also facing the mock-up, Figure 36). The conditions involving physical targets required the researcher to manually place and remove the target, the position indicated using projection. During this time (when the researcher had to interact with the mock-up), the participants were asked to close their eyes, while a board was placed in front of their eyes to prevent peeking. To keep the conditions as similar as possible, the VR counterpart steps displayed a black screen, using a fade to black to make it less abrupt. To standardize the time measurements and workload, each of the steps that required attention were preceded by a 3 second countdown, and this was indicated visually on the helmet, but also verbally for all conditions (to balance the mental workload). The times per trials were recorded to evaluate their impact on the performance.

6.3.2. Apparatus

The software was implemented using the approach described in Chapter 5. The hardware layout was comprised of off-the-shelf components: 4 Optitrack flex 3 cameras, an LG projector PF80G and a HTC Vive set (helmet, controllers and lighthouses). The tracking of the controllers was performed using Optitrack (Figure 37) instead of the HTC tracking. This decision was based on the fact that HTC Vive uses sensor fusion for tracking, and as a result the world position of the components is not known accurately enough for our case.

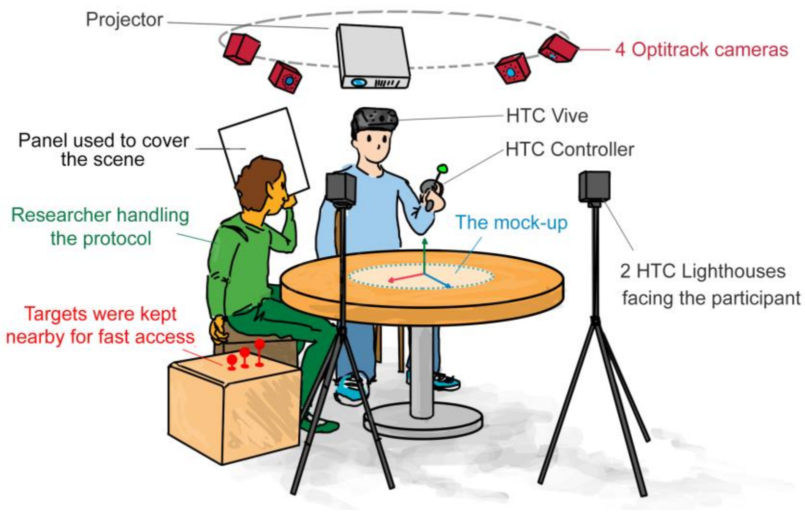


Figure 36: Experiment setup: HTC Vive lighthouses were placed in front of the participant, while the Optitrack cameras and the LG projector where placed around and over the participant. The researcher conducting the study was sitting to the right of the participant at all times.

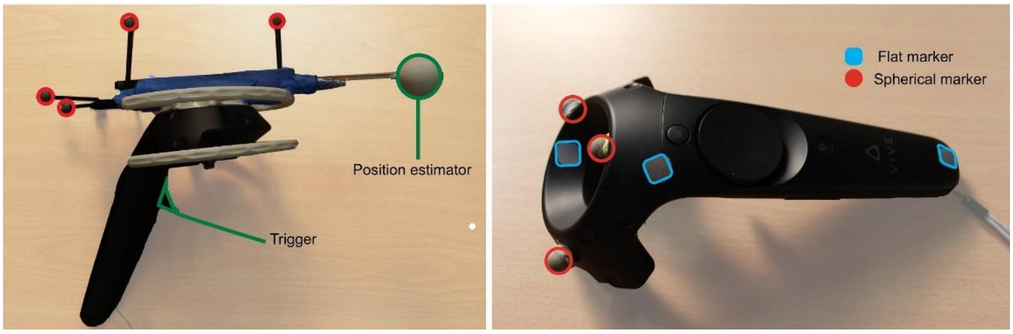


Figure 37: Controller with estimation attached to it (left). Controller used to align Optitrack and HTC Vive spaces, markers placed to easily identify origin and orientation of the controller (right)

6.3.3. Calibration

To guaranty the quality of the results, the system was calibrated at least once per day. First, the Optitrack volume was calibrated over and around the table, and the origin was set to the centre of the table. Then, the projector was calibrated using OpenCV camera calibration, by matching 3D points in the Optitrack frame of reference with 2D points in the projector's image plane (as explained in

). Finally, the alignment between Optitrack coordinate system and HTC Vive was performed computing the 3D to 3D transform using a controller with infrared markers as reference (Figure 37-right).

System accuracy: The Optitrack calibration reported an error at sub-millimetre scale (under 0.2mm), while the projection calibration showed in average a reprojection error of 3.8px (causing a mean error of 1.8mm at the centre of the mock-up). The HTC Vive to Optitrack registration error was in the order of 1cm, which only affected the head position, since the controller was tracked using Optitrack to guaranty higher precision, as previously explained.

6.4. Study 1: Egocentric Estimation

The first study focused on comparing performances and subjective similarities between the physically augmented and virtual scenes. The objective was to test if:

- **H1:** People using a mixed reality environment are able to perceive information in one space, virtual or physical, and to use this information to perform tasks in the other space.
- **Operational hypothesis:** The error when performing a position estimation task (i.e., placing an element at a previously indicated location) in hybrid conditions (combining *both* SAR *and* VR) should be *no bigger than* in pure conditions (*either* SAR *or* VR).

6.4.1. Participants

In order to recruit participants for the study we made a public announcement and posted it on the mailing lists of both the institute and the university. A total of 18 participants (11 male, 7 female) volunteered to take part in the study. Their ages ranged from 21 to 51 (26 ± 9), all of them except one were right handed, and they all had normal or corrected-to-normal vision. Most of the participants were students or members of the research institute. Most of the participants play video games: 6 of them play frequently, 9 of them play occasionally, while 3 of them never play. Most of them (11) played sports frequently while growing up, 6 occasionally, and one never. 10 still do sports frequently now, 5 occasionally, and 3 never. 13 of them already had experience with HMD based virtual reality, and 10 of them with AR applications. This demographic information was obtained in order to detect influencing factors.

6.4.2. Procedure

Participants were welcomed and asked to sign a consent form that explained the objectives of the study along with clarifications regarding the anonymity of the data, known risks of VR and that they were volunteers (i.e., free to stop at any time). Once the consent form was signed, participants filled a demographic questionnaire: age, gender, sports and video game habits, and experience with AR and VR. Then, they filled two mental task tests to know more about their profile: 3D mental rotation test[213], and 2D spatial orientation [79]. These questionnaires took around 20 minutes to complete.

The evaluation involves a position estimation task, comprised of 4 runs with 12 trials each, using a within-participant counterbalanced conditions. Four conditions were considered (Table 3): seeing the target either in SAR or VR, and pointing either in SAR or VR; the hybrid conditions (*SAR_VR* and VR_SAR) required the participants to either place or remove the helmet, and for this reason they were alternated to reduce discomfort (i.e., one trial of *SAR_VR*, then one trial of *VR_SAR*, and so on). To keep the trial length homogeneous, then the combined *SAR_VR* and *VR_SAR* was divided in 2 runs, for a total of four runs, in 3 groups of conditions to counterbalance (Table 3). Each run was followed by the NASA TLX [76] (standard questionnaire used to estimate the effort required to complete a given task), and a custom questionnaire for the conditions involving the HMD.

Regarding the trials, the first trial of each run was explicitly indicated to the participants as a training trial, and was discarded; given that the alternated condition had a pause, the seventh trial of each run was also discarded, leaving a total of 10 usable trials per run. At each trial, the participant was presented with a target for 5 seconds, and then had to estimate where they considered the target was previously located, by placing an estimation artefact of the same size (Figure 37). Once the estimation was confirmed, the target was displayed again, along with an indicator of the estimation position at the table level. The reason why the estimation feedback was placed at the table level was to keep both *SAR* and *VR* conditions equivalent (since it is not possible to make 2 physical objects intersect to provide feedback for the SAR condition).

Finally, a short unstructured interview was conducted, to know more about how the participants felt during the experience. The evaluation took around 45 minutes, given a total time of around 65 minutes for the whole experience.

CONDITION		See target	Make estimation
Type	Code	(Egocentric view)	(Egocentric view)
Spatial AR (SAR)	SAR_SAR	SAR	SAR
Mixed Reality (MR)	SAR_VR	SAR	VR
	VR_SAR	VR	SAR
Virtual Reality (VR)	VR_VR	VR	VR

Table 3: Conditions for Study 1. Four conditions were grouped in three runs, based on their type. Only egocentric perspective was involved

* * * * *

6.4.3. Measurements

For each trial we registered in a log file the positions (target, estimation, head position) and times (total, estimation_start and estimation_end). From this, all the metrics were obtained.

Accuracy - Absolute error: The absolute error is the distance between the target location and the estimation location in world coordinates (expressed in centimetres (cm)).

Accuracy - Failure count: we consider failures those trials with estimation error over $6cm$. This decision was based on the properties of scene ($6cm$ is the size of one cell of the mock-up, and twice the diameter of the target), over that limit we consider the participants forgot or confused landmarks. Failures were counted, but not considered when computing mean error distance.

Accuracy - Subjective error: The subjective error is computed as the signed distance between the target location and the estimation, taken from the participant's point of view.

Cognitive Load: After each run, the participants were asked to complete the NASA TLX.

Subjective Experience: evaluated with a custom 7-point Likert scale questionnaire, and completed after the conditions involving the HMD (Question listed in Figure 42).

6.4.4. Results

In this section, we present the analysis of the data collected, including accuracy, workload and subjective experience.

Data Analysis

To ensure independence, trials were reduced to one sample per participant per factor combination (using mean). In the cases the data presented a non-normal distribution, we used the Aligned Rank Transform [231] to correct our data (indicated with an ART subscript when presenting the results). Then, we used ANOVA on the data (corrected or otherwise), and used Bonferroni as post-hoc analysis. Bivariate correlations were computed using Pearson when non-categorical variables where involved. All the data analysis was performed using SPSS 23, with the pre-processing of the data performed in Microsoft Excel 2013.

The obtained results are displayed at two levels: (i) p-values for statistically significant differences, paired with mean values and confidence intervals (grouping the trials per participant), and (ii) distribution box-plots (not grouped). This was done as an effort to complement the p-values[47], and allow the reader to have their own interpretation of the data. To ease the readability of the text, we will refer to the charts instead of indicating mean and deviations numerically.

Accuracy Results

Failure count: The number of estimations with an error over 6cm (i.e., over twice the diameter of the target) was overall low. As can be observed in Table 4, the failure count ranged between *4%* and *7%* depending on the condition. As mentioned in measurements, these estimations were not considered when computing the mean error.

CONDITION	FAILURE	
	Count	Rate
SAR_SAR	7/180	3.9%
SAR_VR	12/180	6.7%
VR_SAR	10/180	5.6%
VR_VR	13/180	7.2%
Overall	42/720	5.8%

Table 4: Failures for Study 1.

Accuracy by condition: The accuracy of the subjects did not seem to be affected by the condition, according to a one-way ANOVA *($F(3,68)_{ART}$ = 0.491, p = 0.690)*. This can also be observed in Figure 38.

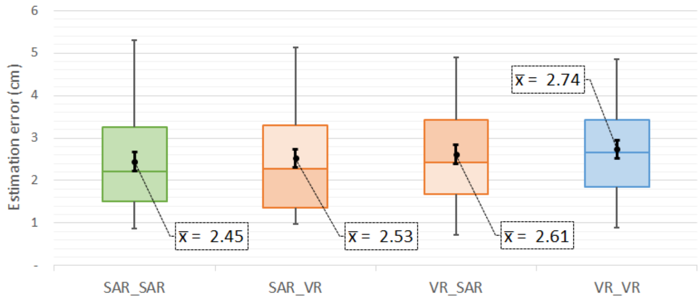


Figure 38: accuracy by condition

Accuracy by region: To better understand the results, the targets and their estimations were divided in clusters based on their distance from the mock-up (inside or outside) and angle from the centre (left, right, top), giving a total of 6 clusters (Figure 39). Results are presented in Figure 40.

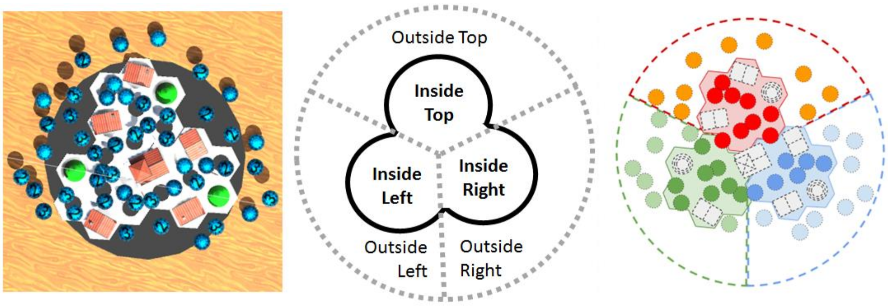


Figure 39: 48 target locations (left), 6 regions (centre), and the 48 targets clustered by region (right). The participant (not represented) would be placed at the bottom of the picture.

We conducted a two-way ANOVA explaining the absolute error by the condition and the region. The results show a significant effect of condition on the absolute error *($F(3,380)_{ART} = 1.724$, $p = 0.162$)*. The ANOVA also showed a significant difference in absolute error between the targets' regions *($F(5,380)_{ART} = 9.793$, $p < 0.001$)*. Finally, the ANOVA rejected the interaction effect between condition and region for the absolute error *($F(15,380)_{ART} = 1.129$, $p = 0.328$)*.

Figure 40 presents the error distribution per region and condition. For the region, the absolute error on targets outside the mock-up is significantly higher than for the targets inside, while estimations for the same region show similar results for different conditions (except for VR_VR).

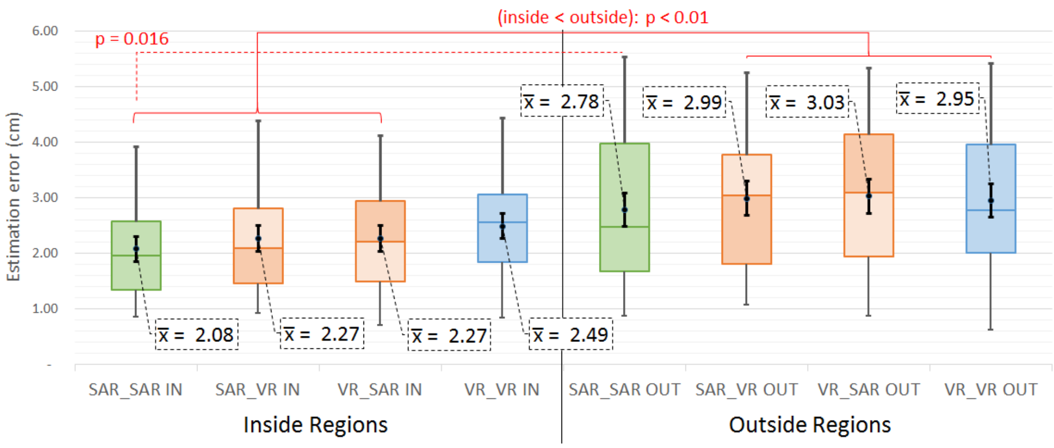


Figure 40: Absolute error by region and condition

The position of the failure count also seem to confirm that result. Indeed, out of the estimations with an error over 6cm, most are positioned in the outside clusters *(37/42)*, half of which are in the outer top cluster *(18/37)*, leaving only *5/42* failures at the inside clusters.

Depth error: There is a strong effect of depth in error on the condition, particularly for VR_VR. A one-way ANOVA showed a significant difference in depth error between conditions ($F(3,60)_{ART} = 5.285$, $p=0.003$). It is possible to look at the depth error as a signed variable (Figure 41). In our case a positive depth error means that the estimation is between the real target position and the user. Participants tended to estimate the target closer to them than it really is, and even more in the VR_VR condition.

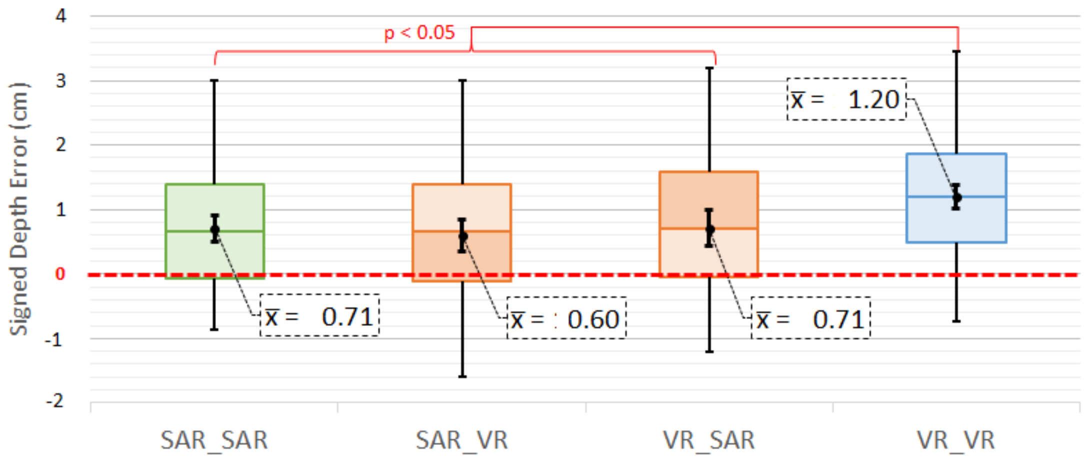


Figure 41: Depth error, a positive error in depth implies that the participants estimated closer to themselves (presented this way for clarity).

Subjective Experience Results

The results obtained by the subjective experience questionnaire (Figure 42, 7-Likert scale) and the comments obtained during the interview were similar. First and foremost, no statistical differences were found between mean scores for MR and VR conditions ($F(1,34)=0.002, p=0.969$, inverting the values of negative questions Q6 and Q7).

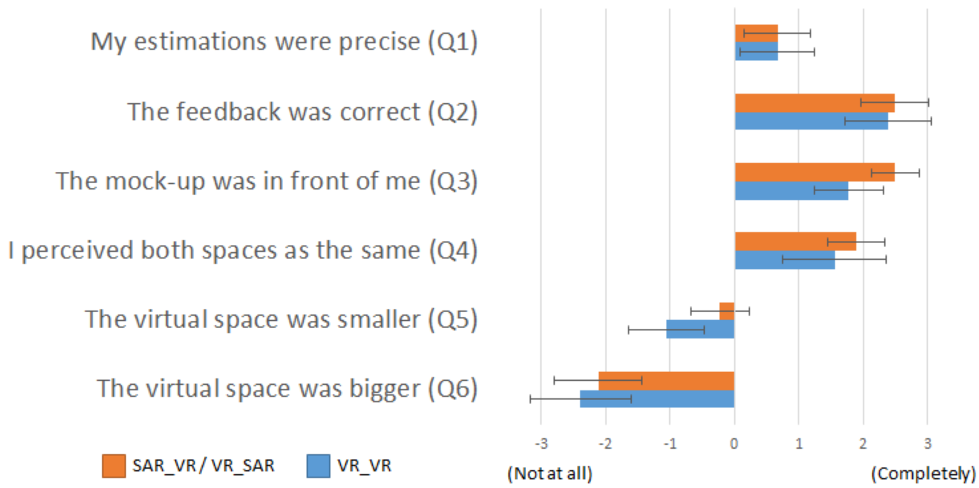


Figure 42: Subjective experience based on a 7-Likert scale questionnaire, values between -3 and 3. Error bars indicate confidence intervals.

A descriptive analysis of the obtained results shows both modalities received rather positive scores (mean scores, *MR: 1.65±1.50*; *VR: 1.64±1.55*, values ranging between -3 and 3), that the participants perceived both spaces as the same once they understood the mapping (Q4), and felt the mock-up was in front of them even while immersed (Q3). Note that even for the VR_VR condition, participants were sitting in front of the mock-up, thus seeing it before and after the run. . All but one of them felt overall precise, and accredited the estimation error to themselves (Q1), rather than the system (Q2); this difference was confirmed verbally during the interview (the participant which questioned the system accuracy tended to face directly down during the experience, thus occluding the HMD tracking). When considering the fidelity of the registration between SAR and VR modalities, the answers present a high variance (Q5), and participants mentioned that considered the fact only after the question was asked to them.

Both during the protocol and the following interview, participants mentioned difficulties with the helmet or the VR rendering. Several participants were initially disoriented by the lack of feedback on where their hands were. Most participants mentioned that the illumination of the virtual scene was different than the physical one, and in particular the shadows were too strong; participants reported that the height of the estimator was harder to see while in VR. Some participants (particularly females) mentioned that the helmet felt heavy, and we had difficulties with some of participant's haircuts.

Workload Results

Regarding the NASA TLX workload, only 17/18 subjects were evaluated, since one of them provided an incomplete questionnaire and had to be discarded. The tasks measured by the NASA TLX are the three groups of the study (SAR, VR, and MR). We conducted a two-way ANOVA on the workload estimated by the subjects explained by the task and the order in which the tasks were passed. The ANOVA showed no effect of the task on the workload $(F(3,54) = 0.453, p=0.716)$, no effect of the order $(F(3,54) = 0.253, p = 0.859)$, and no interaction between task and order $(F(7,54)=0.400, p=0.898)$. The obtained results are displayed later on when discussing the workload for both studies (Figure 47).

Influencing Factors

We found some influence of the demographic data collected and the mental rotation tasks on the accuracy of the subjects. The questionnaires were used as a way to know more about the population rather than to correct the results and further experimentation should be conducted to obtain reliable conclusions, but tendencies can still be observed.

Mental tests: The mean error shows an inverse correlation with *spatial orientation* $(r=-0.387,p=0.01)$, and no significant correlation with *mental rotation* capabilities $(r=-0.045,p=0.721)$; as reported [79], both tests show correlation with each other $(r=0.601,p=0.01)$. Even when both mental rotation tests show correlation with playing sports (particularly *playing while growing up*: Mental Rotation: $r=0.425,p=0.01$; Spatial Orientation: $r=0.530,p=0.01$), no correlation was found between *sports* and *estimation error*.

Gender: Females obtained higher values for absolute error *(females: 2.76±0.43; males: 2.38±0.50)*, to a significant extent according to a 2-way ANOVA by gender and condition $(F(1,64)=10.586, p=0.002)$, while there was no interaction between gender and condition. The same effect appears for the signed depth error *(females: 0.98±0.57; males: 0.64±0.53)*, also significant $(F(1,56)=7.482,p=0.008)$. This is consistent with studies involving depth estimation in VR [6].

Additionally, differences between genders were found in relationship with the mental tasks: the correlation with mental rotation between spatial orientation and estimation error is significative only for males *(males: $r=-0.705, p=0.015$; females: $r=-0.087, p=0.852$)*; this could indicate a sample effect explaining the difference in performance between genders. We consider the differences in accuracy between genders not large enough in practice, even when significant.

No trial duration influence, no order effect: Even when the time taken to complete one trial was different for each condition *($F(3,68)=21.692, p<0.001$),* no correlation was found between this time and the estimation error *($r=-0.022, p=0.852$)*, perhaps because the differences are small in practice *(SAR_SAR: 16.8±1.83s; SAR_VR: 20.41±5.74s; VR_SAR: 17.68±2.58s; VR_VR: 14.56±1.40s)*. Skill transfer between VR and physical environments has been found in the past for other tasks (e.g.,[126]), yet such an effect was not detected *($F(2,60)=0.798, p=0.455$)*, nor an interaction between order and condition *($F(6,60)=0.689, p=0.659$)*.

6.4.5. Study 1 - Conclusion

The results of the first study indicate that both spaces are perceived and interacted-with in a complementary manner. The error is mostly influenced by the region where the participant is targeting, more so than the condition involved. Regarding the MR conditions (SAR_VR and VR_SAR), the participants' accuracy was not significantly different with the SAR condition (no HMD), nor between each other. It is also worth noticing that even when in average participants estimated closer to themselves, the VR condition is the only showing a significantly depth compression [91] when compared to the control condition without HMD (SAR); this is the case even when both mixed conditions required to perceive the space using VR (either to memorize the target location, or to estimate the location).

6.5. Study 2: Complementary Views

In the second study we focused on the change of scale and point of view, to test if:

- **H2:** Users are able to complement their perception of a space when being provided information from different viewpoints.
- **H3:** Complementary views can reduce the estimation error, particularly when dealing with far objects / lack of landmarks from the user's POV.
- **Operational hypothesis:** similar estimation error between conditions would imply similar perception, using the mixed condition as a referent to study the complementarity of modalities. To this end, accuracy should be: 1) not worse for MR than for VR, 2) not worse than for the first study (**H2**), and potentially 3) better than in study one (**H3**).

6.5.1. Participants

For Study 2, we wanted to ensure it was possible to can compare the obtained results with Study 1, so we performed recruitment from 2 sources: **new participants** obtained through a public announcement published on institute and university mailing lists, and **repeating participants** obtained by contacting the participants that performed the first study.

A total of 20 participants (15 male, 5 female) volunteered to take part in the study, 9 of which were took part of Study 1. Their ages ranged from 20 to 58 (26.8±8.4), two of them were left handed, and they all had normal or corrected-to-normal vision. Most of the participants were students or members of the research institute. Most of the participants play video games: 6 of them play frequently, 5 of them play occasionally, while 9 of them almost never play. Part of them (7) played sports frequently while growing up, 11 occasionally, and two never. 4 still do sports frequently now, 12 occasionally, and 4 never. 12 of them already had experience with HMD-VR, and 12 of them with mobile AR.

6.5.2. Procedure

The second study followed the procedure used in Study 1, with some minor corrections. The protocol involved showing from an egocentric point of view a location on the table where the user will be teleported, as an arrow oriented towards the centre of the mock-up. The number of locations were 6, located every 60 degrees around the mock-up at a distance of 25 cm from its centre. Once the location was presented, the participant was then "teleported" to that location thanks to the VR helmet, where he or she could see the target location for 7 seconds. After the 7 seconds, the participant was presented with the scene once again from an egocentric point of view, and then estimate the target's location. As with Study 1, feedback was displayed regarding the target location and the participant's estimation.

Since changing the point of view can only be done while wearing the helmet, we considered as independent variable the display modality used outside (i.e., to see the target location, and to perform the estimation). As a result, 2 conditions were considered (Table 5), each of them consisting of two series of 12 trials each. The extension to 2 series was in order to observe if there is a learning effect. Given that the task was considered harder than in the first study, the first 2 trials were explicitly discarded as rehearsal trials, given a total of 10 usable trials per run. These conditions were counterbalanced within participant, and within group (i.e., the conditions were alternated for ***NEW*** and ***REPEAT*** participants independently).

Most of the questionnaires and forms were shared with the first study. The half of the participants that took part of the first study did not fill the entry questionnaires, only the consent form. The only different questionnaire was the subjective experience questionnaire, which was extended to include questions regarding the change in point of view.

CONDITION		See destination (Egocentric view)	See target (Exocentric view)	Make estimation (Egocentric view)
Type	Code			
Mixed Reality (MR)	SAR_VR_SAR	SAR	VR	SAR
Virtual Reality (VR)	VR_VR_VR	VR	VR	VR

Table 5: Conditions for Study 2. Two conditions, each performed twice. The task involved a change in viewpoint (change of op position and scale), only possible using the HMD

6.5.3. Scene Corrections

Based on the comments from participants of the first study, the virtual scene was improved by decreasing the intensity of shadows. An iconic avatar was added at the location of the participant, to give a frame of reference when immersed.

In order to keep a constant distribution of landmarks, the mock-up was rotated 180 degrees, while the target locations was re-randomized (Figure 43). Each target was associated with one of the nearest POV locations, taking special care on preventing total occlusions. Since the change of perspective added a digital arrow to provide orientation, this could modify the available landmarks around the mock-up. This was taken into account when considering the accuracy similitude between studies, as explained in the *Results* section.

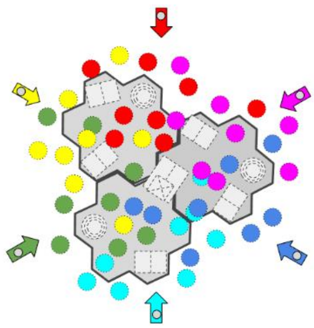


Figure 43: target and location distribution. The participant (not shown) would be located at the bottom of the figure.

6.5.4. Measurements

As with Study 1, for each trial we registered on a log file the positions (target, estimation, head position, *and teleport position*) and times (total, estimation_start and estimation_end). All the metrics were computed as in Study 1 (namely, absolute and depth error, error count, time, and cognitive load).

6.5.5. Results

This section presents the obtained results for the second study, and when relevant, it compares these results with the counterparts from Study 1.

Accuracy Results

Failures: Regarding the failures (estimations with an error above 6cm), the total count adds up to 46 of the estimations. Participant #19 accounted a total of 11 of these errors (27.5% of their own estimations), presenting an outlier behaviour, and for this reason was excluded from the evaluation. The remaining 35 failures were similarly distributed among SAR_VR_SAR and VR_VR_VR conditions.

CONDITION	FAILURE	
	Count	Rate
SAR_VR_SAR	18/380	4.7%
VR_VR_VR	17/380	4.4%
Overall	35/760	4.6%

Table 6: Failures for Study 2, after excluding participant #19.

Condition, viewpoint and target location: The accuracy of the subjects was not found significantly different between conditions $(F(1,36) = 0.397, p = 0.533)$, as seen in Figure 44-top. The impact of which viewpoint was used to observe the target showed no statistical significance $(F(5,210) = 0.996, p = 0.421)$. When looking at the spatial distribution of the error, results show a more uniform distribution than in the first study. No significant statistical differences were found for the accuracy between targets inside and outside the mock-up $(F(1,72) = 0.005, p = 0.943)$ (Figure 44-bottom), nor detectable interactions between condition and location $(F(1,72) = 0.088, p = 0.768)$. Regarding the failures, they happened similarly often outside (17/35) and inside (18/35).

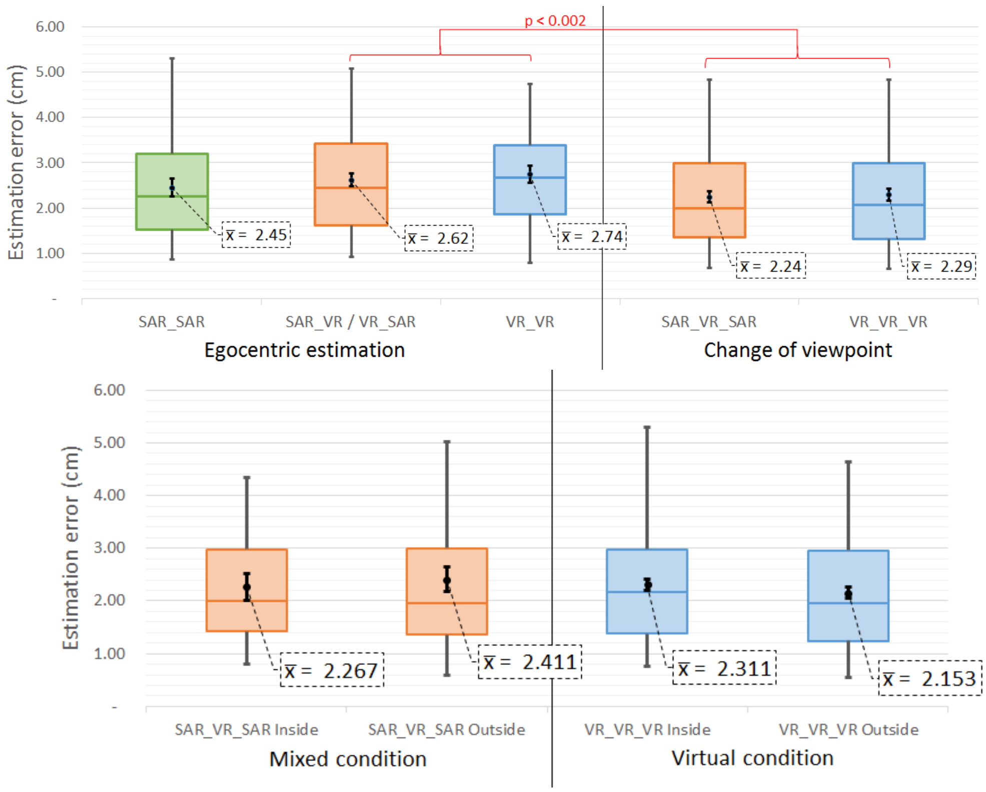


Figure 44: Estimation error per condition vs Study 1 (top), and taking into account the region (bottom).

Participant group and accuracy between studies: There were no statistical differences between *NEW* and *REPEAT* participants *($F(1,34)_{ART}$ = 0.007, p = 0.933)*, nor for between *REPEAT* group and the other participants of Study 1 *($F(1,16)$=0.007, p = 0.933)*. When comparing accuracy, a significant difference in accuracy was found between studies *($F(1,35)_{ART}$ = 4.569, p=0.04)*: overall, the first study showed a higher estimation error than the second study. We remind the reader that for Study 1 a strong region effect was found; when comparing both studies, no statistical differences can be found for the internal regions *($F(5,104)_{ART}$ =0.898,p=0.485)*, while the external regions present significant differences *($F(5,104)_{ART}$=4.296,p=0.001)*, up to various extents (compare Figure 44-bottom with Figure 40 in page 72). The accuracy in Study 2 (for both internal and external targets) is then similar to the accuracy for internal targets in Study 1.

Depth error: The error from the participant's perspective did not show significant differences between conditions *(F(1,36)=2.048, p=0.161)*, nor from the exocentric perspective (Figure XX). When comparing the results with Study 1, a significant difference can be found *(F(5,104)=16.114, p<0.005)*, given that Study 2 presents less distance compression on all conditions. This can be observed comparing Figure 45 and Figure 41 in page 73: both studies present similar distributions from the egocentric perspective, yet there is a shift towards zero for the Study 2.

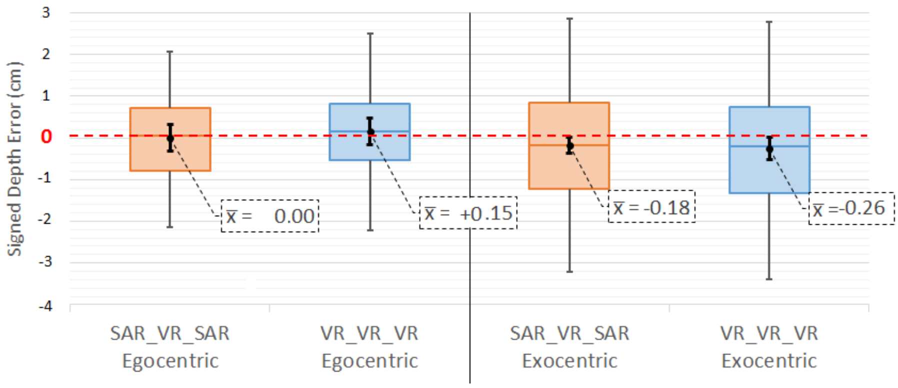


Figure 45: Depth error. The error per condition for Study 2, computed from both the egocentric (left) and exocentric (right) viewpoints. Note the distribution around zero, in contrast with Study 1 (see Figure 41 in page 73).

Order effect: The order of the conditions has a significant impact on the accuracy *(F(1,34) = 7.901, p = 0.008)*. The post-hoc analysis shows tendencies towards improvement for the second condition, both for SAR_VR_SAR *(p=0.054)* and VR_VR_VR *(p=0.074)*, as can be seen in Figure 46.

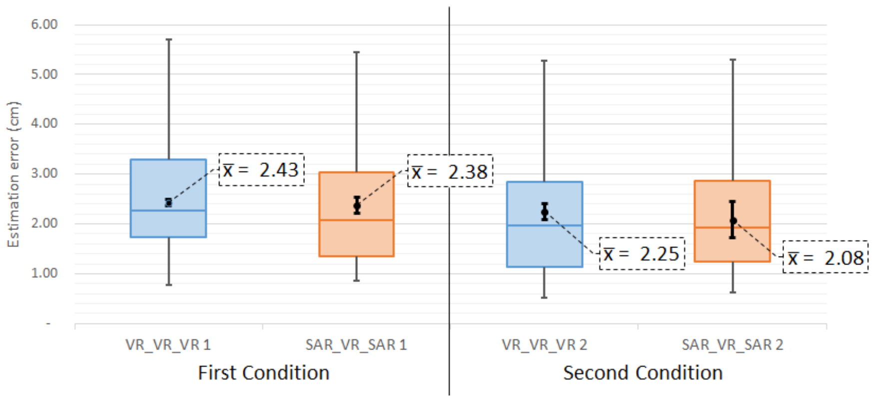


Figure 46: Estimation error over time.

Workload Results

When studying the impact of condition on the workload, no significant differences were found by *condition* $(F(2,72)=1.221, p=0.273)$, nor by *order* $(F(3,72)=0.078, p=0.972)$, nor an interaction *between order and condition* $(F(3,72)=0.753, p=0.524)$. When comparing against the workload reported in the first study, a strong tendency was found $(F(12,127)=2.380, p=0.055)$; the post-hoc analysis reflected this tendency only between SAR_SAR and VR_VR_VR $(p=0.099)$; this can be observed when looking at the workload distribution (Figure 47).

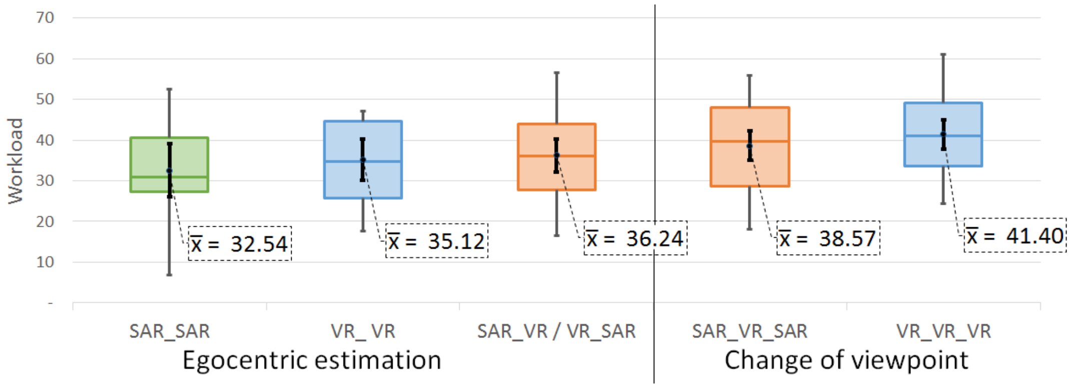


Figure 47: Workload results from the NASA-TLX questionnaire. Both the results for the first and second study are presented, from lowest to highest.

Subjective experience

No statistical differences were found for means scores for the subjective questionnaire between conditions $(F(1,38)=1.099, p=0.301)$. A descriptive analysis of the results (Figure 48) shows that both conditions present positive mean scores $(MR: 1.04 \pm 0.53; VR: 0.9 \pm 0.50)$ when inverting scores for negative questions (Q8, Q9, Q10).

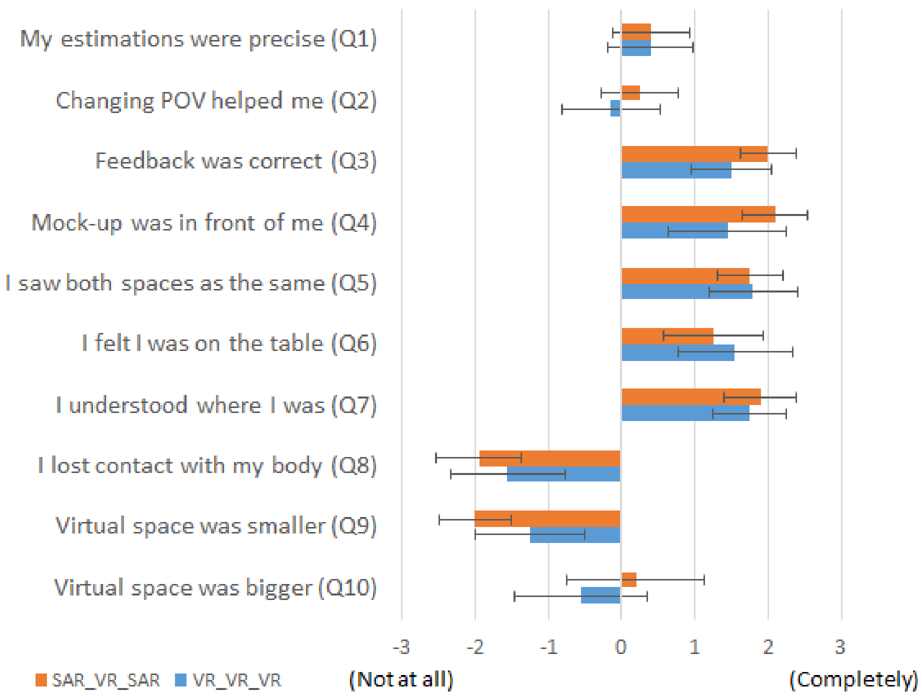


Figure 48: Results obtained for the subjective experience questionnaire for the second study, using a 7-likert scale. The results are presented on a scale between -3 to 3 for clarity.

Questions regarding the feeling of presence (Q6) and complementarity between scenes (Q5) scored consistently high. Participants did not feel they lost contact with their body (Q8) or location (Q4), albeit the virtual condition scored slightly worse on these questions. The feedback was considered precise (Q3), yet slightly less precise in the virtual condition. Participants did not feel particularly precise (Q1), nor considered the change in POV was particularly useful (Q2).

When studying the answers for Q2 (*changing scale is useful*) by participant, *NEW* participants tended to give a slightly positive answer *(MR: 0.73±1.27$, VR: 0.63±1.12)*, while *REPEAT* gave a slightly negative answer *(MR: -1.2±1.09, VR: 0.22±1.2)*.

Questions regarding the perceived scale (Q9: *virtual space smaller*; Q10: *virtual space bigger*) had an ambiguous nature (since the task involved a change of scale), and the answers varied from participant to participant. The results for Q9 and Q10 are similar for both condition orders.

6.5.6. Influencing Factors

Mental tasks, video games: Opposite to the findings from the first study, an inverse correlation was found between *estimation error* and *mental rotation* only for the virtual condition *(VR: $r=-0.564$, $p=0.012$)*, and no significant correlation was found between *estimation error* and *spatial orientation* capabilities. In this study, no significant correlation was found between the tests *($r=0.183$, $p=0.452$)*. Additionally, estimation error presented an inverse correlation with playing video-games *($r=-0.539$, $p=0.017$)*, which extended mostly to VR *(VR: $r=-0.475$, $p=0.04$)*, as only a tendency towards significance was found for MR *(MR: $r=-0.394$, $p=0.095$)*.

Order effect and trial duration: The tendency towards an order effect was previously discussed. Time varied between conditions *($F(1,36)=4.5$, $p=0.041$; MR: 19.3±2.2s; VR: 17.9±1.7s)*, and an inverse correlation between time and error was found (*$r=-0.4$,$p=0.013$)*. It seems that the longer the participants take to estimate, the more precise they are.

6.5.7. Study 2 - Conclusion

Participants seem to be able to correctly complement the viewpoints. Accuracy was not worse than for Study 1 (considering the SAR_SAR condition as a control task), even when the task involved can be considered harder. The similitude in accuracy between internal and external regions can be caused by the digital arrows acting as effective landmarks, or perhaps by the redistribution of targets.

This is supported by comparing the results with the ones obtained for Study 1: accuracy was found overall higher, and a tendency towards accuracy improvement for external targets (i.e., those placed outside the mock-up) was found.

Even when participants were overall at least as precise, they did not report a subjective improvement regarding Study 1 (not only in average, but particularly the participants that were part of both studies). This seems to indicate that they expected higher accuracy, perhaps caused by seeing the scene from closer.

Overall, the obtained results support **H2** *(Users are able to complement their perception of a space when being provided information from different viewpoints).* It is not possible to make strong statements regarding **H3** *(Complementary views can reduce the estimation error, particularly when dealing with far objects / lack of landmarks from the user's POV);* instead, it can be said that the change in perspective *in combination with the addition of digital landmarks* seems to reduce the estimation error for far away objects / lack of *available* landmarks from the users POV.

6.6. Understanding the accuracy results

In order to put the obtained results in perspective, Figure 49 shows three cases of distances between target and estimation. For both studies, when presented with enough landmarks, participants were able to obtain some degree of intersection between target and estimation (between case A and B of Figure 49). This was consistently achieved in most cases for most participants, disregarding modality (SAR or VR) or the viewpoint used to obtain the information (egocentric, exocentric). For both studies, around 25% of the estimations where under 1.5cm (case A), around 75% of the estimations were under 3cm of error (case B), and the remaining estimations were in most cases under 6cm (between 93% and 97% estimations are better than case C, depending on the condition).

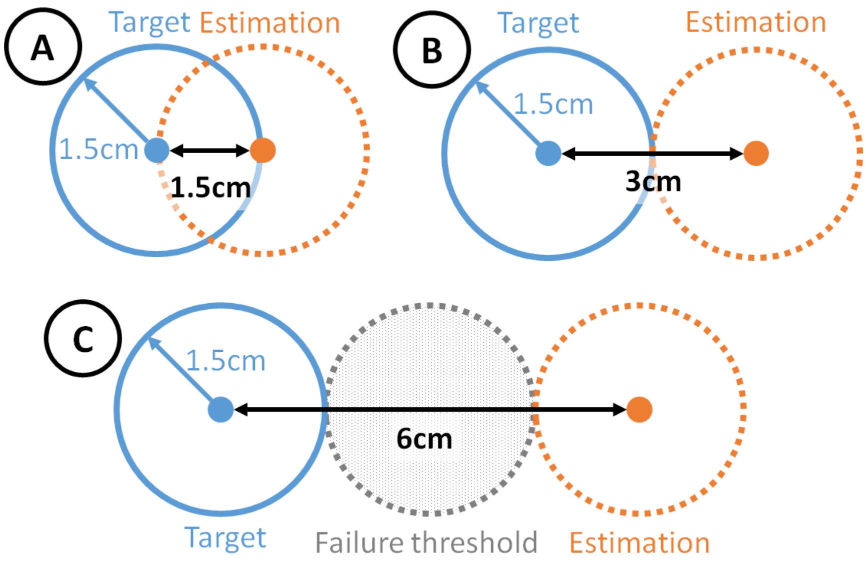


Figure 49: Three instances of estimation: touching the center of the target (A), touching the target (B), and the failure threshold (C).

In addition to the rather robust estimation capabilities for all conditions, there could be in place a skill transfer, as shown by Study 2. This might imply that operating in different modalities could be equivalent in practice. It is worth noticing that participants considered their performance not good enough, in contrast with the perceived high precision of the system. This indicates that there is room for improvement at the users' end before the precision of the system becomes an issue.

Finally, the pure VR conditions tended to have slightly higher error, and a significantly higher distance compression. It seems that alternating physical and digital can reduce this effect, which we consider this aligns with the findings of [193], and is a strong indication towards the construction of a unified mental model by the participants.

* * * * *

6.7. Study Design

In order to provide some additional information regarding the study design, here we present the considerations and limitations of the study.

6.7.1. Considerations

When designing the protocol, we took the next aspects into account:

Homogeneity: At the protocol design stage, when facing a trade-off caused for differences between SAR and HMD-VR, we opted for the feature that could be implemented in both, prioritizing homogeneity between conditions over usability. For instance, when showing the estimation feedback only the base was indicated (since it is not possible to intersect two solid physical spheres), or the lack of explicit height indication when making the estimation while in VR (since it would provide an advantage over the SAR condition). Such self-imposed limitations do not need to be preserved when designing hybrid interfaces, allowing designers to improve each modality by addressing their limitations independently.

Non-ecological, yet prioritizing comfort: We are aware that putting and removing the helmet is uncomfortable and rather non-ecological, and we are not proposing this as the correct usage of this kind of systems, but instead as a way of evaluating the capabilities of the participants of combining both spaces (it would be better to use ST helmets [20], yet the render quality is not yet comparable to HMD-VR). Given this, during the first study we alternated SAR_VR and VR_SAR conditions to minimize discomfort. The consequence of this decision was the lack of individual answers for the post-run questionnaires (subjective questions, and workload), while no significant differences were found for the available measurements. Even when our objective is to move towards a unified space (and bidirectional interaction would be ideal), these conditions could be studied independently in the future.

Risk of survival bias: For the second study we decided to also contact the participants from the first study, which had the risk of presenting a survival bias. We consider that the impact of this decision is negligible, since we are not asking questions regarding the enjoyment of the experience (e.g., we did not use the SUS [24]), and we found no statistical differences for accuracy between NEW and *REPEAT* participants, nor between *REPEAT* and other participants of Study 1.

6.7.2. Limitations

Some limitations should be explicitly mentioned, to place the obtained results in context, and to be taken into account for future evaluations.

The depth estimation error is a known effect in VR called distance compression [118]; since participants reported virtual elements as smaller while in VR, the field of view was not matching the natural one. This could have been mitigated by taking into account Inter-Pupillary distance. As mentioned on the *subjective experience results* of Study 1, ergonomics are also a factor. Even when the HMD used is an improvement versus previous generations, it seems to still present difficulties for the general population, and particularly for females (similar results are reported in ***Chapter 8***).

There were several effects detected (gamers, sports, gender), but the protocol was not prepared to take them into account, and the population was not balanced in order to reach conclusions. Still, the found effects resonate with the literature [6,128]. A particular effect that could have been explicitly taken into account is the variability in working memory of the participants. Additional questionnaires could be used to measure it (e.g., the complex figure test [29]), and longer memorization times per run could be also used, at the cost of longer sessions.

Finally, the virtual scene had a non-photorealistic shading, with a schematic reconstruction, even when the technology used could support more realistic rendering. Participants did not mention difficulties with this, yet the illumination was reported as a recurrent source of errors for the first study (in particular, the shadows were too strong, and were perceived incorrectly as feedback). We consider that increasing the visual similarity between physical and virtual scenes could improve the performance of the participants, as discussed in the previous Chapter.

6.8. Conclusion

This chapter presents the results of two user studies that focus on the user's accuracy on mixed reality systems. The first study considered variations -- and similarities -- between mixed reality conditions from an egocentric viewpoint, while the second study evaluated the complementarity of egocentric and non-egocentric viewpoints for target estimation.

6.8.1. Results

The obtained results indicate that, as with other spatial tasks, the accurate perception of the space is supported by the presence of landmarks. Participants showed a remarkable capability to transfer information between SAR and VR modalities, even between ego/exocentric POVs. Additionally, perceiving the scene from closer seems to increase the participants' expectations on their accuracy.

It is worth mentioning that depth compression was significantly higher than in purely physical scenarios only for egocentric tasks that happened solely in VR. Hybrid MR conditions do not seem to suffer from this anymore than in purely physical tasks in the case of egocentric estimation. All these results indicate that the participants were able to construct a unified mental model from the heterogeneous representations and views.

6.8.2. Moving towards concrete applications

The end of this Chapter also means the end of ***PART II*** of this manuscript. So far, we discussed the general notion of combining technologies and degrees of augmentation, and showed that humans can properly interact with heterogeneous representations. The natural following question: *What can we do with this kind of hybrid mixed realities?* This question is addressed in ***PART III***, where we explore potential applications.

PART III

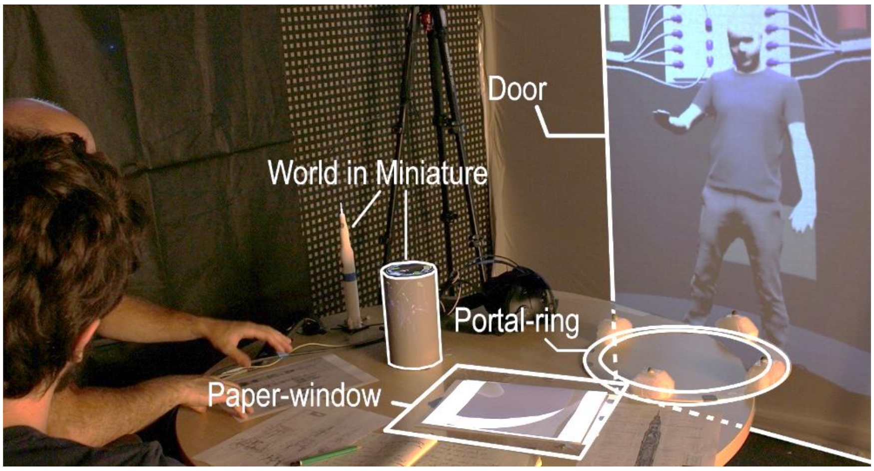
Collaboration, wellbeing, and other possible applications

PART III: Collaboration, wellbeing, and other possible applications

7. Asymmetric Collaboration in Mixed Reality

Reducing the distance between remote locations (physical and virtual locations alike)

"Organic as a dandelion seed, the ship of our imagination will carry us to worlds of dreams and worlds of facts."
–Carl Sagan, Cosmos



About this section:

This chapter presents the extension of One Reality to asymmetric collaborative scenarios, in the context of the Aerospace Industry. When different roles work at different locations (either physical or virtual), the awareness, communication, and ideally navigation between places can greatly improve the collaboration. We propose tools to progressively perform these tasks, much in the philosophy of One Reality.

The works described in this section were accepted as a short paper at ACM VRST'2017 **[VRST17]**. This project is the result of a collaboration with ***Damien Clergeaud***.

7.1. Context: Asymmetric hybrid collaboration in the industry

In the industrial context, Virtual Reality is more and more used to support the design and iteration of products (CAD reviews, ergonomic studies, assembly task design, and many others). Compared to physical prototypes, virtual environments enable engineers to test solutions faster, safer, and cheaper. Moreover, in some domains such as in the case of the aerospace industry, all the products are designed digitally with CAD software, which facilitates the use of VR. On the other hand, the design of complex artefacts involve multiple iterations based on decisions discussed by multiple stakeholders, which still require physical meetings.

The need of both physical meetings and VR prototyping leads to asymmetric collaboration. Indeed, the combination of both physical and virtual spaces could not be replaced by co-locating all the participants: immersing all the users in a virtual environment would break the meeting dynamics, while keeping the operators in the meeting room would greatly reduce the experimentation possibilities. As a result, collaborators with different roles at different locations must effectively communicate in order to successfully achieve their objectives and sometimes have to move between virtual and physical spaces.

The question regarding asymmetric collaboration was brought to us by engineers from **Airbus Group**, a mayor aerospace manufacturing company. They use VR to design assembly tasks directly with the final operator immersed in the digital mock-up, while experts supervise and guide the process from a meeting room. In order to follow the operator's work, experts can observe him or her through a monitor located at the meeting room, while an audio channel supports the communication between the two locations. For instance, experts could design a procedure to mount a particular system in the launcher while the operator is following the procedure to complete its design.

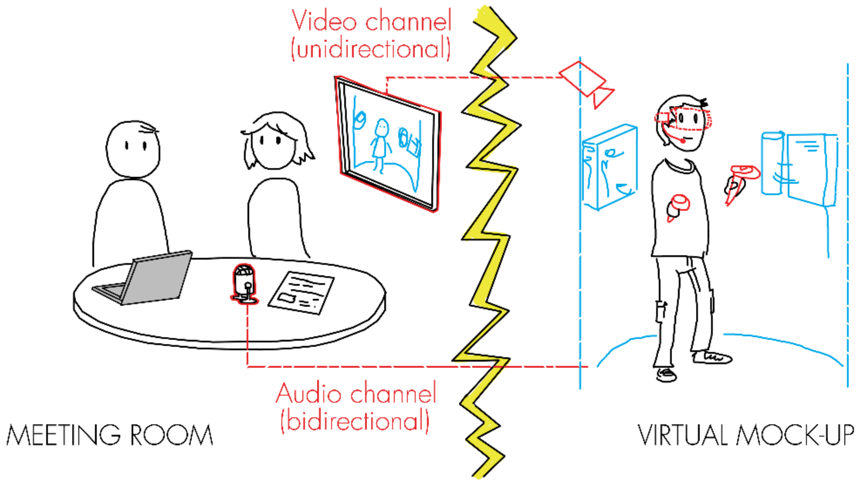


Figure 50: Current method used by Airbus. The collaborative design process involves discussions in a physical meeting room and simultaneous experimentation in VR. Both locations are connected though an audio channel (bidirectional) and a view of the virtual scene (unidirectional)

The engineers from three different branches of **Airbus Group** agreed on the limitations of the current approach for asymmetric collaboration. Their main concerns orbit around the fact that the viewpoint and the audio channel do not support awareness and communication properly. First, it is not uncommon that the experts forget the presence of the operator, since they work in parallel. Second, they struggle to provide instructions for precise or complex tasks. Finally, there are instances where being temporarily present at the remote location (in either direction) could ease the understanding between participants.

The purpose of this work is to provide tools that address the limitations of the current approaches for asymmetric collaboration, in order to ease the communication between these remote locations. This involves 1) supporting awareness between spaces, 2) tools that support remote operations, and 3) ways to visit the remote location. To do so, we choose to focus on the creation of Through-The-Lens technique (TTL) [197] artefacts for both the physical and virtual locations.

The contributions of the current work are: 1) the conception and prototyping of tangible Mixed Reality lenses -- and their virtual counterparts -- for asymmetric collaboration, 2) the description of the system used to support these artefacts, and 3) preliminary feedback from future end users of the aerospace industry.

7.2. Specific Related Work

This work combines technologies for telepresence and mixed reality with through the lens techniques. Through-The-Lens techniques (TTL) allow connecting spaces together. More formally, TTL enable users to simultaneously explore a virtual environment from two or more different viewpoints as described by Stoev et al. [197]. If the second viewpoint is linked to the location to another immersed user, it provides a new channel for communication, as proposed by Kunert et al. with Photoportals [119]. Photoportals are a mix between viewports and portals with a collaborative purpose, enabling users to create flat (2D) or volumetric (3D) viewports to observe a remote location or to directly teleport themselves through it. It can also be used to retrieve remote objects, reducing the perceived distance between locations. We propose to extend this line of research by supporting the asymmetrical collaboration between physical and digital spaces, using both traditional VR and tangible metaphors, as described in the next section.

7.3. Proposed Method

In order to facilitate the asymmetric collaboration between physical and virtual locations, we propose to support three core features (Figure 51): 1) awareness through overview of the remote scene and collaborators; 2) the ability to remotely interact and manipulate elements; and 3) the capability to navigate between locations. With these objectives in mind, we created artefacts that support the required features, using either SAR or VR. SAR and TUIs were used in the instances where immersion would break the dynamics of the local interaction. For the virtual locations, the interaction supports both tangible and purely virtual artefacts; these alternatives were created in order to enable more immersion or to free the hands respectively.

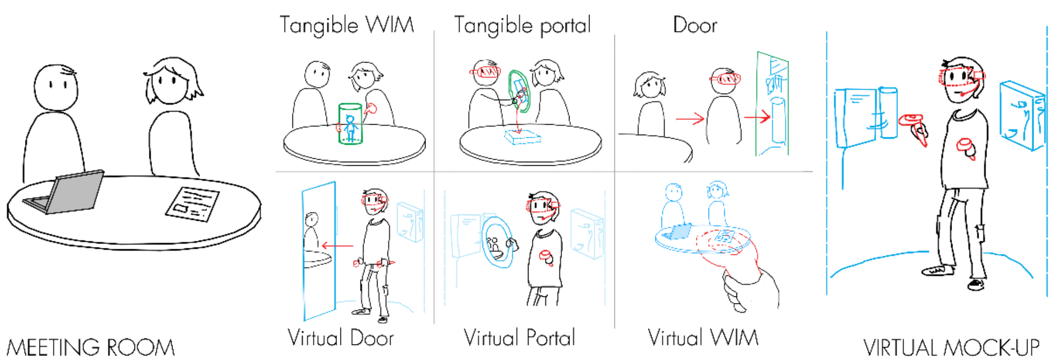
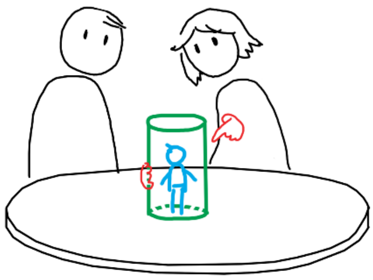


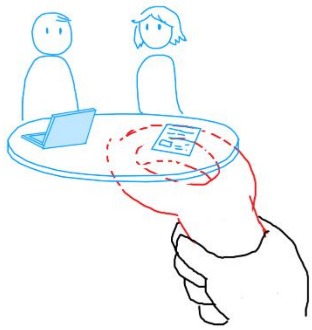
Figure 51: The proposed artefacts bridge the distance between the local and remote locations, enabling progressive immersion and interaction capabilities on the other side, up to navigating there.

7.3.1. Awareness through Scaled Representations

In order to enable the communication between the two locations, it is first necessary to have an overall understanding of the whole space and the location of each collaborator. For this reason, we decided to use World-In-Miniature [196] as the starting point for interaction and communication. WIMs provide a general overview of the context of the remote collaborators, while also supporting coarse interaction through pointing. Collaborators are displayed using 3D avatars, as animated rigged 3D scans [183].



We implemented independent scaled down mock-ups for physical and virtual locations. The physical counterpart is supported by a physical mock-up (which is possible to 3D print thanks to the already available CAD information) with spatial augmentation generated via projection. Spatially augmented mock-ups work as ambient displays [230], with always-available-yet-subtle information, which makes them ideal to keep the users aware of the remote activity without being too distracting. In VR, the WIM displays the remote table and its surroundings, immersed users can hold a tangible prop to observe the remote place, and leave the prop on the workbench nearby. Alternatively, immersed users can interact with a virtual WIM (i.e., without a physical support).



When building physical mock-ups, both limitations and opportunities arise given the intrinsic physical properties. In the case of the launcher mock-up, two complementary views are necessary: an overview of the whole launcher and a detailed overview of the working piece. This was reproduced in our case with two separated mock-ups (Figure 52): the first one showing the location of the active section, while the second one shows combined inside-out and top-down views of the active room.

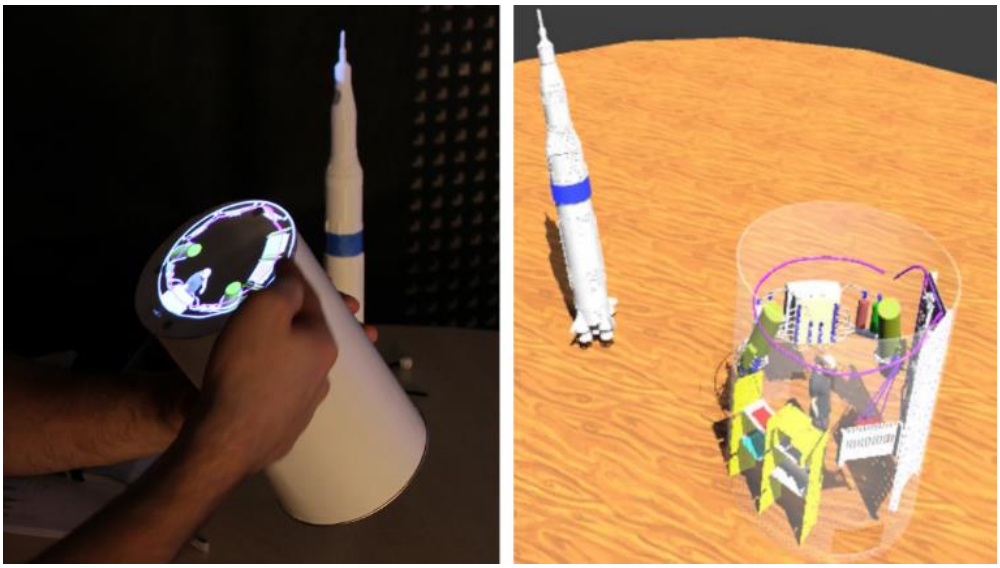
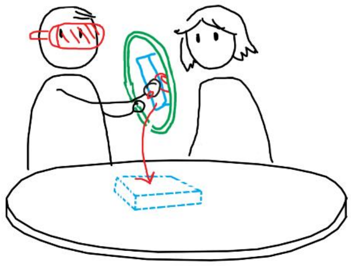


Figure 52: The tangible world in miniature provides a general view of the remote scene (left), and can be also similarly manipulated while wearing the HMD (right).

7.3.2. Interaction through Windows



Windows provide lenses to other spaces from given points of view, enabling interaction through them, as with traditional desktop screens. The main advantage of windows and their relative small size is the possibility of using several of them simultaneously. In our context, windows provide a powerful solution to bridge both spaces.

We support two types of windows: *interactive-paper* and *portal-rings*. The former provides a see-through view of the scene, allowing to display mid-air information at the meeting room location much like a see-through tablet, as explained in ***Chapter 5***. Portal-rings are hollow frames that display the remote location, and enable interaction *through* them. Users can pass their hands to interact with the other side, and move elements between locations (Figure 53). Windows then provide the possibility to observe and interact with the other side, allowing a better communication between locations.

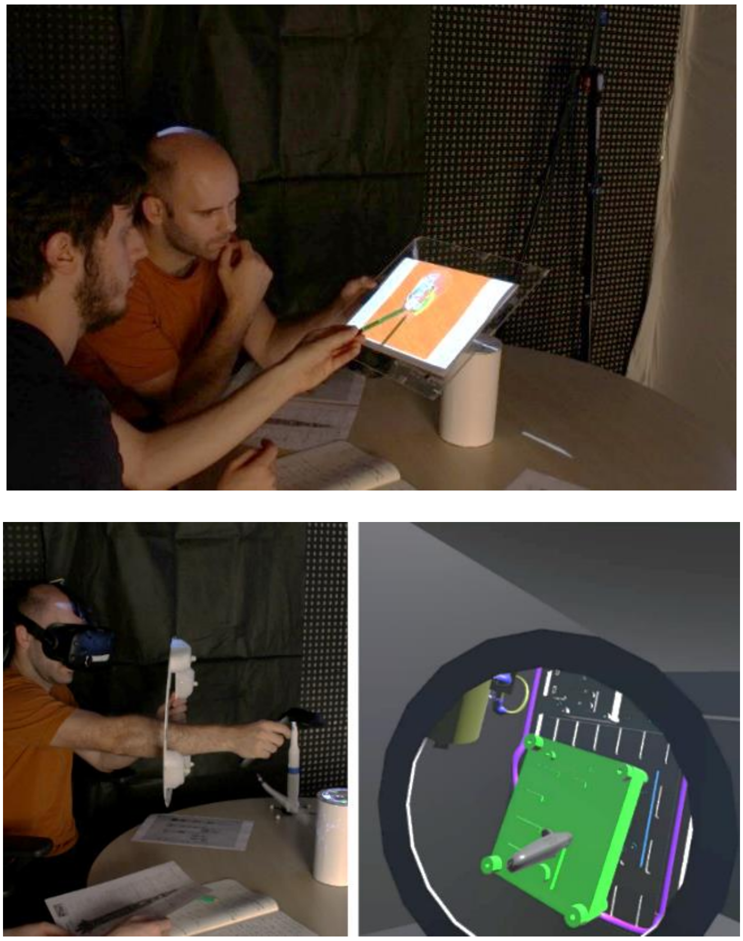


Figure 53: It is possible to interact with the remote scene using see-through and the WIM (top); alternatively, elements can be manipulated remotely and moved between locations through the window metaphor of TTL technique (bottom).

7.3.3. Navigation through Doors

Doors are similar to windows, but besides providing a viewpoint of the remote location, they also enable to go there. Doors and teleportation are fairly common in VR, in particular to deal with limited physical spaces. In our context, we place a virtual door at a physical wall of the meeting room thanks to projection (Figure 54). Regarding the navigation, the user immersed in VR does not need additional tools, as portals are a commonly used technique. For the users at the meeting room, the door is displayed on a wall using projection while the navigation is performed in VR, by wearing an HMD. Alternatively, the door could be located on a nearby space with CAVE capabilities, thus the process of walking to the door could provide a seamless transition into a dynamic VR space.

Once users transited to the remote location, they can directly interact with its virtual version. Displaying tele-presence collaborators in VR is done by showing their avatar, while in the meeting room is necessary to use an additional display (either HMD or paper-window).

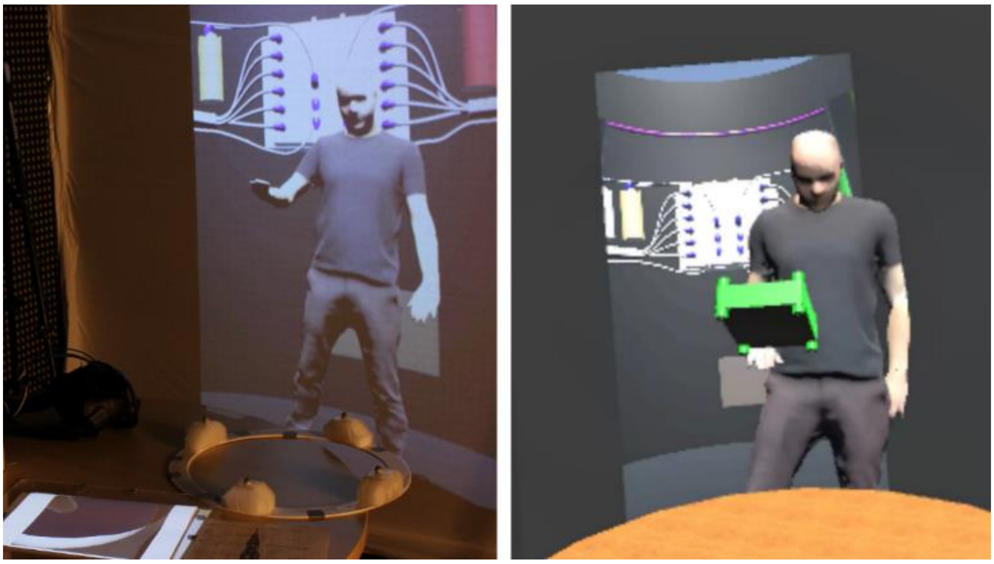


Figure 54: Collaborators can see each other through doors, similar with windows (left), yet they have the additional benefit of being able to go to the remote location using the door (right). It is clear that we cannot display the virtual avatar with the current technology (albeit mid-air displays exist), so the users at the meeting room must wear a HMD to see their collaborator.

* * * * *

7.4. System Implementation

The prototype was implemented as an extension of the system presented in ***Chapter 5***. It is comprised of peer applications connected using UDP and OSC (real-time control protocol built over UDP). In each location, the hardware comprised of Optitrack Flex cameras, HTC vive, and optionally projectors. Both applications are identical, except by the hardware configuration; this design decision supports the creation of a topology of locations interconnected through artefacts, yet the artefact configuration and networking require further consideration. Since the setups are identical, the immersed user could have a transition space by adding projectors, up to the point of having a CAVE-style space with augmented props.

7.5. Feedback from Partners

After the design and implementation of the prototype, we performed a demo session with the three engineers that brought the use case to us. The engineers were presented with the features of the system one by one, following the structure of this paper, from the meeting room side (since the VR interaction is closer to their expertise); one of the researchers was immersed inside the launcher mock-up from a nearby room.

The physical WIM was perceived as a good way to be aware of the remote location, placing the operator literally at the centre of the table. The use of 3D printing created mixed emotions, and the conclusion is that their utilisation will depend of the use case (e.g., good for training since it is reused, not good for daily meetings). The manipulation of the WIM while immersed was enjoyed yet rapidly discarded as impractical given its weight, preferring the use of the pure virtual WIM. Regarding the portal-ring, the engineers found the possibility of taking objects from the other location of great use, and the gesture of bringing the object through the portal was intuitive. On the other hand, holding the ring was once again seen as unpractical, and they preferred the virtual version.

The use of paper-windows was rapidly understood and they envisioned this the easiest to adopt through the use of tracked tablets. This was particularly appreciated when interacting with the WIM in particular, and more generally with all the digital objects, like the ones taken through the portal-ring.

The door was the most appreciated artefact, since it allows the operator to come into the meeting room, while the navigation into the other direction was not consider novel. When presented with the door, they envisioned how to improve their current workflow using it. Concretely, operators follow strict step-by-step protocols to prevent mistakes, so they need to keep the virtual scene uncluttered; being able to do back and forths to the meeting room enables them to focus on one task at a time, while the experts can coordinate the whole process. In their context, the use of the door to move elements between locations was better suited than portal-rings.

Regarding the technologies involved, the engineers mentioned limitations on both. The VR helmet was perceived as cumbersome to utilise and equip. They considered the extension of the system to support a lighter augmented reality helmet such as HoloLens, while they also were aware of the loss of resolution and FOV implied. Regarding SAR, on one hand they liked the un-instrumented interaction, yet the use of flat augmented surfaces was limited, and they would prefer richer rendering. An interesting conclusion they reached is that not all experts will use the same modalities, and the limitations of one technology can be mitigated with the other.

7.6. Conclusion

This work was inspired by the current limitations in asymmetric collaboration expressed by industrial engineers. Based on the literature, we created artefact prototypes to improve the awareness, communication and interaction between locations. Finally, we presented these prototypes to engineers in order to get feedback.

From the proposed solutions, the use of windows and doors where the two preferred features. Windows were considered to complement the current approach, while also being easy to adopt. Regarding the doors, they envisioned how they could improve the existing workflows.

Interestingly, the tangible physical TTLs were instantly understood, yet quickly discarded for practical reasons. We consider this does not suggest that tangibles should be avoided when designing prototypes, quite the opposite. Their use in early stages can lead to a better understanding of the interaction metaphors (and perhaps during the training sessions), to then be replaced with purely virtual widgets.

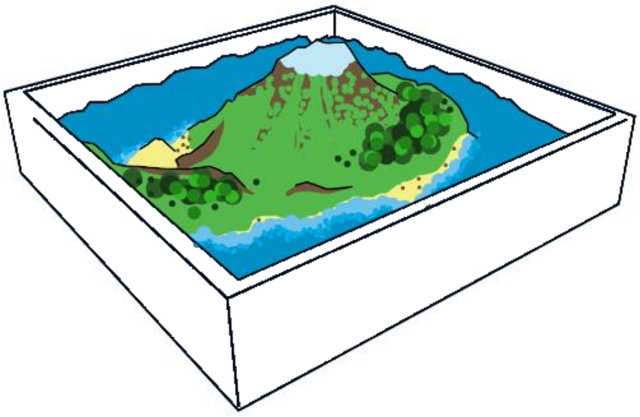
We consider that the proposed interaction techniques and artefacts can be extended to other contexts, and are not limited to SAR-VR collaboration. Indeed, the presented interactions could be used between any two locations, each of them either physical (digitalized) or virtual. Our conclusion is that, instead of trying to provide a single space that supports all the interactions, it is preferable to have several spaces with specific aims (much as we do in the physical world), even when these locations are only virtually separated.

PART III: Collaboration, wellbeing, and other possible applications

8. Introspectibles and Inner Garden

Artefacts that support introspection and mindfulness exercises

"We are a way for the cosmos to know itself."
— Carl Sagan, Cosmos



About this section:

This Chapter presents the general notion of *Introspectibles* (augmented tangible artefacts for introspection), and then provides a concrete instance of this with Inner Garden (a sandbox designed for introspection and mindfulness practices). *Inner Garden* was created using the same technologies presented so far, designed through iterations with meditation experts and practitioners. Our takeaway from this process is that, when designing carefully, it is possible to use technology to support wellbeing practices (which is counterintuitive at first).

Inner Garden was published first as a work-in-progress at ACM TEI'16 [**TEI16a**], and after at CHI'17 [**CHI17**] (honourable mention), while the *Introspectibles* have been discussed in workshops [**CHI4G16**, **ETIS16**]. These projects are the result of the collaboration with ***Renaud Gervais*** and ***Jérémy Frey***.

8.1. The Context that led to Introspectibles

This chapter discusses the usage of physiological sensors in combination with tangible interaction as a way to foster wellbeing. To this end, we must first introduce Positive Computing and Physiological computing:

- **Positive computing** is based in the notion that computers are not simply devices we interact with, but can also be used as agents of positive changes in our lives. As a field, it sits at the intersection between HCI [30,166], positive psychology [180] and design [43].
- **Physiological computing** [30] uses physiological data from the users as one of the interaction components of the system, both *actively* (the user can use the physiological activity to control the system in real time) and *passively* (we can adapt the system state using the user's physiological data, even if the user is not aware of its current internal state).

8.1.1. Quantification vs Qualitative information

As with other kind of data, we can compute statistics and feed them back to the users. The Quantified Self (QS) [232] is an increasingly popular movement that consists in keeping a log of different metrics related to health or physical activity, which can be used to gain insights about one's own body. Lately, extensions of the QS to cognitive tasks has also been proposed [120]. We consider that the current approach to observing the collected resembles too much to performance measurements: numbers, charts and progressions over time (such a *Terminal World* view of the human).

A more calm approach is to provide qualitative, iconic or enactive representations of the physiological information. Different works relate to the tangible and social representations of this data, for example creating a 3D printed object based on a running session [110]. Bodily signals have been used in the context of calm technologies, for example SWARM [226] is a wearable with sensing capabilities to mediate affect. Similar in spirit with our system is BreathTray [139], an ambient desktop widget that help users to control their breathing patterns. Sonne and Jensen [189] created a game for children suffering from ADHD (Attention Deficit Hyperactivity Disorder) based on slow breathing patterns.

8.1.2. Physiology and Tangibles

The idea behind Introspectibles is the creation of physical artefacts with a digital overlay so as to keep user's awareness of the environment and themselves. This approach was first proposed by ***Renaud Gervais*** and ***Jérémy Frey***, who created tangible avatars on which physiological information is displayed in real time. This information can range from low level EEG information [63], to higher level states (such as relaxation and tiredness) [67]. They created a toolkit that allows others the creation of tangible animated representations of real-time physiological data. Extending their research, we worked together to define Introspectibles [**CHI4G16**, **ETIS16**]. This concept is better understood through an example.

> It is important to clarify that this collaboration raised from the complementarity of the profiles of the members: ***Jérémy Frey*** worked during his studies in HCI and BCI (Brain Computer Interface), and his strengths are in signal processing and partially in fabrication, and is very interested in interoception [54]. ***Renaud Gervais*** has a stronger design and wellbeing orientation, with technical knowledge in SAR and TUI. I learned a lot of what I know from them, with prior knowledge in interactivity and game design.

8.2. Happiness and Mindfulness

As technology becomes more and more ubiquitously available, it could be used not only to fulfil tasks, but also to help us to increase our subjective wellbeing (that is, feel happier). Yet, with the rapid advance of technology, we are not happier than 20 years ago [80].

Our position regarding this issue is that modern interfaces lack the *calm approach* [167] – instead of being constantly competing for the user's attention, technological devices should be available only when required –, putting humane experiences before the efficiency and connectivity. Regarding this topic, **slow design** can a source of inspiration. Slow design focuses on designing activities that are worth paying attention to, and that happen at a deliberately slow pace. Instead of trying to finish a task as fast as possible, making the experience enjoyable by deliberately extending the length of the pleasurable aspects, and capturing the attention of the user.

8.2.1. What is Happiness?

Subjective wellbeing (or *happiness*) is a multifaceted concept often described as a combination of *hedonism* (i.e., maximizing pleasure) and *eudaimonia* (i.e., developing our human potential) [170]. Increasing happiness can be then achieved by a combination of both *enjoyable experiences now*, and *activities that help fulfil human aspirations* (as explained by the *self-determination theory* [66] or *intrinsic aspirations* [170]), the latter having more lasting effects than the former. There is a multitude of factors that have been shown to contribute to happiness and could benefit from technology [30], for instance: positive emotions, motivation and engagement, self-awareness, and mindfulness. Among these, this work focuses on mindfulness practices.

8.2.2. And what is Mindfulness?

Mindfulness can be considered as the opposite of mind wandering, which has been demonstrated to cause unhappiness [111]. Kabat-Zinn defines it as *“the awareness that emerges through paying attention on purpose, in the present moment, and non-judgmentally to the unfolding of experience moment by moment”* [100]. While mainly used in Buddhist traditions, it has been adapted and brought to a more western audience a few decades ago. Since then, it has been empirically tested and proved to be a good tool to train attention [23], for health care, mental illness and education [225]. More specifically, it has been integrated in an eight-week program course – Mindfulness-Based Stress Reduction (MBSR) [73] – to help people cope with stress, pain and depression [73]. More recently, the Mindfulness-Based Cognitive Therapy (MBCT) program has been proposed and has proven effective in treating depression, preventing relapse and promoting wellbeing [179].

While the mental and physical health benefits of mindfulness have been demonstrated, mindful practices are not easy to integrate into a daily routine. One of the most common said practices is mindfulness meditation. Similar to physical exercises, attention training requires practice and efforts, however, unlike physical activity – for which we can design playful, competitive or compelling settings that contributes to motivate its practice (e.g. sports) [74] –, meditation is a tricky activity to make engaging. Trying to contemplate without judgment or specific goals feels idle, like *actively doing nothing*. For novices, mindfulness meditation can be deceptively hard to practice for more than a few minutes without being distracted, yet involving any external source of motivation would be contradictory with mindfulness itself. The challenge we face in then:

How to design artefacts that are both intrinsically engaging and foster mindfulness practices?

8.3. Introducing Inner Garden

Inner Garden is a multi-modal tangible artefact that takes the form of an augmented sandbox depicting a small world (Figure 55), designed to implicitly support mindfulness exercises. The garden combines Physiological Computing with Spatial Augmented Reality, Tangible Interaction and immersive Virtual experiences. The Inner Garden's metaphor is anchored in reality using the user's physiological activity and actual sand. It is designed both as an ambient display for self-monitoring and as a support for more involved mindfulness exercises: *breath* and *body* awareness. The user can shape the terrain with their hands or tools, and affect its evolution with their internal state. By contemplating and interacting with the garden, the user receives biofeedback (e.g. the breathing controls the waves, changing the sea level and its sound accordingly). This gives a gentle-yet-constant reminder of their own bodily activity, providing an anchor to the present moment. Using a head-mounted display, the user can travel *inside* their own garden, for a moment of solitude and meditation, still accompanied by the biofeedback.

The garden was strongly influenced by video-games, particularly god-games like Black and White [129] where the user can influence the virtual world without having complete control over it. However, a game is often based on goals, explicit constraints and rules, which are mostly antithetical with the main concept of non-striving promoted by mindfulness [30]. We instead opted to create a toy [36], which uses the users' self-motivation and curiosity as main driving forces.

The system was created over multiple design iterations. During the process, we presented it to five experts, either meditation experts or medical practitioners initiated in meditation practices, and collected their feedback. Moreover, we also collected feedback from 12 potential end users of varying levels of experience.

The three main contributions of this work are: (1) a prototype leveraging Spatial Augmented Reality, Tangible Interaction, Virtual Reality and Physiological Computing for the purpose of supporting mindfulness practices (namely, *breath* and *body awareness*), (2) design considerations and takeaways from iterations with experts, and (3) interviews with meditation practitioners that tested the system.

Figure 55: Inner garden is an augmented sandbox, creating a landscape connected to the user's internal state

8.4. Specific Related Work

This work sits at the intersection of different fields: it builds on top of technologies for mindfulness, borrows insights from games, and takes the shape of a mixed reality sandbox.

8.4.1. Technologies for Mindfulness

Most technologies that support mindfulness activities rely on guided sessions [78,187] or leverage social networking features [136], most often delivered through the web or mobile applications. Recently, Virtual Reality (VR) has been used to deliver more immersive experiences, using either guided sessions [38], brain computer interfaces [3] or both [116]; other VR applications provide procedural landscapes based on the user inner activity [41,203]. Further along the spectrum of reality-virtuality, we find the work of Vidyarthi et al. [215], who proposed *Sonic Cradle*, where users sit on a hammock in a sound-proof room where their breathing patterns were amplified through sounds. This installation focuses on the embodiment of inner activity to anchor the user into reality, yet it requires a dedicated room, thus it is not part of the user's daily environment. Finally, there are mobile applications that explicitly explore attention exercises [173].

It is possible to divide the previous technologies into 3 groups: 1) always-available mobile applications, 2) immersive VR sessions, 3) and dedicated rooms with embodied installations. With Inner Garden we propose a middle ground between these options, with a locally-available ambient display which provides an embodied experience, with optional immersive sessions.

8.4.2. Engagement: Gamification, Toys, and Toy-ish Video Games

Gamification has caught a lot of attention from research [178] and commercial [88] communities in the past few years. The core principle is to take what makes a game engaging, and use it outside of the traditional gaming context; this can be useful when trying to design an engaging artefact. Gamification studies why people play games; it was initially focused on *extrinsic motivation* – based on external rewards such as prizes or scores –, and lately moved towards *intrinsic motivation* – activities that are interesting by themselves. Both extrinsic and intrinsic motivation have been used to foster engagement [103] and healthy habits [74]. In our case, extrinsic motivation can be clearly seen as contradictory with mindfulness, as it is based on goal setting and comparison with others. On the other hand, intrinsic motivation models such as *need driven* (competence, autonomy and relatedness) [155] or *RAMP* (Relatedness, Autonomy, Mastery and Purpose) [132] overlap with the *Ryff's* *eudaimonic* *intrinsic aspirations* (positive relations with others, autonomy, self-acceptance, environmental mastery, purpose in life, personal growth) [170] (Figure 56). These related intrinsic motivators could be used to increase the engagement of an artefact for mindfulness, as long as they do not contradict mindfulness principles; additionally, it is possible to directly drive inspiration from games that apply these principles.

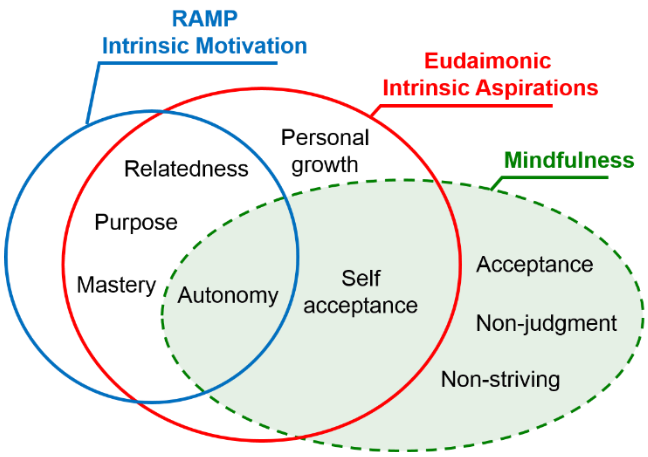


Figure 56: overlapping between RAMP intrinsic motivation model, the eudaimonic intrinsic aspirations, and the aspects of mindfulness. This shows that gamification is not opposite to mindfulness, and if designed carefully other features of gamification could be used to design mindful applications.

Toys have been used in the literature to study self-motivated activities. Paulos et al. [152] created toys to encourage children to explore their physical environment. Another example is the work of Karlesky et al. [104] who created seemingly meaningless tangible toys in order to explore the interaction that happens in the margins of creative work.

Some commercial video games remove goals and striving from the gameplay, shifting the main driving force from challenge and reward to wonder. Open worlds such as *Minecraft* [138] and *Proteus* [51] provide the player a range of control over the environment, without any explicit goals. Minimalistic games such as *Viridi* [90] present gardening as a relaxing activity without an end goal. In artistic games like the iconic *Mountain* [149], the players are invited to simply contemplate the beautiful landscape, with a minimum amount of interactivity. Some of these games also possess reduced to non-existent Graphical User Interface, forcing the user to discover the functionality, and to move the focus from the interface to the experience. It could be argued that some of the mentioned games fall into the toy category [36].

8.4.3. Augmented Sandboxes

This project uses as a support a projection-augmented sandbox, a concept previously explored in the literature. *Illuminating Clay* [154] has been the first system to combine the two elements. It uses clay to represent a landscape which can be directly shaped with the hands. The result of landscape analysis functions is displayed directly on the clay. In the artistic project *EfectoMariposa* [70], the projected simulated world is alive and evolves on its own (elements in the mini-world evolve without any user intervention). Dynamic augmented worlds can also be navigated virtually: in *MadSand* multiplayer game [169], one of the players creates the mini-world using an augmented sandbox while a second user navigates it in real time using a computer.

As was previously discussed, interfaces involving SAR and TUI make abstract topics easier to grasp by enabling the user to experiment hands on [182]. In our case, we present the user their own inner state, giving life to a small world: an embodied meditation metaphor.

* * * * *

8.5. Inner Garden: Introspection and mindfulness

Inner Garden is inspired by both the reflective and metaphoric nature of Zen gardens as well as the playful and experimental nature of sandboxes. Zen gardens are all about careful placement of elements and are often used for contemplative and meditative purposes. On the other hand, sandboxes call for interaction and experimentation. We purposely used a gardening metaphor to encourage the *continuous practice* of mindfulness; Inner Garden grows with time and with each meditation session a user does. Indeed, mindfulness is not about the quantification of the current experience (e.g. how you performed when watering your garden today), but is instead about the ongoing process of being able to identify a wandering mind and returning it to the present moment (e.g. watering your plants every day). The gardening metaphor puts the user in charge of a living artefact, and it provides a source of *relatedness* (a driving human aspiration) [42].

The garden's sandbox contains polymeric sand (sand with an added polymer that keeps the *"wet sand texture"*), which can be reshaped at any time using bare hands or tools and determines the topology of the terrain. The real-time simulation handles the generation and evolution of the living mini-world, which is then rendered on the sand using projection. The lowest areas are filled with virtual water, which level is mapped to a breathing sensor attached to the user. Breathing patterns therefore generate waves, producing subtle sound effects audible even when the user is not directing his or her attention on the garden. The user can simply contemplate the mini-world, the sea moving with his or her breathing and the clouds slowly passing by. Finally, the user can decide to go *inside* the mini-world for a moment of solitude [64], selecting a location by placing a mini-avatar in the sandbox (Figure 57), and then using a VR helmet. During the immersive session, the user will find him- or herself sitting at the corresponding location in the garden, facing a campfire that intensifies slightly when the user is breathing out (Figure 57-right). At all times, biofeedback is part of the nature itself, as a gentle reminder to focus on his or her internal state.

Figure 57: users can shape the terrain with their hands (left), chose a location (centre) and then immerse themselves thanks to an HMD (right). Biofeedback is present at all times, gently influencing the mini-world.

* * * * *

The living mini-world is composed of different elements, each one evolving at a different pace. It consists of the terrain, the sea, life in the mini-world, weather conditions and time. These layers are depicted in Figure 58, and based on *EfectoMariposa* [70].

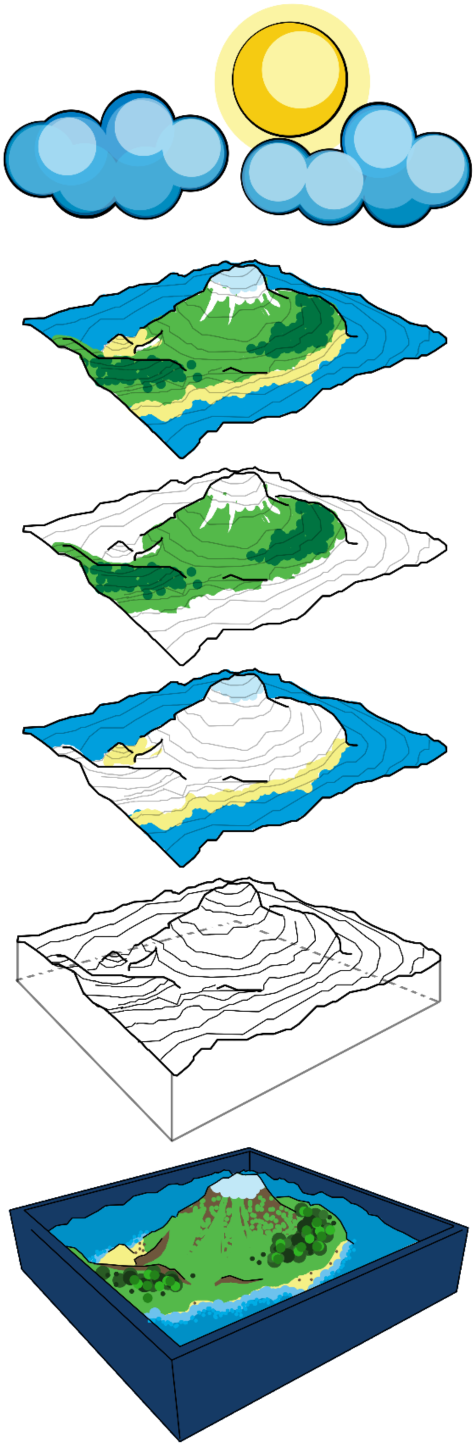


Figure 58: the layers of the simulation, from the bottom up: terrain, sea, flora and fauna, combined together and affected by the weather and time.

Terrain: The sand defines the topology of the mini-world. According to the elevation and slope, sections of terrain with specific properties are defined. For example, if the slope is too steep, no plants will grow; if the terrain is too high, snow will accumulate.

Sea: The sea level divides what is above and what is underwater. This level is a fixed value that is set according to the height of the sandbox. Sections of terrain that surround the sea level will not contain flora, having a beach instead.

Life: Two types of life form exists in the mini-world: flora and fauna. The flora consists of grass and trees that grow over time. The fauna, while not explicitly made visible in the mini-world, manifests its presence with sounds (birds or crickets).

Weather: The weather elements that are simulated in the mini-world include wind and clouds. The wind is one of the central aspects as it is directly linked to breathing. It also impacts other elements of the mini-world. Namely, clouds move according to the wind's speed. In VR, it also directly affects the movement of the trees and the intensity of the campfire.

Time: The mini-world has its own internal clock and the passing of time can be made constant (e.g. following the local time) or variable. The clock defines the amount of light and the sun position, which in turn affects how the clouds shadows appear. The day/night cycle also influences the active fauna (birds during the day, crickets during the night).

Time can be used in different ways for different purposes. For one, time of day has aesthetic properties. For example, some persons may prefer the intimate setting of a starry sky along a campfire while others may enjoy a more energizing virtual sun bath. The clock can also be used as an ambient timer for a meditation exercise. For instance, a user can start the exercise at the beginning of the garden's day and stop when the night comes, 10 minutes later (instead of using an alarm to indicate the end of the session).

8.5.1. Physiological Signals

The system supports physiological signals using the TOBE framework [67]. Inner Garden evolves and reacts to the user's real-time physiological measurements. It acts as both an instant biofeedback device as well as a long term and motivating ambient support. The measured physiological activities and their mapping to the dynamic elements of the mini-world open a wide number of design possibilities. We present the ones we selected and the rationale behind them (Figure 59):

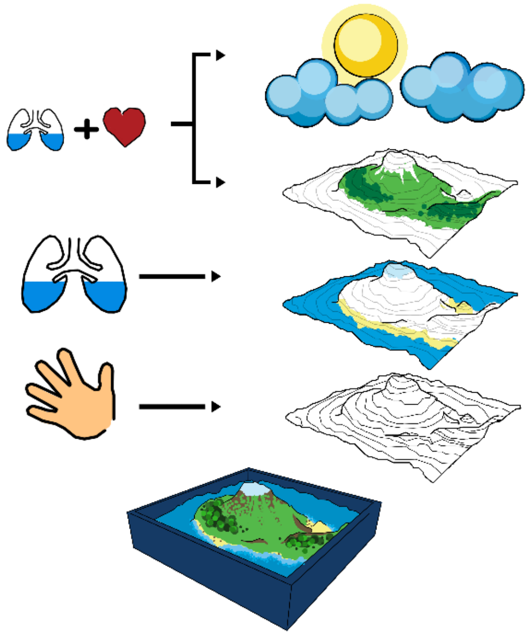


Figure 59: The layers of the Inner Garden and their connection with the user. From bottom to top: the sand is manipulated by hand; the level of the sea is controlled by breathing; vegetation growth and weather depend on cardiac coherence.

Breathing: The oscillating breathing patterns are mapped to the water level (which creates waves), and to the wind strength (creating gusts of wind). These elements are naturally periodic in the real-world, creating a coherent experience. All the experts we interviewed found the water biofeedback interesting since they use similar metaphors in their meditation sessions.

Cardiac coherence (CC): This metric is the positive correlation between the fluctuation of the heart-rate and the breathing cycles, when taking regular, slow and deep breaths. The resulting state has positive impact on wellbeing [134] and it is associated with a relaxed state. CC is a metric that varies slowly over time, therefore we used it as a subtle indicator of the overall “healthiness” of the garden. When in good health, the amount of clouds reduces, flora's growth speed is increased and sounds caused by the fauna are more present. Note that there is no obvious “unhealthy” state; fauna might hide, but trees will not start dying as a result of a low CC score. This is to avoid the introduction of judgmental effect. CC might be correlated with a relaxed state, but this is not a requirement for practicing mindfulness.

We also considered and discarded two additional sources of feedback:

Heartbeats: We avoided direct feedback of the heart beats, since the user cannot control them directly. Moreover, because the rhythm is naturally fast, we considered it could lead to unnecessary discomfort. One of the interviewed experts pointed out that since mindfulness is about acceptance, observing a state like stress should not trigger “negative” biofeedback loops, yet their design should be considered carefully, especially for novice users. We envision the possibility of using heart beats to controls different rhythmic components of the mini-world, such as ambient music.

Electroencephalography (EEG): EEG measures brain activity, and it can detect high level mental states, such as emotional valence and arousal (i.e. if the user is experiencing positive or negative emotions, and how intense they are) [235] or the level of attention. Previous versions of Inner Garden included EEG, but it was excluded from the current version given both the inaccuracy of current methods and the difficulties to equip and calibrate them outside of the lab. In a foreseeable future the detection of high level mental states will be more robust, and EEG electrodes could be embedded in the Head Mounted Displays, enabling its use during immersive sessions, as done in [116].

8.5.2. Designing for Mindfulness

While designing Inner Garden, we came across different problems and considerations about supporting mindful practices and experiences. As part of the design iterations we tested the system with five experts, either in meditation or medical practitioners initiated in meditation practices. Their diverse backgrounds (Table 7) allowed us to gather feedback from different point of views.

EXPERT ID	BACKGROUND
E1_Teacher	Transcendental and mindfulness meditation teacher
E2_Psychomotor	Psychologist and psychomotor therapist
E3_MBSR	Head of MBSR centre
E4_Psychologist	Psychologist interested in cardiac coherence
E5_Buddhist	Lama of a Buddhist centre (Buddhist monk)

Table 7: The experts interviewed, along with their backgrounds

Each of the experts was invited for a private session. We described the system and then equipped them with the breathing sensor. After an incremental demonstration of the available features, they were left free to experiment, usually making spontaneous comments in the process. After that, we performed an unstructured interview to know more about their experience and their opinion about the system as a tool for mindfulness. After each session the system was updated taking into account the obtained comments. The resulting design considerations build on the work of Calvo and Peters [30], and are presented below.

2.3.1 Distraction vs Guidance

There are two main approaches to support mindfulness, either provide direct guidance (e.g. external stimuli that point towards the breathing) or to train the required abilities (e.g. requiring focus on the breathing even in the presence of distractions). We chose to include both in a very subtle manner. The sea and wind present aural and visual feedback as a constant reminder of the breathing while animal sounds provide distraction from it. Moreover, since CC is based on consistent slow breathing cycles, the stop in animal sounds provides a feedback that the breathing frequency is becoming inconsistent.

The pace at which the simulation runs and at which change happens needs to be carefully considered. Like guided meditations, some moments where nothing is happening or heard is required for people to focus, before offering a reminder to come back to the breath. E2_Psychomotor found some of the sounds a little too erratic and distracting while E1_Teacher and E3_MBSR considered these as interesting elements that could provide exercise to train attention. This tends to show that the pace at which events happens should be slow and progressive. One of the main considerations of E1_Teacher for using Inner Garden in his teaching was how to frame an exercise without any external control. He suggested using the day/night circles to define the length of the exercises, while preventing ending the exercises abruptly (e.g. at nightfall). One interesting aspect for him was the use of naturally occurring events as a timer allowed the user to continue with the exercise if desired.

2.3.2 Keeping it Minimalist

A minimalist design can create an environment that has very few distractions, enabling the user to focus. Carefully chosen ambient biofeedback that is thematically built into the system – i.e. nature elements moving on a deserted island – provides a unified minimalist experience. E1_Teacher found the setup visually appealing but sometime hard to focus on. The constantly changing terrain was too detailed and dynamic for him. He suggested more minimalistic textures and slower movements. To address this issue, we reduced the overall speed of the clouds, the range of the sea waves, and the contrast of the textures.

2.3.3 Non-judgment and Non-striving

One of the core aspect of mindfulness is trying to be non-judgmental of your thoughts and adopting a non-striving attitude. We found these aspects were the trickiest part to design for, as one of the main motivation for creating a system supporting mindfulness is in making the activity more attractive and engaging. Usually, to increase engagement, goal-setting is efficient. However, as previously mentioned, it is relatively incompatible with a non-striving attitude. This is why we based its design on exploratory open-ended toys.

Even when goal-setting would be contradictory with mindfulness, E2_Psychomotor mentioned that having a way to tell how well the exercise was performed could help novice users. One way to get around goal setting is to provide a sense of progression related to the practice of the activity and avoid any explicit evaluation of how well a session went. This is also why we consciously avoided any quantified data and instead provided qualified feedback that is also fleeting – i.e. there is no record of your past breathing patterns or how your cardiac coherence evolved throughout the session. Moreover, to make the practice of mindfulness motivating every day, we added a localized biomass intensifying mechanism. Every time the user chooses to meditate at a specific area in the garden, it acts as a “watering system” for the biomass located there. That is, grass will grow greener and the trees taller, leveraging once again the gardening metaphor. It is based on the completion of an exercise (a given amount of time), and it avoids any negative progression: E1_Teacher insisted that the garden should not dry out if the user misses a day or two of practice – it should simply stay in the same state. This is also backed by the literature, that indicates that positive-only feedback (or progression in our case) is better for novices [77], and would satisfy the eudaimonic personal-growth aspiration [170].

2.3.4 Promoting Acceptance

Exercising acceptance – accepting what is happening, in contrast with the desire or ability to change it – is a key aspect of mindfulness. To train acceptance, the user must acknowledge that the garden will never be complete, and every time he or she modifies the terrain, the life will reset. The user must accept that there are consequences of his or her actions that are out of his or her control. Another way acceptance is exercised is with the immersive session. The user selects where to do the session with a tangible avatar before starting. Then, they can contemplate the surroundings, but cannot navigate the virtual space.

The biofeedback is itself not filtered allowing the user to face his or her own internal states, e.g. if a user starts breathing inconsistently, the water will reproduce this behaviour. E3_MBSR considered that the user should exercise the ability to observe negative states (such as stress) without triggering a negative feedback loop, which is related with eudaimonic's self-acceptance aspiration. Even when this seems opposite to what E1_Teacher recommended in the previous section (no negative progression), they are complementary aspects: the user is exposed to both positive and negative states, but only the positive states (daily practice) have a lasting impact on the system.

2.3.5 Promoting Autonomy

This aspect is a core component in both intrinsic motivation and eudaimonic development. It is important to allow the user to control the boundaries of the experience along with when and how the feedback occurs. Ambient information is well suited for this, since it does not impose on the user. Therefore, the user has control over how fine grained he or she needs the information retrieval to be. A more directed attention enables more precise information to be derived while a soft focus will let unconscious mental processes do the monitoring. Moreover, sound is a modality that is easy to either tone down or disable by the way of volume control, if the feedback turns out to be too invasive. When we asked the experts about offering the possibility of controlling the moment of the day in the garden to suit individual preferences in meditation settings, they kept referring to such possibilities as “very ludic”, but were concerned that too much control and options will prevent the user from exercising acceptance. E1_Teacher and E3_MBSR said that such customization could be useful to define a starting point for the meditation session.

2.3.6 Using Tangible Interaction

A lot of attention-directing exercises leverage the body and its sensation to stay in the present moment (e.g. yoga, guided meditation, tai chi). E5_Buddhist was concerned about the increasingly prevalent disconnection from the body in favour of the world of the mind and thoughts. This expert especially liked the use of the sand in Inner Garden as a tangible, direct and visceral way to reconnect to physical sensations, potentially “bypassing” the mental. E2_Psychomotor and E4_Psychologist found the SAR installation could be particularly useful in physiotherapy, in order to let the patient “get back” to their body. While trying the system, all experts enjoyed doing back and forths between performing modifications on the landscape and periods of contemplation.

2.3.7 Choosing the Right Reality

The use of Virtual Reality (VR) or See-Through Augmented Reality (ST-AR) helmets can provide immersive experiences that can be leveraged in interesting ways for mindfulness purposes, and were used in the past [3,38,116] for this objective. For example, these technologies can prove useful when looking for moment of solitude [64], by blocking or tuning out external stimuli. However, completely traveling to an alternate reality seems to go against the “acceptance” of reality described earlier. On this matter, E4_Psychologist also raised concerns on using a VR experience in cases where user has psychological troubles to differentiate what is real and what is not, such as persons suffering from schizophrenia. For these reasons, we made sure that the system was designed around a SAR experience. SAR is interesting because it directly uses the physical world as support and provides a shared experience that does not require users to wear any equipment, unlike ST-AR systems. From this experience anchored in reality, we added a VR modality that allows users to travel to a spot inside their garden. Because of this AR-VR transition, the VR experience is anchored on a real-world element: before the visit, the user built the virtual world with his or her own hands, and placed an avatar on the location they want to travel to. Moreover, even if the virtual world is completely computer generated, the biofeedback is always present, keeping the user aware of his or her own body. For those cases where VR is not recommended, such as with young children, SAR is a very interesting alternative to explore.

During the iterations we experimented with an intermediate Augmented Virtuality modality, where it is possible to see and touch the sand while wearing the VR helmet. This was accomplished by rendering the scene from the user's head position, and using a Leap Motion (attached to the front of the helmet) to provide feedback about the hands position (Figure 60). The resulting hybrid modality mixes both the immersive experience with the tactile feedback from the sand. Experts were concerned that this feature will render the system “too ludic” for mindfulness (i.e., too much playfulness can disperse the user's attention, which is the opposite of the system's objective). This concern might be due to the high degree of interactivity with the virtual elements, which drives away attention from the physically based elements. This modality, while promising, requires further consideration.

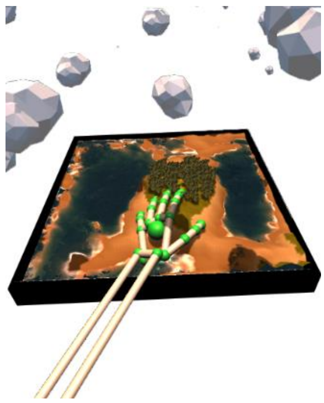


Figure 60: augmented virtuality might not be suited for mindfulness practices.

8.5.3. Design Overview

The starting point of Inner Garden took inspiration from Zen gardens, and it was built on top of existing mindfulness applications, leveraging tangible artefacts. The exchanges with experts either validated or complemented the design considerations we found in the literature, and enabled us through an iterative process to create an engaging artefact for mindfulness. We consider the lessons learned in the process can be of use not just for the reproduction of the system (detailed in the next section), but also when facing the design of novel positive computing artefacts.

8.6. System Implementation

The final version of the system is composed of five modules: projector-camera calibration, surface scanning, physiological controller, simulation and rendering (Figure 61).

The components were implemented using 3 frameworks: vvvv [238], Unity3D v5 [211] and OpenViBE[164]. The whole installation runs on two computers: the graphical pipeline and simulations runs at 80 fps on an Intel i7 Windows 10 desktop computer equipped with an NVidia GTX980Ti graphic card, while the physiological controller runs on a dedicated laptop computer running Kubuntu 14.04 in order to reduce signal acquisition latency.

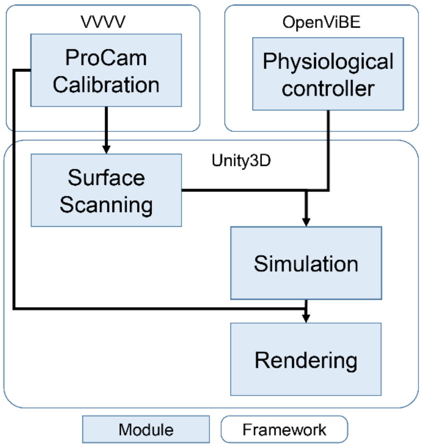


Figure 61: Inner Garden modules. The projector camera-calibration is performed offline, and the physiological controller runs on a dedicated computer. The main application performs the surface scanning, simulation and rendering, and it was implemented using Unity3D.

8.6.1. Projector-Camera Calibration

The SAR setup is comprised of an ASUS short-throw projector to augment the sand and a Microsoft Kinect v1 (Figure 62) to capture the sand's topology. The projector's properties were calibrated using OpenCV [22] camera calibration tools (2D to 3D calibration), while the Kinect extrinsics were calibrated using the Kabsch algorithm (3D to 3D alignment: rotation and translation).

The sandbox was used as the common frame of reference. The calibration is performed offline. We used a 40 × 40 × 4cm sandbox. Different shapes could be studied, notably in the case of multiple users where circular shapes could ease collaboration.

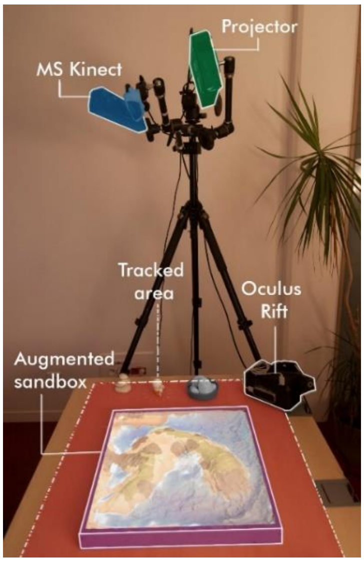


Figure 62: Installation setup: Projector (green), Kinect (blue) tracking the augmented sandbox (purple) and tokens over the table (red zone). The VR headset equipped with a leap motion stands at the back of the table.

8.6.2. Surface Scanning

The acquisition of the Kinect information (Figure 63) is performed on a worker thread on the CPU, while the computation of topological information is performed on the GPU. First, the Kinect's depth information is converted into a point-cloud (sparse world positions) using the depth camera's field of view. Then, the Kinect's extrinsic calibration is applied, to obtain the point-cloud in the sandbox referential. Being in the same coordinate system, it is possible to identify the points that are inside of the box, or on areas of interest around it. Then, topological information can be computed: for every point of interest, it is possible to know the height from the base, and the associated normal along with the time from its last modification. Because the Kinect's depth information is noisy, changes under a threshold of 2 cm are ignored.

The detection of the tangible avatar is done using the height information. The avatar can be detected in one of two locations: 1) on a predefined area near the sandbox, or 2) on the surface of the terrain. In the first case, the terrain is updated continuously. The moment the avatar is removed from the preselected position, the update of the terrain geometry is paused. From this point on, significant changes on the height map are assumed to belong to the avatar

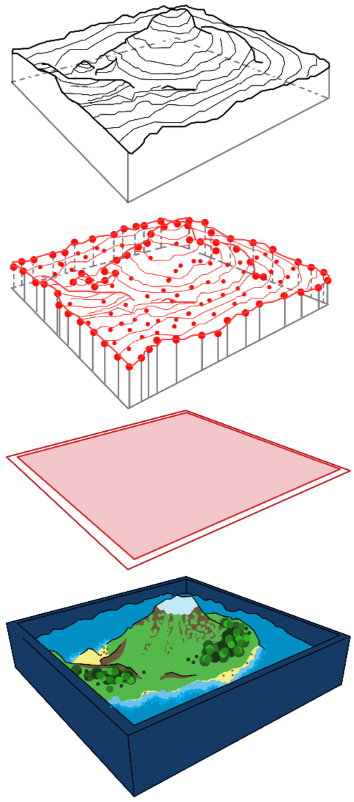


Figure 63: From the bottom up, how the tracking information is performed. Starting from a redefined region of interest, the Kinect point cloud is segmented, then the per-point height and normal are computed.

8.6.3. Physiological Controller

The physiological controller uses the TOBE framework [67], including both the sensors and the signal processing software. In terms of sensors, we used a Mio Fuse smartwatch for measuring heart beats and a homemade belt based on a stretch sensor to measure breathing. The latter is comprised of a conductive rubber band that was mounted as a voltage divider and connected to an instrumentation amplifier. It was then connected to a custom printed circuit board (PCB) that is directly embedded into the breathing belt. The PCB has a Bluetooth 4 module to stream the data to a laptop computer to be processed using OpenViBE, where the value is normalized using a moving window. Finally, the normalized information is sent to Unity3D using the LSL protocol, a network protocol dedicated to the streaming of physiological data[5].

8.6.4. Simulation

The simulation handles all the different elements of the garden's world (Figure 58). The generation of the mini-world's topology is created using the result of the surface scanning step. The topological information is used to classify the terrain in different region types (e.g. under water, above water,

[5] Connecting OpenViBE with Unity using LSL: https://github.com/sccn/labstreaminglayer

with snow, too steep to grow life). For each point of the sandbox, the growth of different kinds of vegetation is controlled by the time since its height changed, along with the region type. Therefore, each point will be assigned a colour that will result in the terrain texture. Then, assets such as trees are instantiated based on the terrain texture. Finally, virtual elements are added, such as clouds, sun position and stars. These virtual elements are visible while in VR, and they influence the illumination in SAR. The rendering is performed following the same approach as in ***Chapter 5***.

8.7. Interviews with Practitioners

To complement the feedback from experts, we conducted interviews with meditation practitioners to collect feedback from potential end users. Participants were recruited through the newsletter of a non-profit association that gathers people interested in meditation. 12 females took part in the study, mean age 45±SD: 11. Most of them (7) practiced meditation regularly – several times a week or every day – but only 3 explicitly mentioned mindfulness practices, in contrast with other meditation practices. The study comprised of 5 stages (Figure 64).

8.7.1. Protocol

For the interviews we used a preliminary version of the system, where the breathing belt was connected to an external OpenBCI board [148], and the cardiac coherence was inferred from the breathing patterns.

The participants entered a dimly lit room and filled out a consent form. Afterwards we introduced the features of the system progressively: shaping the sand, dynamic world, breathing (they were equipped at this moment), and finally the head mounted display. As we explained the breathing, the users could see their breathing projected onto the sand, they were also encouraged to touch it and play for a few moments, in order to see the connection.

After the system introduction, the participants filled several questionnaires so we could know more about their profile. The period during which they filled these questionnaires was used as the breathing baseline. The questionnaires were:

1. A demographic questionnaire
2. The five facet mindfulness questionnaire (FFMQ) [9]: a questionnaire evaluating how "mindful" they are during their everyday life
3. The state-trait anxiety inventory (STAI-YA) [191]: and a questionnaire assessing their present anxiety

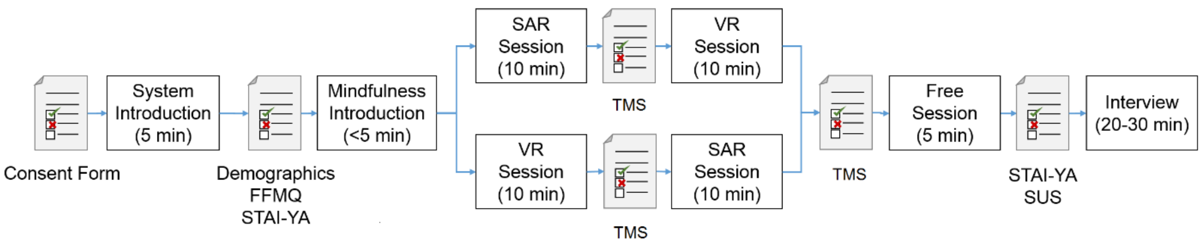


Figure 64: The 5 stages of the study and the questionnaires used. The stages involved: description of the system, two consecutive meditation sessions – SAR and VR, the order of the conditions counterbalanced between participants –, a free session, and finally an informal interview. Various questionnaires, measuring mindfulness and qualities of the system, were given during the study. Overall, a session lasted 1.5 hour.

Then we instructed the participants about the principles of mindfulness. On par with the instructors we met, we told participants that newcomers to mindfulness meditation could focus their attention on their breathing, taking deep and slow breath. Then the first meditation session occurred, using one of the SAR or VR modality to interact with the system. In the VR condition participants would pick their location using the token (as shown in Figure 57, page 101); in the SAR condition participants could freely manipulate the sand while meditating.

After 10 minutes – the length suggested by E1_Teacher – we interrupted the meditation exercise and participants filled the Toronto Mindfulness Scale (TMS), a questionnaire assessing how one was mindful during a specific exercise [121]. They had the opportunity to rest for a little while before we prompted them with a second meditation session, using the other interaction modality. Again, after 10 minutes they were interrupted and answered the TMS.

Before the exit questionnaires, participants were invited to use freely the system, without any guidance: they could shape sand, use VR or focus on breathing as they wished. When they finished to do so – usually after 5 min –, they filled three questionnaires:

1. The STAI-YA questionnaire, the same as the beginning of the study
2. The system usability scale (SUS) [24], which, as its name indicates, helps to evaluate the general usability of systems and services
3. Custom questions regarding the acceptability of specific aspects of the system

Finally, we conducted an informal interview with the participants, freely exchanging about the system for 20 to 30 minutes depending on the person. Questionnaires Results

Concerning the profile of the participants, the FFMQ questionnaire corroborated the meditation practices declared by the participants at the beginning of the study. Regular practitioners scored above the average of our population – on a scale ranging from 39 to 195, the mean of the 5 traits measured by the questionnaire was 146.2 ± 14.75. A Wilcoxon Signed-rank test yielded no significant difference $(p = 0.13)$ between the two interaction modalities concerning the TMS questionnaire – overall mean 34.91 ± 7.37, scale ranging from 0 to 48. This result may imply that the system induces a mindful state, regardless of the chosen modality. As for the STAI-YA questionnaire given before and after the exposition to the system, the same statistical test showed a significant difference; participants were less anxious at the end of the study (25 ± 5.23) as compared to the beginning $(31.50 \pm 10.71$, $p < 0.01)$ – scores between 20 and 80. Even though mindfulness meditation is not about reducing stress per se, our system may induce calmness. This is reflected by the breathing rate of the participants that was significantly lower during both mediation sessions, as compared to a baseline recorded at the beginning of the session (9 breaths per minute vs 17). The average SUS score was 80.82 ± 11.78 – scale between 0 and 100. Overall, this indicates that the system has good usability in both its SAR and VR modalities.

8.7.2. Feedback Discussion

The feedback we received was converging around recurring themes: the SAR and VR modalities and the multisensory experience.

4.3.1 VR and SAR

The participants were divided between which modality (SAR or VR) they liked the most – this was not caused for an order effect, since only 7 out of 12 participants preferred the first modality they tried. About half the participants mentioned VR as the best part of the experience, praising the immersion and the sense of control over the environment. Some participants, used to meditate with their eyes closed to fully concentrate, highlighted that the VR world enabled them not to be disturbed by external distractions while having something to look at. The other half of the participants (5/12) mentioned problems related with the heaviness of the VR headset and found the graphics too simplistic. One of the recurring comments was about the limitation of not being able to move in VR and explore the mini-world. This design choice was intentional, in order to prevent both simulator sickness and to foster acceptance, as discussed previously. However, it would be interesting to explore the potential benefits of virtual meditative walks [72]. The fact that the participants were split is a good indicator that providing complementary modalities can be used to suit personal preferences regarding meditation practices. Interestingly, when asked about the quality of the system's graphics, none of the participants mentioned the augmented sandbox; they only considered the VR as being a graphical display. This could indicate that using the physical world as a support can make technology more approachable, if not transparent. This is of special interest given the demographics (middle aged females, two of them explicitly stated their reluctance towards technology before starting the study).

Not all the users saw the SAR and VR components of the systems as parts of a whole. Notably, they saw discrepancies between what they thought they were building in the sandbox and what they witnessed in VR. The users that did see a connection (4/12) found the parts complementary, and highlighted being able to create the virtual world from the real one. The sea was mentioned as the strongest link between real and virtual. Visiting the virtual world can also increase the impact of the projected augmentation: one participant mentioned she could not understand the cloud shadows on the sandbox until she visited VR, and that after she imagined the clouds floating above the sandbox.

4.3.2 Multisensory Experience

We were interested in how the participants felt about the multisensory aspects of the system, namely the tactile, visual and auditory components. All the participants enjoyed playing with the sand. They were particularly pleased by the texture of the white polymeric sand. Similar to the experts, several of them felt that playing with sand had a “grounding” effect, beneficial to their mindfulness. We also had comments about how shaping sand was a freeing activity, reminiscent of when they were children (a known aspect, used in Sandplay [101]). To explore other possibilities, we provided samples of alternative materials (Figure 65). After trying the polymeric sand, the participants were not particularly inspired by the proposed substitutes, even regular sand (perhaps they were biased towards the first material they tried). On the other hand, they found the idea of using pebbles or rocks an interesting option to create heterogeneous landscapes, such as in traditional Zen gardens.

Figure 65: Shapes and textures presented to the subjects during the experiment. Top row, from left to right: big stones, pebbles, and texture cards. Bottom row, from left to right: normal sand, polymeric sand (natural), polymeric sand (white), insta-snow.

When asked if they were interested in customizing the appearance of the terrain, both the texture and the moment of the day or year, most participants liked the idea, but the ones with experience in mindful meditation were fast to notice that too much customization would be counterproductive for mindfulness, since it would not exercise their acceptance. They instead considered the possibility of an initial parametrization, and then letting the garden evolve slowly. This matches with the opinions gathered from the experts.

The participants were very pleased with the sea sound and being able to control it with the breathing, and felt it was synchronized with their inner state (9/12). They presented mixed emotions regarding the animal sounds: two of the participants, with backgrounds in music, found the overall sound-samples too monotone, and suggested adding a richer variety. This could be addressed using procedurally generated soundtracks.

For the study, the system layout was designed so that the participants sat comfortably on a rigid puff, while facing the sandbox that was placed over a table. Several participants found this position unusual for meditation, and asked if they could sit on the floor, or with their legs crossed. One participant also asked to move to a nearby couch for the VR session. Even when we tried to provide a comfortable environment for the participants, it would be interesting to evaluate it in a more ecological context.

4.3.3 Other Comments

Even when sensors we used were not designed for comfort, the participants did not mention them as inconvenient. The breathing belt limited the user's movement, which was not a problem while sitting. This could be still addressed using sensors such as Spire[6]. Finally, some participants confirmed what we observed on the questionnaires: the system created a sense of calm and feeling centred, in particular thanks to the sea and the sand. One of the participants (mindfulness meditation practitioner) sent us an email hours after the session to thank us, and commented she felt more relaxed during her work day. This might be caused simply because we provided a calm context to take a break from their routine, nevertheless we find these results promising.

8.7.3. Study Limitations

Even when the system was designed with novices in mind, the study was conducted with initiated meditators. The rationale behind this decision was that we first wanted to get feedback from practitioners to make sure that our design (and the use of technology) was not antithetical with their practices, to complement the feedback from experts. Even when we used questionnaires to measure the impact of the system, our main interest was to know more about their subjective experience. The preliminary results seem to indicate we are on the right track, laying the bases for a long term study.

Mindful meditation training requires longer periods of time, and measuring the impact of the system would require following novices along the training process. In the future we would like to conduct a long term study with novices, which will include a control condition, in collaboration with the mentioned meditation teachers. This would be vital to determine if the system can help or complement existing mindfulness exercises.

[6] We contacted the company beforehand, but their API was closed for real time measurements, which is a requirement for our system. Website: https://spire.io/

8.8. Conclusion

In this Chapter we introduced *Introspectibles:* mixed reality systems that leverages Tangible Interaction and Physiological Computing in order to support introspection. As an instance of Introspectibles, we then presented Inner Garden: a mixed reality sandbox designed to support mindful experiences. Our design considerations were based on the literature, complemented and validated with feedback of experts. Finally, we tested the system with meditation practitioners of different levels of experience (from initiated to daily meditators), which found the design engaging while also being well suited for mindfulness. Preliminary quantitative results seem to indicate that the system foster a calm and mindful state on the users.

We envision Inner Garden as an ambient device that would both provide a gentle reminder to practice mindfulness in daily life – like a plant that requires watering in an apartment – and a compelling tool to support mindfulness exercises. One of the main challenges will be keeping the system engaging, for example by introducing subtle changes over time such as seasons. Another interesting dimension to study is the potential of such a medium for social interaction around wellbeing. For example, multiple members of the family could take care of the same garden together, thus taking care of each other in the process. Both E1_Teacher and E3_MBSR perform group meditations as part of their practices and were interested in how it will be extended for multi-users, and explicitly proposed their interest to install the system on their practice. They were especially interested in its use to introduce novices to the practice of mindfulness. In the future we would like to study these aspects with their collaboration.

In the end, Inner Garden was not designed as the sole source of mindful experiences, but as an engaging way to pay attention to foster body and breath awareness. It serves as a first introduction to mindfulness, and ideally leads the users to train and develop a mindful perspective that will extend outside the garden, into their everyday life. For these purposes, tangible interaction is a promising approach. It has this interesting property of not looking like “real” digital technology, being closer to our physical and humane selves.

On a more general note, we find great promise in the usage of tangible interaction as a way to support wellbeing practices. TUIs in combination with Physiological computing can be used not only to control systems, but to ground the user onto the *physical world*.

PART III: Collaboration, wellbeing, and other possible applications

9. Objects that tell stories

Speculations of possible usages of the proposed approach to augment the essence of *things*

"Imagination will often carry us to worlds that never were [and perhaps never will]. But without it we go nowhere."
– Carl Sagan, Cosmos

About this section:

This brief chapter presents additional projects, the main idea explored involves the augmentation of the behaviour of everyday objects through the use of the same technologies presented so far.

9.1. More possibilities

In the spirit of the projects presented so far, there are other ideas that were explored but did not reach the academic publication process. These ideas are briefly presented in this Chapter, as we consider they show the potential of the proposed approach. The general concept behind these unfinished projects involves the exploration of augmented objects, as a medium to store and trigger immersive experiences.

9.2. Souvenirs: objects that help us remember

> *"Time moves in one direction, memory in another."*
> *– William Gibson*

Humans have the capability to use their environment as a support for information, what is referred to as extended cognition [199]. We can do this through language, and that is considered one of the most important human revolutions: because we can write, our ideas and knowledge can outlive our memory, and even ourselves. But not all memories are explicitly stored in external elements, sometimes an object is used as trigger of memories. These objects are often called *souvenirs* (a noun derived from the verb "to remember" in French). We collect these items when we go to places or experience moments we want to keep with us. And being objects, we can place them nearby, or hide them from sight if we want to forget… always keeping the power to bring that moment back. This souvenirs have a very intimate effect: the remembering process is only internal, we cannot share it unless we try to put into words what is happening.

With the invention of digital media, now we can (and do) store most of our lives digitally. Yet the media presents a very fragmented representation of our experiences. Indeed, showing someone a picture of our holidays does not have the same effect on them than it does on us. What if we could use technology to address this limitation, to use objects to relive an important moment? This is the idea behind *augmented souvenirs:* A physical object that is associated to a given location at a moment in time.

The reproduction of the information could be done using techniques similar to the ones presented in the previous chapters. The main limitation of such a project is then the acquisition of information and its linking to the physical object. There are two possible solutions: the usage of an external system that records at all times, much as the work presented by Rekimoto [107], or the creation of dedicated artefacts, much in the line of traditional postcards.

Figure 66: we use objects as a way to externally store memories, from gift store objects to found objects in the street. A simple rock can be enough to remind us a trip to the mountains.

9.2.1. Augmenting Post-cards

An augmented postcard (once again, the word augmented not as in a technology but instead "augmenting what makes a postcard a postcard") could record a short 360 degree video and audio of a moment, using e-ink or OLED to display it as a still frame or small loop. By interacting with the postcard (Figure 67), the audio could be triggered (a technology already available for traditional postcards), and opening the postcard can lead to a see-through effect (a navigation of the video much as it is possible nowadays using smartphones). Using augmented spaces as presented in ***One Reality***, the information could be extended from the post-cart onto the environment (perhaps using surround sound). Finally, the scene could be accessed using immersive HMDs, transporting the user into the moment the memory was recorded.

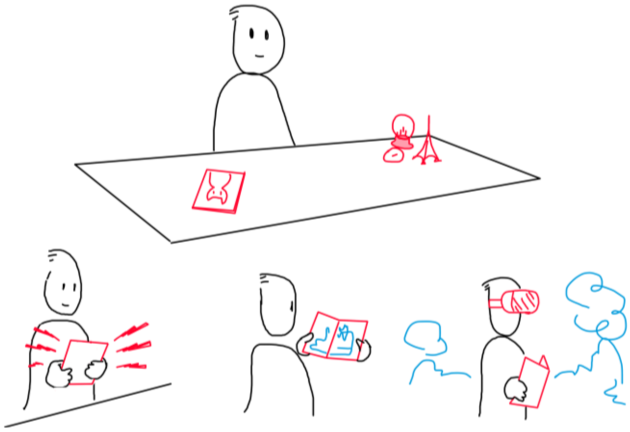


Figure 67: Augmented postcard behaviour: the postcard is always available (top), manipulation the postcard can trigger a first layer of information (audio, and fixed-view video, bottom-left), the card can be used as a see-through device (bottom-middle), or as a trigger for an immersive HMD application (bottom-right).

9.2.2. Arbitrary Souvenirs

Much like the postcard, arbitrary objects serve as triggers of memories. The augmentation of these objects is possible, yet the authoring and connection of the object and the memory are open issues.

If the authoring process is performed manually, then the artefacts can be used to trigger not only past memories but anything else, allowing the creation of storytelling and application implementation. This leads to the next project.

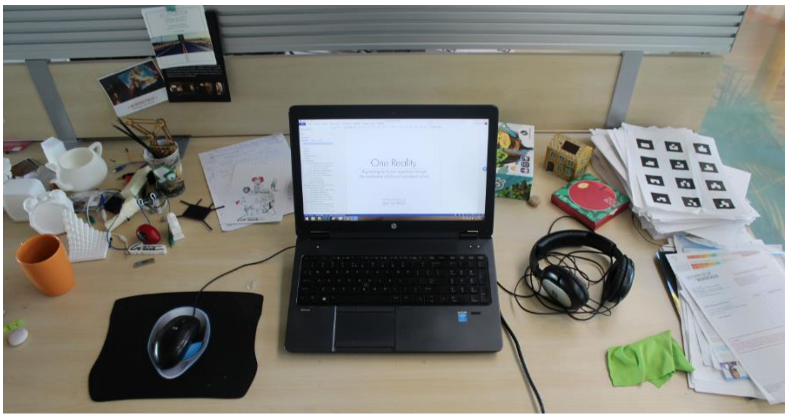


Figure 68: This messy place is my desk at work (after some sorting), it looks like that because every object reminds me of a project, or a place, or a moment. It would be ideal if I could share their stored memories with others.

9.3. Bookest: Books that are very good at *being books*

> *"What an astonishing thing a book is. It's a flat object made from a tree with flexible parts on which are imprinted lots of funny dark squiggles. But one glance at it and you're inside the mind of another person, maybe somebody dead for thousands of years. (…) A book is proof that humans are capable of working magic."*
>
> *- Carl Sagan. Cosmos, Part 11: The Persistence of Memory (1980)*

9.3.1. Books are more than just information containers

Books are not just artefacts to preserve and share knowledge and stories, they can give place to magical experiences. Computational devices offer a faster and more efficient way of obtaining information, yet it can be argued that the experience of reading a book cannot be always replaced by reading the digital counterpart, even with e-readers such as kindle (**Figure 69**).

The same way we can use mixed reality and embedded systems to store and replay events from the past, we can use them to create new stories. This way, it would be possible to extend the expressive characteristics of books and the reading experience. As the reader must be aware by now, the notion of augmented books is not new, as MagicBook [20] was mentioned several times during this manuscript as one of the main referents of hybrid mixed reality systems. The idea presented in this section involves not only "using a book as a display", but also to mention other aspects that could be augmented, in particular by driving inspiration from pop-up books.

First we must consider what defines a book (so to say, their *bookness*). Some features come to mind: their construction (a set of sheets stuck together, fragile against fire and water), in combination with the information dimensions (the information is hierarchical, it can be accessed linearly or randomly), their objective (sharing knowledge or trying to generate an emotional effect), and their specificity (each book is designed with a topic, the content has an intimate relationship with its layout).

9.3.2. A book from the future

When considering the potential of novel technologies, science fiction can be a source of inspiration. Neal Stephenson's Diamond Age [194] follows a girl and her computational book as they grow together. The book is presented as a storyteller and educator, while the reading process takes the form of a dialog. In order to support the interaction, the book has many amazing features: it can know where the user is looking, it can both listen and narrate, it can open and close by itself, and it can change its contents dynamically. Nowadays, 20 years after the first edition of Diamond Age, such an artefact could be prototyped.

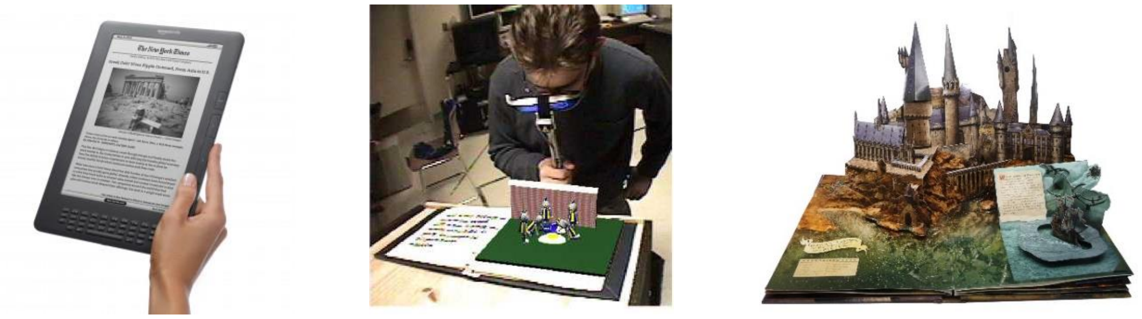


Figure 69: *unfair Comparison between an e-book (left), MagicBook (centre)* [20] *and a pop-up book (right)*

9.3.3. Current book technologies

Books can be technically complex. Among other features, they can have cut-outs, moving parts, interaction with light, sound and buttons, volumetric pieces and even circuits (**Figure 70**).

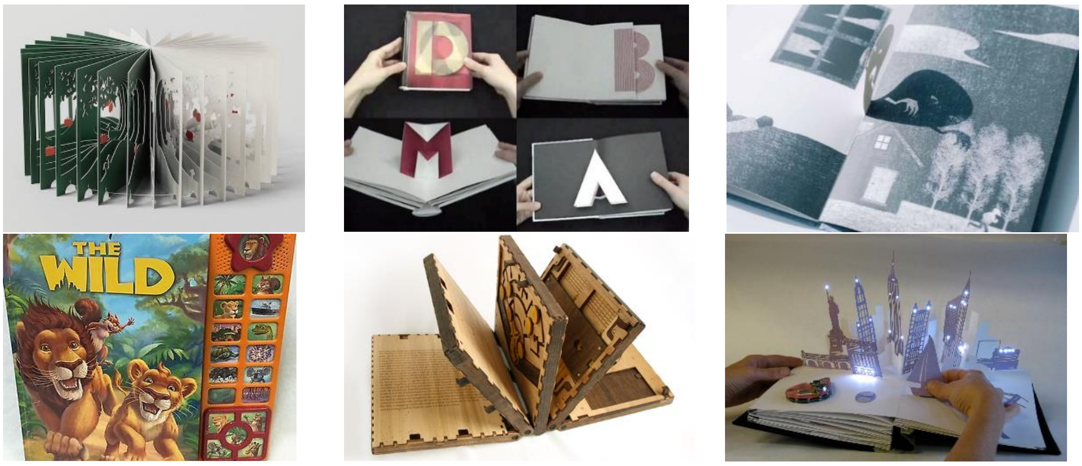


Figure 70: *Some examples of books that involve innovative technologies. Top row: 360 book (laser cutting), ABC3D (pop-up), Motion Silhouette (pop-up and shadows). Bottom row: Disney's The Wild (sound), Codex Silenda (laser-cutted mechanical puzzle book), Electronic Popables (electronic circuits and pop-up)* [156].

9.3.4. Dynamic content and awareness

One limitation of books is that once printed their content is static. Paper craft technologies enable interactivity to a certain extent. Paper electronics [192] in combination with printed screens [147] could provide a middle point between digital and printed. It is possible to envision more complex usages, such as the combination of ink and printed screens to create dynamic images, similar to *projectibles* [97], or to create dynamic narratives much in the spirit of *Choose Your Own Adventure* (books with many reading paths).

Pop-up technology provides mechanical interaction. Using jamming interfaces such as Electronic Popables [156] provided interactivity using traditional pop-up contraptions and electronic circuits. Foldio [146] technology could contribute this field in two ways: 1) interactors could enrich the catalog of possible pop-up elements[31], and 2) actuators could contain pop-up designs inside. The fact that the actuation technology is not particularly fast could be used as a feature, in the spirit of slow design.

Once a book has dynamic features, it is possible to make it react to the user. For example, it would be possible to use physiological sensors to measure stress or tiredness in real time. The book could then close itself if it considers the user should not read anymore. Certain events could also be triggered using Eye-tracking (actions that happen only when the reader reaches certain paragraph, or when the user is *not looking*). Actuation could play a role into making the reading activity challenging (e.g., the book could act as if a sudden storm started, when the reader starts reading an intense paragraph).

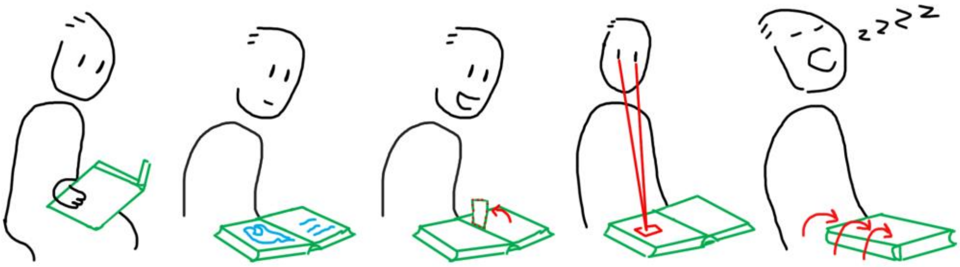


Figure 71: an augmented book could have dynamic printing, pop-ups, eye tracking and actuation based on physiology

9.3.5. The reading space

Even when reading transports us to a different realm, the space where we read has an important influence on the experience. Notably, books and some e-books lack self-illumination, so ambient light and lamps play a key role. This opens possibilities, for example the creation of digital text tools such as highlight and dictionary [175], or the control of the ambient light according to the mood of the text currently being read (**Figure 72**).

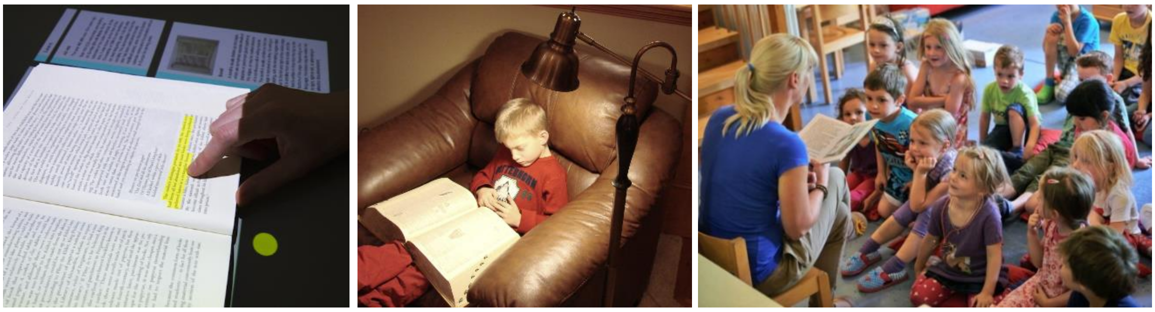


Figure 72: *Left: Enhanced reading to support learning* [175]. *Centre: an augmentation opportunity. Right: the context of supporting storytelling with dynamic ambient light, as described at the TEI'16 student session (sadly, it was not part of the proceedings).*

As augmentation and paper has been extensively explored, such possibilities are not new, yet a collaboration between a book and its environment in the process of creating of an experience is very promising.

9.4. Closing remarks

General purpose technology cannot replace the experience we get from function-specific artefacts. This does not mean that technology cannot be used to rethink classical artefacts, as long as we preserve what makes them unique.

These projects are part of many unfinished ones, yet it is never good to extend a presentation too much. For this reason, we are now at the end of ***PART III***, and we will move directly to the ***Conclusion***.

PART IV

Conclusion and Future Paths

10. Conclusion and future paths

10.1. About this work

In this manuscript a general vision is documented, along with the implementation of several systems and artefacts aligning with that vision, and their evaluation to different extents. The presented works provide an incremental approach to the state of the art. The contribution is perhaps the way knowledge from different fields is put together as a way to take one step closer to the vision that *physical and digital will be (perhaps soon) one and the same.* This vision is tempered by the question *what is the future we are trying to build?*

10.2. Contributions

The contributions presented in this manuscript involve:

- **Tangible Viewports:** the reframing of desktop computers (and the programs living inside of them) as part of the working environment, allowing the seamless transition between the screen and objects placed in front of it.
- **One Reality:** the progressive increase in immersion as a way to gain digital liberties when interacting with the physical space, up to the point of framing any virtual space in relationship with the physical one.
- **Position estimation in mixed reality:** the evaluation of user's capability to transfer information across display modalities (SAR and VR), and perspectives (egocentric and exocentric). The results show very robust capabilities, and seems to point that users can create correct mental models of mixed reality spaces.
- **An Asymmetric collaborative scenario:** the exploration of a collaborative scenario involving a physical meeting room (where experts discuss actions) and a virtual mock-up (where an operator is working). This scenario was inspired from a real use-case, provided and discussed with Airbus engineers.
- **Introspectibles and Inner Garden:** the usage of the previously presented technologies in combination with physiological computing, as a way to support wellbeing practices. Inner Garden was created based on iterations with meditation experts and practitioners.
- **Objects that tell stories:** the exploration of the notion of augmenting everyday objects to improve their capabilities to tell us stories, either by showing recordings from the pasts, or by providing constructed narratives.

10.3. Limitations

This manuscript presented our efforts towards merging physical and digital realms, yet the efforts were, at best, proofs-of-concept exploring the potential of such possibility. The main limitation of the current works is their feasibility outside of the lab context.

It is my personal position that the future of such systems lies at some point between spaces that can support the interaction, and objects that are aware of their properties and are capable of changing them. From an engineering standpoint, this is still far from becoming an everyday reality. For this reason, the next section describes possible steps that can improve the current contributions.

10.4. Considerations regarding Future work

Without repeating what was previously discussed, the improvement of the presented contributions can be achieved in five steps (that can be addressed in order, or independently):

1. **Improving the underlying technologies:** Since the devices presented are lab proofs-of-concept, there are many ways in which the underlying technical solutions could be improved.
2. **Disappearing into the environment:** an important step towards improvement would be to integrate the solutions into artefacts and spaces
3. **Consider deployment:** in order these designs to have an impact in society, we must take deployment as one of the main research steps, otherwise they remind simple mental exercises.
4. **Perform ecological evaluations:** once the systems can be deployed and are robust enough, proper ecological evaluations should be performed. Taking as an example *Inner Garden*, the impact of a system that is designed to foster everyday practices cannot be measured with a single session, yet long term evaluation requires all the previous steps to be achieved.
5. **Move towards embedded systems:** one of the main ways to create more unified experiences is to use, whenever possible, objects that contain the required computational capabilities (processing, sensing and displaying) instead of spatial augmentation. Perhaps computational matter is a concept that we cannot yet materialize, but we can move towards computational objects.

10.4.1. Creating augmented worlds

An aspect not directly addressed by this manuscript is the authoring process itself. As can be seen at the different projects (and Appendix A), the technical challenges are many. It would be then of vital importance to create tools that ease this process for non-experts. Empowering the users by allowing them to become creators is the best way to accelerate and democratize the advance of technology, as can be seen with the Do-it-yourself (DIY) movement.

10.4.2. Augmented fabrication, fabricated augmentation

Augmentation and fabrication are complementary, as the former creates the illusion that digital content is placed in relationship with the physical environment, while the latter allows to materialize digital content. This complementarity is also reflected in the expressivity, materiality and time constraints, particularly when complex interaction is considered.

It would be then of interest to allow the progressive transition between physical and digital prototypes, from early stages of sketching, iteration over fabricated pieces into final products, as briefly described at the Conclusion of **Chapter 4**.

Fabrication that can benefit from augmentation, and vice versa. Most of the tangible props used for the presented projects involve fabricated pieces. Any improvement of the fabricated support could greatly improve the augmentation process, allowing for instance the self-tracking of artefacts [125], or the usage of printed screens combined with projected pixels (using a method similar to [97]).

10.4.3. Deploying Introspectibles

The deployment outside of the lab of the *Introspectibles* is currently being explored by ***Ullo***[7], a Start-up associated with Inria-Potioc (and ***Jérémy Frey*** is one of its members). Their focus is to provide assistive technology for well-being for hospitals and retirement homes.

The notion of deploying artifacts designed for wellbeing is not trivial, as it involves in some cases at risk populations, sensible expectations and sensible data. Beyond the ethical challenges, the form factor of such devices needs to be thoroughly designed. One of the main difficulties of the SAR setups presented in this document is the calibration, as no fully automatic solution was implemented. The creation of fully embedded objects could mitigate this issue.

Technical and application considerations need to be taken into account together. An example of this is the fact that both young kids and senile elderly adults have the tendency to eat elements in front of them. This would be problematic with Inner Garden, even when the sand is nontoxic, it is not designed to be eaten. An alternative to this would be to use food-based materials (e.g., replacing the sand with mash potatoes, or using eatable markers [207]). Like this example, many other unseen challenges will surely arise when the designs will meet the end users.

10.4.4. Asymmetric collaboration

Another relevant step towards deployment is the collaboration with ***Airbus***. If the technology involved is problematic for homes (both caused by its price and complexity), it might still be suited for industrial scenarios where spatial-VR is already used, as communication and prototyping can reduce costs.

As with the deployment of Introspectibles, there are serious considerations to take into account when the final result of using the application has the characteristics of a commercial airplane or rocket engine. Properly adapting the current interaction techniques to support correct formal communication will be essential to allow the usage of such augmented/virtual spaces in the wild.

10.5. A recapitulation

The main idea behind the works presented in this manuscript is that technology can be used as a way to help us improve our lives. To that end, we must carefully design and embrace what we are trying to accomplish, from more adapted interfaces, to more empowering creation tools, and more humane applications.

[7] Ullo website: http://www.ullo-world.fr/

That’s it, Thank you

If you reached this point, you know most of what there is to know of the results of my studies for the past three years. I would like to thank you for reading this manuscript, which is extense and covers many diverse topics, not all of them interesting for everybody.

If after reading all this, you are still wanting to know more on how to make mixed reality systems happen, you can read ***APPENDIX A***. If you would like to have a more personal note about the process (and failures), I invite you to read ***APPENDIX B*** (do not worry, it is shorter than most Chapters).

11. PRERSONAL REFERENCES

Indexed publications

CHI18 J.S. Roo, Jean Basset, Pierre-Antoine Cinquin and M. Hachet. "*Understanding Users' Capability to Transfer Information in* Mixed Reality: Position Estimation across Modalities and Perspectives." Submitted to CHI'18, Montréal, Canada. April 2018. ***(Accepted)***

VRST17 D. Clergeaud*, J.S. Roo*, M. Hachet and P. Guitton. "Towards Seamless Interaction between Physical and Virtual Locations for Asymmetric Collaboration." In VRST'17, Gothenburg, Sweden. November 2017. **First and second author share co-first authorship*

UIST17 J.S. Roo and M. Hachet. *"One Reality: Augmenting How the Physical World is Experienced by Combining Multiple Mixed Reality Modalities."* In UIST '17, Quebec, Canada. October 2017.

CHI17 J.S. Roo, R. Gervais, J. Frey and M. Hachet. *"Inner Garden: Connecting Inner States to a Mixed Reality Sandbox for Mindfulness."* In CHI '17, Denver, USA. May 2017. ***(Top 5%)***

3DUI17 J.S. Roo and M. Hachet. *"Towards a Hybrid Space Combining Spatial Augmented Reality and Virtual Reality."* In 3DUI '17, Los Angeles, USA. March 2017. ***(Best tech-note award)***

TEI16b R. Gervais, J.S. Roo and M. Hachet. *"Tangible Viewports: Getting Out of Flatland in Desktop Environments."* In TEI '16, Eindhoven, Netherlands. February 2016.

TEI16a J.S. Roo, R. Gervais and M. Hachet. *"Inner Garden: an Augmented Sandbox Designed for Self-Reflection."* In TEI '16, Eindhoven, Netherlands. February 2016. *(Work in Progress)*

Other documents (not indexed)

CHI4G16 R. Gervais, J.S. Roo, J. Frey and M. Hachet. "*Introspectibles: Tangible Interaction to Foster Introspection".* In CHI 2016 Computing and Mental Health Workshop (CHI '16 Workshop). May 2016.

ETIS16s J.S. Roo, R. Gervais, J. Frey and M. Hachet. "*Augmented Human Experience: Spatial Augmented Reality and Physiological Computing."* In ETIS – European Tangible Interaction Studio Winter School, Fribourg, Switzerland. ETIS'16. February 2016.

IHM15s J.S. Roo and M. Hachet. "Interacting with Spatial Augmented Reality". At *Rancontes Doctorales* (IHM'15), Toulouse, France. October 2015.

12. REFERENCES

1. Eric Akaoka, Tim Ginn, and Roel Vertegaal. 2010. DisplayObjects: prototyping functional physical interfaces on 3d styrofoam, paper or cardboard models. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction* (TEI '10), 49–56. https://doi.org/10.1145/1709886.1709897

2. Ismo Alakärppä, Elisa Jaakkola, Ashley Colley, and Jonna Häkkilä. 2017. BreathScreen. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4424–4429. https://doi.org/10.1145/3025453.3025973

3. Judith Amores, Xavier Benavides, and Patricia Maes. 2016. PsychicVR: Increasing mindfulness by using Virtual Reality and Brain Computer Interfaces.

4. Michel-Ange Amorim. 2003. "What is my avatar seeing?": The coordination of" out-of-body" and" embodied" perspectives for scene recognition across views. *Visual Cognition* 10, 2: 157–199.

5. Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. 2010. Google street view: Capturing the world at street level. *Computer* 43, 6: 32–38.

6. Robert S Astur, Maria L Ortiz, and Robert J Sutherland. 1998. A characterization of performance by men and women in a virtual Morris water task:: A large and reliable sex difference. *Behavioural brain research* 93, 1: 185–190.

7. S Audet and M Okutomi. 2009. A user-friendly method to geometrically calibrate projector-camera systems. *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*: 47–54. https://doi.org/10.1109/CVPRW.2009.5204319

8. Cédric Bach and Dominique L Scapin. Obstacles and perspectives for evaluating mixed reality systems usability.

9. Ruth A Baer, Gregory T Smith, Jaclyn Hopkins, Jennifer Krietemeyer, and Leslie Toney. 2006. Using self-report assessment methods to explore facets of mindfulness. *Assessment* 13, 1: 27–45.

10. Ravin Balakrishnan and Gordon Kurtenbach. 1999. Exploring Bimanual Camera Control and Object Manipulation in 3D Graphics Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '99), 56–62. https://doi.org/10.1145/302979.302991

11. D Bandyopadhyay, R Raskar, and H Fuchs. Dynamic shader lamps : painting on movable objects. *Proceedings IEEE and ACM International Symposium on Augmented Reality*: 207–216. https://doi.org/10.1109/ISAR.2001.970539

12. Albert Barille. 1987. Once Upon a Time... Life.

13. Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. 1998. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)* 5, 3: 185–223.

14. Hrvoje Benko, Edward W Ishak, and Steven Feiner. 2004. Collaborative mixed reality visualization of an archaeological excavation. In *Proceedings of the 3rd IEEE/ACM international Symposium on Mixed and Augmented Reality*, 132–140.

15. Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 199–208.

16. Hrvoje Benko, Eyal Ofek, Feng Zheng, and Andrew D Wilson. 2015. FoveAR: Combining an Optically See-Through Near-Eye Display with Projector-Based Spatial Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (UIST '15), 129–135. https://doi.org/10.1145/2807442.2807493

17. Hrvoje Benko, Andrew D Wilson, and Federico Zannier. 2014. Dyadic projected spatial augmented reality. In *UIST '14*, 645–655.

18. Warren Berger. 2010. *CAD Monkeys, Dinosaur Babies, and T-Shaped People: Inside the World of Design Thinking and How It Can Spark Creativity and Innovation*. Penguin.

19. Mark Billinghurst, Hirokazu Kato, Kiyoshi Kiyokawa, Daniel Belcher, and Ivan Poupyrev. 2002. Experiments with face-to-face collaborative AR interfaces. *Virtual Reality* 6, 3: 107–121.

20. Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The magicbook-moving seamlessly between reality and virtuality. *IEEE Computer Graphics and applications* 21, 3: 6–8.

21. Oliver Bimber and Ramesh Raskar. 2006. Modern approaches to augmented reality. In *ACM SIGGRAPH 2006*

Courses, 1.

22. G Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

23. Tracy Brandmeyer and Arnaud Delorme. 2013. Meditation and neurofeedback. *Frontiers in psychology* 4.

24. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189: 194.

25. Sophia (Sophia Agnes) Brueckner. 2014. Out of network : technologies to connect with strangers. Retrieved October 18, 2017 from https://dspace.mit.edu/handle/1721.1/95592

26. Jerome Seymour Bruner. 1966. *Toward a theory of instruction*. Harvard University Press.

27. D Alex Butler, Shahram Izadi, Otmar Hilliges, David Molyneaux, Steve Hodges, and David Kim. 2012. Shake'n'sense: reducing interference for overlapping structured light depth cameras. In *CHI '12*, 1933–1936.

28. Andreas Butz, Tobias Hollerer, Steven Feiner, Blair MacIntyre, and Clifford Beshers. 1999. Enveloping users and computers in a collaborative 3D augmented reality. In *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, 35–44.

29. P Caffarra, G Vezzadini, F Dieci, F Zonato, and A Venneri. 2002. Rey-Osterrieth complex figure: normative values in an Italian population sample. *Neurological Sciences* 22, 6: 443–447.

30. Rafael A Calvo and Dorian Peters. 2014. *Positive computing: technology for wellbeing and human potential*. MIT Press.

31. David A Carter and James Diaz. 1999. *The elements of pop-up: A pop-up book for aspiring paper engineers*. Little Simon.

32. Jessica R Cauchard, Markus Löchtefeld, Pourang Irani, Johannes Schoening, Antonio Krüger, Mike Fraser, and Sriram Subramanian. 2011. Visual Separation in Mobile Multi-display Environments. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (UIST '11), 451–460. https://doi.org/10.1145/2047196.2047256

33. Damien Clergeaud and Pascal Guitton. 2017. Design of an annotation system for taking notes in virtual reality. In *3DTV-CON 2017: 3DTV Conference 2017: Research and Applications in Future 3D Media*.

34. Joanna Cole and Bruce Degen. 1994. The Magic School Bus.

35. Marcello Costantini, Ettore Ambrosini, Claudia Scorolli, and Anna M Borghi. 2011. When objects are close to me: affordances in the peripersonal space. *Psychonomic bulletin & review* 18, 2: 302–308.

36. Chris Crawford. 2003. *Chris Crawford on game design*. New Riders.

37. Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. 1993. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 135–142.

38. Cubicle Ninjas. 2016. Virtual Reality Meditation.

39. Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. 2007. MonoSLAM: Real-time single camera SLAM. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29, 6: 1052–1067.

40. Nicolas J Dedual, Ohan Oda, and Steven K Feiner. 2011. Creating hybrid user interfaces with a 2D multi-touch tabletop and a 3D see-through head-worn display. In *ISMAR 2011*, 231–232.

41. Deep VR. 2016. Deep VR.

42. Pieter Desmet and Marc Hassenzahl. 2012. Towards happiness: Possibility-driven design. In *Human-computer interaction: The agency perspective*. Springer, 3–27.

43. Pieter M A Desmet and Anna E Pohlmeyer. 2013. Positive design: An introduction to design for subjective well-being. *International Journal of Design, 7 (3), 2013*.

44. Sebastian Deterding. 2012. Gamification: Designing for Motivation. *interactions* 19, 4: 14–17. https://doi.org/10.1145/2212877.2212883

45. Philip K. Dick. 1968. *Do androids dream of Electrc sheep*. Doubleday.

46. Tanja Döring, Axel Sylvester, and Albrecht Schmidt. 2013. A design space for ephemeral user interfaces. *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction - TEI '13*: 75. https://doi.org/10.1145/2460625.2460637

47. Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern Statistical Methods for HCI*. Springer, 291–330.

48. Andreas Dünser, Raphaël Grasset, and Mark Billinghurst. 2008. *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand.

49. Andreas Dünser, Karin Steinbügl, Hannes Kaufmann, and Judith Glück. 2006. Virtual and Augmented Reality As Spatial Ability Training Tools. In *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI* (CHINZ '06), 125–132. https://doi.org/10.1145/1152760.1152776

50. C Eames and R Eames. 1977. Powers of ten [Motion picture]. *United States: IBM*.

51. David Kanga Ed Key. 2013. Proteus.

52. Mike Eissele, Oliver Siemoneit, and Thomas Ertl. 2006. Transition of mixed, virtual, and augmented reality in smart production environments-an interdisciplinary view. In *Robotics, Automation and Mechatronics, 2006 IEEE Conference on*, 1–6.

53. Monika Elepfandt and Marcelina Sünderhauf. 2011. Multimodal, touchless interaction in spatial augmented reality environments. In *International Conference on Digital Human Modeling*, 263–271.

54. Norman Farb, Jennifer J Daubenmier, Cynthia J Price, Tim Gard, Catherine Kerr, Barney Dunn, Anne Carolyn KLein, Martin P Paulus, and Wolf E Mehling. 2015. Interoception, Contemplative Practice, and Health. *Name: Frontiers in Psychology* 6: 763.

55. Kenneth P Fishkin, Thomas P Moran, and Beverly L Harrison. 1999. Embodied user interfaces: Towards invisible user interfaces. In *Engineering for Human-Computer Interaction*. Springer, 1–18.

56. KennethP. Fishkin. 2004. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing* 8, 5: 347–358. https://doi.org/10.1007/s00779-004-0297-4

57. George W Fitzmaurice, Hiroshi Ishii, and William A S Buxton. 1995. Bricks: Laying the Foundations for Graspable User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '95), 442–449. https://doi.org/10.1145/223904.223964

58. Sean Follmer, Daniel Leithinger, Alex Olwal, Nadia Cheng, and Hiroshi Ishii. 2012. Jamming user interfaces: programmable particle stiffness and sensing for malleable and shape-changing devices. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, 519–528.

59. Paul J. Ford. 2001. A further analysis of the ethics of representation in virtual reality: Multi-user environments. *Ethics and Information Technology* 3, 2: 113–121. https://doi.org/10.1023/A:1011846009390

60. Clifton Forlines, Daniel Vogel, and Ravin Balakrishnan. 2006. HybridPointing: fluid switching between absolute and relative pointing with a direct input device. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, 211–220.

61. Clifton Forlines, Daniel Vogel, Ravin Balakrishnan, Renaud Gervais, Jérémy Frey, Martin Hachet, Ken Hinckley, Randy Pausch, John C Goble, Neal F Kassell, Miguel A Nacenta, Samer Sallam, Bernard Champoux, Sriram Subramanian, Carl Gutwin, Brett Ridel, Patrick Reuter, Jérémy Laviole, Nicolas Mellado, Nadine Couture, Xavier Granier, Bruce H Thomas, Michael Marner, Ross T Smith, Neven Abdelaziz Mohamed Elsayed, Stewart Von Itzstein, Karsten Klein, Matt Adcock, Peter Eades, Andrew Irlitti, Joanne Zucco, and others. 2014. Spatial augmented reality：A tool for 3D data visualization. In *INTERACT* (CHI '06), 45–50. https://doi.org/10.1145/1124772.1124817

62. Jérémie Francone and Laurence Nigay. 2011. Using the user's point of view for interaction on mobile devices. In *Proceedings of the 23rd Conference on l'Interaction Homme-Machine*, 4.

63. Jérémy Frey, Renaud Gervais, Stéphanie Fleck, Fabien Lotte, Martin Hachet, and others. 2014. Teegi: tangible EEG interface. In *UIST 2014, ACM User Interface Software and Technology Symposium*.

64. Ben Fullerton. 2010. Designing for solitude. *interactions* 17, 6: 6–9.

65. Chris Furmanski, Ronald Azuma, and Mike Daily. 2002. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, 215–320.

66. Marylène Gagné and Edward L Deci. 2005. Self-determination theory and work motivation. *Journal of Organizational behavior* 26, 4: 331–362.

67. Renaud Gervais, Jérémy Frey, Alexis Gay, Fabien Lotte, and Martin Hachet. 2016. TOBE: Tangible Out-of-Body Experience. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. https://doi.org/10.1145/2839462.2839486

68. Renaud Gervais, Jérémy Frey, and Martin Hachet. 2015. Pointing in Spatial Augmented Reality from 2D Pointing Devices. In *INTERACT*, 8. Retrieved from https://hal.archives-ouvertes.fr/hal-01153647

69. William Gibson. 1995. Neuromancer. 1984. *New York: Ace*.

70. Patricio Gonzalez. 2011. Efecto Mariposa.

71. James G Greeno. 1994. Gibson's Affordances. *Psychological Review* 101, 2: 336–342. Retrieved October 18, 2017 from http://ftp.idiap.ch/pub/courses/EE-700/material/31-10-2012/gibsonAffordances.pdf

72. Diane Gromala, Xin Tong, Amber Choo, Mehdi Karamnejad, and Chris D Shaw. 2015. The virtual meditative walk: virtual reality therapy for chronic pain management. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 521–524.

73. Paul Grossman, Ludger Niemann, Stefan Schmidt, and Harald Walach. 2004. Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of psychosomatic research* 57, 1: 35–43.

74. Kristoffer Hagen, Konstantinos Chorianopoulos, Alf Inge Wang, Letizia Jaccheri, and Stian Weie. 2016. Gameplay As Exercise. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), 1872–1878. https://doi.org/10.1145/2851581.2892515

75. Chris Harrison, Hrvoje Benko, and Andrew D Wilson. 2011. OmniTouch: Wearable Multitouch Interaction Everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (UIST '11), 441–450. https://doi.org/10.1145/2047196.2047255

76. Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52: 139–183.

77. John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1: 81–112.

78. Headspace inc. 2016. Headspace - Treat your head right.

79. Mary Hegarty and David Waller. 2004. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 2: 175–191.

80. John F Helliwell, Richard Layard, and Jeffrey Sachs. 2012. World happiness report. *Earth Institute*.

81. Valentin Heun, James Hobin, and Pattie Maes. 2013. Reality Editor: Programming Smarter Objects. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (UbiComp '13 Adjunct), 307–310. https://doi.org/10.1145/2494091.2494185

82. Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. 2012. HoloDesk: direct 3d interactions with a situated see-through display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2421–2430.

83. James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 2: 174–196.

84. David Holman, Jesse Burstyn, Ryan Brotman, Audrey Younkin, and Roel Vertegaal. 2013. Flexkit: a rapid prototyping platform for flexible displays. In *Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology*, 17–18.

85. David Holman, Audrey Girouard, Hrvoje Benko, and Roel Vertegaal. 2013. The Design of Organic User Interfaces: Shape, Sketching and Hypercontext. *Interacting with Computers* 25, 2: 133–142.

86. Kasper Hornbaek. 2015. Flexible Displays, Rigid Designs? *Computer*, 3: 92–96.

87. Eva Hornecker and Jacob Buur. Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. https://doi.org/10.1145/?????????

88. Kai Huotari and Juho Hamari. 2012. Defining Gamification: A Service Marketing Perspective. In *Proceeding of the 16th International Academic MindTrek Conference* (MindTrek '12), 17–22. https://doi.org/10.1145/2393132.2393137

89. Hikaru Ibayashi, Yuta Sugiura, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Masaaki Mochimaru, and Takeo Igarashi. 2015. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 Emerging Technologies*, 8.

90. Ice Water Games. 2015. Viridi.

91. Victoria Interrante, Brian Ries, and Lee Anderson. 2006. Distance perception in immersive virtual environments, revisited. In *Virtual Reality Conference, 2006*, 3–10.

92. Hiroshi Ishii, Dávid Lakatos, Leonardo Bonanni, and Jean-Baptiste Labrune. 2012. Radical Atoms: Beyond Tangible Bits, Toward Transformable Materials. *interactions* 19, 1: 38–51. https://doi.org/10.1145/2065327.2065337

93. Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms.

In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI '97), 234–241. https://doi.org/10.1145/258549.258715

94. Robert J K Jacob, Audrey Girouard, Leanne M Hirshfield, Michael S Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. Reality-Based Interaction: A Framework for Post-WIMP Interfaces. Retrieved October 18, 2017 from http://www.cs.tufts.edu/~jacob/papers/chi08.pdf

95. Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2013. Evaluating the efficiency of physical visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2593–2602.

96. Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. 2007. An interactive 360 light field display. In *ACM SIGGRAPH 2007 emerging technologies*, 13.

97. Brett R Jones, Rajinder Sodhi, Pulkit Budhiraja, Kevin Karsch, Brian Bailey, and David Forsyth. 2015. Projectibles: Optimizing Surface Color For Projection. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 137–146.

98. Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units. In *UIST '14*, 637–644.

99. J Adam Jones, J Edward Swan II, Gurjot Singh, Eric Kolstad, and Stephen R Ellis. 2008. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, 9–14.

100. Jon Kabat-Zinn. 2003. Mindfulness-based interventions in context: past, present, and future. *Clinical psychology: Science and practice* 10, 2: 144–156.

101. Dora M Kalff. 2003. *Sandplay: A psychotherapeutic approach to the psyche*. Temenos Press.

102. Shaun K Kane, Daniel Avrahami, Jacob O Wobbrock, Beverly Harrison, Adam D Rea, Matthai Philipose, and Anthony LaMarca. 2009. Bonfire: A Nomadic System for Hybrid Laptop-tabletop Interaction. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (UIST '09), 129–138. https://doi.org/10.1145/1622176.1622202

103. Dominic Kao and D Fox Harrell. 2016. Exploring the Effects of Encouragement in Educational Games. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), 1906–1914. https://doi.org/10.1145/2851581.2892335

104. Michael Karlesky and Katherine Isbister. 2014. Designing for the physical margins of digital workspaces: fidget widgets in support of productivity and creativity. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, 13–20.

105. Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2017. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE transactions on visualization and computer graphics* 23, 3: 1222–1234.

106. Shunichi Kasahara, Ryuma Niiyama, Valentin Heun, and Hiroshi Ishii. 2013. exTouch: Spatially-aware Embodied Manipulation of Actuated Objects Mediated by Augmented Reality. In *TEI '13* (TEI '13), 223–228.

107. Shunichi Kasahara and Jun Rekimoto. 2014. JackIn. In *Proceedings of the 5th Augmented Human International Conference on - AH '14*, 1–8. https://doi.org/10.1145/2582051.2582097

108. Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head. In *SIGGRAPH Asia 2015 Emerging Technologies on - SA '15*, 1–3. https://doi.org/10.1145/2818466.2818486

109. Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology - VRST '15*, 217–225. https://doi.org/10.1145/2821592.2821608

110. Rohit Ashok Khot, Larissa Hjorth, and Florian'Floyd' Mueller. 2014. Understanding physical activity through 3D printed material artifacts. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 3835–3844.

111. Matthew A Killingsworth and Daniel T Gilbert. 2010. A wandering mind is an unhappy mind. *Science* 330, 6006: 932.

112. Kiyoshi Kiyokawa, Mark Billinghurst, Sohan E Hayes, Anoop Gupta, Yuki Sannohe, and Hirokazu Kato. 2002. Communication behaviors of co-located users in collaborative AR interfaces. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, 139–148.

113. Kiyoshi Kiyokawa, Hidehiko Iwasa, Haruo Takemura, and Naokazu Yokoya. Collaborative immersive workspace through a shared augmented environment.

114. Kiyoshi Kiyokawa and Haruo Takemura. 2005. A tunnel window and its variations: Seamless teleportation techniques in a virtual environment. In *HCI International*.

115. Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. JackIn space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*, 14.

116. Ilkka Kosunen, Mikko Salminen, Simo Järvelä, Antti Ruonala, Niklas Ravaja, and Giulio Jacucci. 2016. RelaWorld: Neuroadaptive and Immersive Virtual Reality Meditation System. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 208–217.

117. D W F Van Krevelen and R Poelman. 2010. A Survey of Augmented Reality Technologies , Applications and Limitations. 9, 2.

118. Ernst Kruijff, J Edward Swan II, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *ISMAR*, 3–12.

119. André Kunert, Alexander Kulik, Stephan Beck, and Bernd Froehlich. 2014. Photoportals: Shared References in Space and Time. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 1388–1399. https://doi.org/10.1145/2531602.2531727

120. Kai Kunze, Mikio Iwamura, Kenji Kise, Seiichi Uchida, and Shinichiro Omachi. 2013. Activity Recognition for the Mind: Toward a Cognitive" Quantified Self." *Computer* 46, 10: 105–108.

121. Mark A Lau, Scott R Bishop, Zindel V Segal, Tom Buis, Nicole D Anderson, Linda Carlson, Shauna Shapiro, James Carmody, Susan Abbey, and Gerald Devins. 2006. The Toronto mindfulness scale: Development and validation. *Journal of clinical psychology* 62, 12: 1445–1467.

122. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. *Construction and evaluation of a user experience questionnaire*. Springer.

123. Jeremy Laviole and Martin Hachet. 2012. PapARt: Interactive 3D graphics and multi-touch augmented paper for artistic creation. *2012 IEEE Symposium on 3D User Interfaces (3DUI)*: 3–6. https://doi.org/10.1109/3DUI.2012.6184167

124. Jinha Lee and Cati Boulanger. 2012. Direct, spatial, and dexterous interaction with see-through 3D desktop. In *ACM SIGGRAPH 2012 Posters*, 69.

125. Johnny C Lee, Scott E Hudson, Jay W Summet, and Paul H Dietz. 2005. Moveable interactive projected displays using projector based tracking. In *UIST '05*, 63–72.

126. Kai S Lehmann, Joerg P Ritz, Heiko Maass, Hueseyin K Çakmak, Uwe G Kuehnapfel, Christoph T Germer, Georg Bretthauer, and Heinz J Buhr. 2005. A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Annals of surgery* 241, 3: 442.

127. Jason Leigh, Andrew E Johnson, Christina A Vasilakis, and Thomas A DeFanti. 1996. Multi-perspective collaborative design in persistent networked virtual environments. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*, 253–260.

128. Marcia C Linn and Anne C Petersen. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development*: 1479–1498.

129. Lionhead Studios. 2001. Black and White.

130. Mark A Livingston, Catherine Zanbaka, J Edward Swan, and Harvey S Smallman. 2005. Objective measures for the effectiveness of augmented reality. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, 287–288.

131. E T A Maas, M R Marner, R T Smith, and B H Thomas. 2012. Supporting Freeform Modelling in Spatial Augmented Reality Environments with a New Deformable Material. In *Australasian User Interface Conference (AUIC 2012)* (CRPIT), 77–86. Retrieved from http://crpit.com/confpapers/CRPITV126Maas.pdf

132. Andrzej Marczewski. 2013. The Intrinsic Motivation RAMP.

133. MR Marner. Physical-virtual tools for interactive spatial augmented reality. Retrieved November 14, 2014 from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Physical-Virtual+Tools+for+Interactive+Spatial+Augmented+Reality#0

134. Rollin McCraty, Mike Atkinson, Dana Tomasino, and Raymond Trevor Bradley. 2009. The coherent heart: Heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order. *Integral Review* 5, 2: 10–115.

135. Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12: 1321–1329.

136. Mindfulnets. 2011. Mindfulnets - Connecting online from within.

137. Mark R Mine, Jeroen van Baar, Anselm Grundhofer, David Rose, and Bei Yang. 2012. Projection-based augmented reality in disney theme parks. *Computer* 45, 7: 32–40.

138. Mojang. 2011. Minecraft.

139. Neema Moraveji, Athman Adiseshan, and Takehiro Hagiwara. 2012. Breathtray: augmenting respiration self-regulation without cognitive deficit. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, 2405–2410.

140. Jens Mueller, Roman Rädle, and Harald Reiterer. 2017. Remote Collaboration With Mixed Reality Displays: How Shared Virtual Landmarks Facilitate Spatial Referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6481–6486.

141. Stefanie Mueller, Martin Fritzsche, Jan Kossmann, Maximilian Schneider, Jonathan Striebel, and Patrick Baudisch. Scotty : Relocating Physical Objects Across Distances Using Destructive Scanning , Encryption , and 3D Printing. https://doi.org/10.1145/2677199.2680547

142. Miguel A Nacenta, Satoshi Sakurai, Tokuo Yamaguchi, Yohei Miki, Yuichi Itoh, Yoshifumi Kitamura, Sriram Subramanian, and Carl Gutwin. 2007. E-conic: A Perspective-aware Interface for Multi-display Environments. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (UIST '07), 279–288. https://doi.org/10.1145/1294211.1294260

143. Miguel A Nacenta, Samer Sallam, Bernard Champoux, Sriram Subramanian, and Carl Gutwin. 2006. Perspective Cursor: Perspective-based Interaction for Multi-display Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), 289–298. https://doi.org/10.1145/1124772.1124817

144. Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, 127–136.

145. Yoichi Ochiai, Kota Kumagai, Takayuki Hoshi, Jun Rekimoto, Satoshi Hasegawa, and Yoshio Hayasaki. 2016. Fairy lights in femtoseconds: aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields. *ACM Transactions on Graphics (TOG)* 35, 2: 17.

146. Simon Olberding, Sergio Soto Ortega, Klaus Hildebrandt, and Jürgen Steimle. 2015. Foldio: Digital fabrication of interactive and shape-changing objects with foldable printed electronics. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 223–232.

147. Simon Olberding, Michael Wessely, and Jürgen Steimle. 2014. PrintScreen: fabricating highly customizable thin-film touch-displays. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 281–290.

148. OpenBCI. 2016. OpenBCI - Open Source Biosensing Tools (EEG, EMG, EKG, and more).

149. David OReilly. 2014. Mountain.

150. Mamoru Oshii. 1995. *Ghost in the Shell*. Shochiku, Japan. Retrieved from http://www.imdb.com/title/tt0113568/

151. Robert Patterson, Marc D Winterbottom, and Byron J Pierce. 2006. Perceptual issues in the use of head-mounted visual displays. *Human factors* 48, 3: 555–573. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17063969

152. Eric Paulos, Chris Myers, Rundong Tian, and Paxton Paulos. 2014. Sensory triptych: here, near, out there. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 491–496.

153. Rosalind W Picard and Jennifer Healey. 1997. Affective wearables. In *Wearable Computers, 1997. Digest of Papers., First International Symposium on*, 90–97.

154. Ben Piper, Carlo Ratti, and Hiroshi Ishii. Illuminating Clay: A 3-D Tangible Interface for Landscape Analysis. Retrieved October 18, 2017 from http://www.cs.uml.edu/~fredm/courses/91.548-spr03/papers/illclay_chi02.pdf

155. Andrew K Przybylski, C Scott Rigby, and Richard M Ryan. 2010. A motivational model of video game engagement. *Review of general psychology* 14, 2: 154.

156. Jie Qi and Leah Buechley. 2010. Electronic popables: exploring paper-based computing through an interactive pop-up book. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, 121–128.

157. Ramesh Raskar and Kok-Lim Low. 2001. Interacting with spatially augmented reality. *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation - AFRIGRAPH '01*: 101. https://doi.org/10.1145/513886.513889

158. Ramesh Raskar, 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, 179–188.

159. Ramesh Raskar, Greg Welch, and Henry Fuchs. 1998. Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR '98)*, 11–20.

160. Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. 2001. Shader Lamps: Animating Real Objects With Image-Based Illumination. In *Rendering Techniques 2001*, StevenJ. Gortler and Karol Myszkowski (eds.). Springer Vienna, 89–102. https://doi.org/10.1007/978-3-7091-6242-2_9

161. Jun Rekimoto, Yuji Ayatsuka, and Kazuteru Hayashi. 1998. Augment-able reality: Situated communication through physical and digital spaces. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, 68–75.

162. Jun Rekimoto and Katashi Nagao. 1995. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of the 8th annual ACM symposium on User interface and software technology*, 29–36.

163. Jun Rekimoto and Masanori Saitoh. 1999. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 378–385.

164. Yann Renard, Fabien Lotte, Guillaume Gibert, Marco Congedo, Emmanuel Maby, Vincent Delannoy, Olivier Bertrand, and Anatole Lécuyer. 2010. Openvibe: An open-source software platform to design, test, and use brain--computer interfaces in real and virtual environments. *Presence* 19, 1: 35–53.

165. Brett Ridel, Patrick Reuter, Jérémy Laviole, Nicolas Mellado, Nadine Couture, and Xavier Granier. 2014. The Revealing Flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. *Journal on Computing and Cultural Heritage (JOCCH)* 7, 2: 6.

166. Giuseppe Riva, Rosa M Banos, Cristina Botella, Brenda K Wiederhold, and Andrea Gaggioli. 2012. Positive technology: using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking* 15, 2: 69–77.

167. Yvonne Rogers. 2006. Moving on from weiser's vision of calm computing: Engaging ubicomp experiences. In *UbiComp 2006: Ubiquitous Computing*. Springer, 404–421.

168. David Rose. *Enchanted objects : design, human desire, and the Internet of things*. Retrieved October 18, 2017 from https://books.google.fr/books?hl=en&lr=&id=PkH6AwAAQBAJ&oi=fnd&pg=PR11&dq=enchanted+objects+rose&ots=0ItoLoDvZ6&sig=mwLWK7MeBmDlexUAxCZq6NeW6DE#v=onepage&q=enchanted objects rose&f=false

169. Royal Institute of Technology (KTH). 2015. Mad Sand.

170. Carol D Ryff and Burton H Singer. 2008. Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of happiness studies* 9, 1: 13–39.

171. Daniel Saakes and Pieter Jan Stappers. 2009. A tangible design tool for sketching materials in products. *AI EDAM* 23, 3: 275–287.

172. Carl Sagan, Steven Soter, and Ann Druyan. 1989. Cosmos: A personal voyage.

173. Kavous Salehzadeh Niksirat, Chaklam Silpasuwanchai, Mahmoud Mohamed Hussien Ahmed, Peng Cheng, and Xiangshi Ren. 2017. A Framework for Interactive Mindfulness Meditation Using Attention-Regulation Process. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 2672–2684. https://doi.org/10.1145/3025453.3025914

174. Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: Enhancing Touch Interaction on Humans, Screens, Liquids, and Everyday Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 483–492. https://doi.org/10.1145/2207676.2207743

175. Yihui Saw. 2014. Enlight: a projected augmented reality approach to science education. Massachusetts Institute of Technology.

176. Ridley Scott. 1982. *Blade Runner*.

177. Sue Ann Seah, Diego Martinez Plasencia, Peter Bennett, Abhijit Karnik, Vlad Otrocol, Jarrod Knibbe, Andy Cockburn, and Sriram Subramanian. SensaBubble: A Chrono-Sensory Mid-Air Display of Sight and Smell. https://doi.org/10.1145/2556288.2557087

178. Sebastian Deterding. 2015. Designing Against Productivity. *MIT Media Lab*. Retrieved October 18, 2017 from https://www.media.mit.edu/videos/wellbeing-2015-11-24-2/

179. Zindel V Segal, J Mark G Williams, and John D Teasdale. 2012. *Mindfulness-based cognitive therapy for depression*. Guilford Press.

180. Martin E P Seligman and Mihaly Csikszentmihalyi. 2000. *Positive psychology: An introduction.* American Psychological Association.

181. Abigail Sellen, Yvonne Rogers, Richard Harper, and Tom Rodden. 2009. Reflecting human values in the digital age. *Communications of the ACM* 52, 3: 58. https://doi.org/10.1145/1467247.1467265

182. Orit Shaer. 2009. Tangible User Interfaces: Past, Present, and Future Directions. *Foundations and Trends® in Human–Computer Interaction* 3, 1–2: 1–137. https://doi.org/10.1561/1100000026

183. Ari Shapiro, Andrew Feng, Ruizhe Wang, Hao Li, Mark Bolas, Gerard Medioni, and Evan Suma. 2014. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds* 25.

184. Toby Sharp, Cem Keskin, Duncan Robertson, Jonathan Taylor, Jamie Shotton, David Kim, Christoph Rhemann, Ido Leichter, Alon Vinnikov, Yichen Wei, and others. 2015. Accurate, robust, and flexible real-time hand tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3633–3642.

185. Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. *Communications of the ACM* 56, 1: 116–124.

186. Mel Slater, Martin Usoh, and Anthony Steed. Depth of Presence in Virtual Environments. Retrieved October 18, 2017 from https://www.researchgate.net/profile/Anthony_Steed2/publication/237129885_Depth_of_Presence_in_Immersive_Virtual_Environments/links/5829e4e308aef19cb80506ea.pdf

187. Smiling Mind. 2010. Smiling Mind - Meditation made easy.

188. Ross T Smith, Guy Webber, Maki Sugimoto, Michael Marner, and Bruce H Thomas. 2013. [Invited Paper] Automatic Sub-pixel Projector Calibration. *ITE Transactions on Media Technology and Applications* 1, 3: 204–213.

189. Tobias Sonne and Mads Møller Jensen. 2016. ChillFish: A Respiration Game for Children with ADHD. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 271–278.

190. Stephen Spielberg. 2003. *Minority Report*.

191. Charles D Spielberger. 2010. *State-Trait anxiety inventory*. Wiley Online Library.

192. Jürgen Steimle. 2015. Printed electronics for human-computer interaction. *interactions* 22, 3: 72–75.

193. Frank Steinicke, Gerd Bruder, Klaus Hinrichs, Markus Lappe, Brian Ries, and Victoria Interrante. 2009. Transitional environments enhance distance perception in immersive virtual reality systems. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, 19–26.

194. Neal Stephenson. 1998. *The diamond age*. Penguin UK.

195. Brett Stevens, Jennifer Jerrams-Smith, David Heathcote, and David Callear. 2002. Putting the Virtual into Reality: Assessing Object-Presence with Projection-Augmented Models. *Presence* 11, 1: 79–92. Retrieved October 18, 2017 from www.tech.port.ac.uk/

196. Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 265–272.

197. Stanislav L Stoev and Dieter Schmalstieg. 2002. Application and taxonomy of through-the-lens techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology*, 57–64.

198. Ivan E Sutherland. 1965. The ultimate display. *Multimedia: From Wagner to virtual reality*.

199. John Sutton, Celia B Harris, Paul G Keil, and Amanda J Barnier. 2010. The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the cognitive sciences* 9, 4: 521–560.

200. Desney S Tan and Mary Czerwinski. 2003. Effects of visual separation and physical discontinuities when distributing information across multiple displays. In *INTERACT*, 252–255.

201. Robert J Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, 1: 87–94. https://doi.org/10.1109/3DUI.2011.5759222

202. Alexander Teibrich, Stefanie Mueller, François Guimbretière, Robert Kovacs, Stefan Neubert, and Patrick Baudisch. 2015. Patching Physical Objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15*, 83–91. https://doi.org/10.1145/2807442.2807467

203. The Mill. 2016. STRATA: A New VR Experiment.

204. Yutaka Tokuda, Mohd Adili Norasikin, Sriram Subramanian, and Diego Martinez Plasencia. 2017. MistForm. In

Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17, 4383–4395. https://doi.org/10.1145/3025453.3025608

205. Giovanni Maria Troiano, Esben Warming Pedersen, and Kasper Hornbæk. 2014. User-defined gestures for elastic, deformable displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, 1–8.

206. Michael Tsang, George W Fitzmaurice, Gordon Kurtenbach, Azam Khan, and Bill Buxton. 2002. Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, 111–120.

207. Takahiro Uji, Yiting Zhang, and Hiromasa Oku. 2017. Edible retroreflector. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17*, 1–8. https://doi.org/10.1145/3139131.3139148

208. Brygg Ullmer and Hiroshi Ishii. 1997. The metaDESK: models and prototypes for tangible user interfaces. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, 223–232.

209. Brygg Ullmer and Hiroshi Ishii. 2000. Emerging frameworks for tangible user interfaces. *IBM systems journal* 39, 3.4: 915–931.

210. John Underkoffler and Hiroshi Ishii. 1999. Urp: a luminous-tangible workbench for urban planning and design. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 386–393.

211. Unity Technologies. 2015. Unity - Game Engine.

212. Daisuke Uriu and Naohito Okude. 2010. ThanatoFenestra: photographic family altar supporting a ritual to pray for the deceased. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, 422–425.

213. S G Vandenberg. 1971. Mental rotation test. *Boulder: University of Colorado*.

214. Bret Victor. Seeing Spaces. Retrieved from https://vimeo.com/97903574

215. Jay Vidyarthi, Bernhard E Riecke, and Diane Gromala. 2012. Sonic Cradle: designing for an immersive experience of meditation by connecting respiration to music. In *Proceedings of the designing interactive systems conference*, 408–417.

216. Daniel Wagner and Dieter Schmalstieg. 2007. Artoolkitplus for pose tracking on mobile devices. In *Proceedings of 12th Computer Vision Winter Workshop (CVWW'07)*, 139–146.

217. James A Walsh, Stewart Von Itzstein, and Bruce H Thomas. Ephemeral Interaction Using Everyday Objects. Retrieved October 18, 2017 from http://crpit.scem.westernsydney.edu.au/confpapers/CRPITV150Walsh.pdf

218. Ranxiao Frances Wang and Daniel J Simons. 1999. Active and passive scene recognition across views. *Cognition* 70, 2: 191–210.

219. Colin Ware, Kevin Arthur, and Kellogg S Booth. 1993. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, 37–42.

220. Christian Weichel, Manfred Lau, David Kim, Nicolas Villar, Hans W. Gellersen, Christian Weichel, Manfred Lau, David Kim, Nicolas Villar, and Hans W. Gellersen. 2014. MixFab. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 3855–3864. https://doi.org/10.1145/2556288.2557090

221. Mark Weiser. 1991. The Computer for the 21 st Century. *Scientific american* 265, 3: 94–105.

222. Mark Weiser. 1993. Some computer science issues in ubiquitous computing. *Communications of the ACM* 36, 7.

223. Mark Weiser and John Seely Brown. 1997. The coming age of calm technology. In *Beyond calculation*. Springer.

224. Pierre Wellner. 1993. Interacting with Paper on the DigitalDesk. *Commun. ACM* 36, 7: 87–96.

225. J Mark G Williams and Jon Kabat-Zinn. 2011. Mindfulness: diverse perspectives on its meaning, origins, and multiple applications at the intersection of science and dharma. *Contemporary Buddhism* 12, 1: 1–18.

226. Michele A Williams, Asta Roseway, Chris O'Dowd, Mary Czerwinski, and Meredith Ringel Morris. 2015. SWARM: An Actuated Wearable for Mediating Affect. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, 293–300.

227. Karl D.D. Willis and Ivan Poupyrev. 2010. MotionBeam. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, 3253. https://doi.org/10.1145/1753846.1753967

228. Karl D.D. Willis, Ivan Poupyrev, Scott E. Hudson, and Moshe Mahler. 2011. SideBySide. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, 431. https://doi.org/10.1145/2047196.2047254

229. Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9, 4: 625–636.

230. Craig Wisneski, Hiroshi Ishii, Andrew Dahley, Matt Gorbet, Scott Brave, Brygg Ullmer, and Paul Yarin. 1998. Ambient displays: Turning architectural space into an interface between people and digital information. In *International Workshop on Cooperative Buildings*, 22–32.

231. Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 143–146.

232. Gary Wolf. 2010. The data-driven life. *The New York Times* 28: MM38. Retrieved from http://nyti.ms/18XZ1Zv

233. Robert Xiao, Chris Harrison, and Scott E Hudson. 2013. WorldKit: rapid and easy creation of ad-hoc interactive applications on everyday surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 879–888.

234. Robert Xiao, Miguel A Nacenta, Regan L Mandryk, Andy Cockburn, and Carl Gutwin. 2011. Ubiquitous cursor: a comparison of direct and indirect pointing feedback in multi-display environments. In *Proceedings of Graphics Interface 2011*, 135–142.

235. Hyun Joong Yoon and Seong Youb Chung. 2011. EEG spectral analysis in valence and arousal dimensions of emotion. In *Control, Automation and Systems (ICCAS), 2011 11th International Conference on*, 1319–1322.

236. Jianlong Zhou, Ivan Lee, Bruce Thomas, Roland Menassa, Anthony Farrant, and Andrew Sansome. 2011. Applying Spatial Augmented Reality to Facilitate In-situ Support for Automotive Spot Welding Inspection. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry* (VRCAI ’11), 195–200.

237. Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, and others. 2014. Real-time non-rigid reconstruction using an RGB-D camera. *ACM Transactions on Graphics (TOG)* 33, 4: 156.

238. vvvv. Retrieved from http://vvvv.org

APENDICES

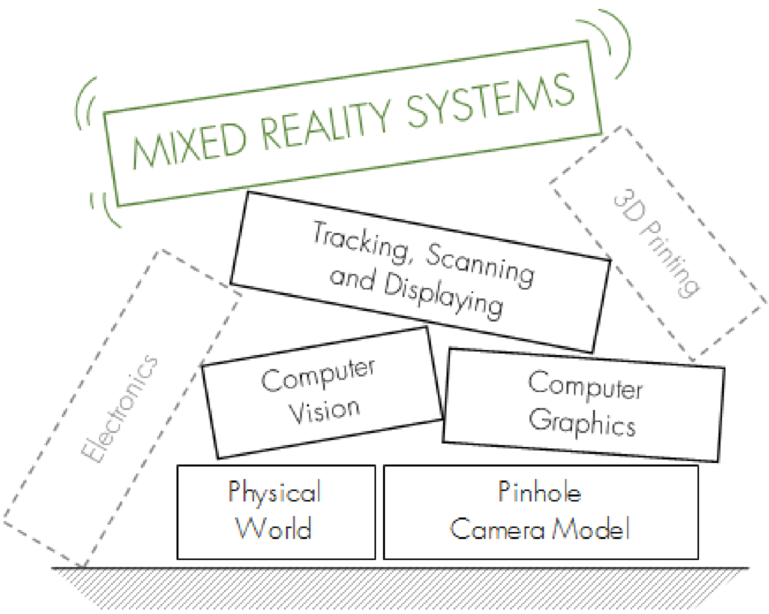
TECHNICAL AND PERSONAL

Appendices: Technical and Personal

A. Implementation

Cameras and Projectors

"Sometimes science is a lot more art, than science…
A lot of people don't get that."
— Rick, Rick and Morty



About this section:

This section introduces the technical knowledge required to be able to create mixed reality applications. This section focus solely on the vision-based augmentation, both from the computer perspective (based in computer vision and computer graphics), and the human perspective (the creation of visual illusions).

The unstable look on the image above is not accidental, as every new layer adds its own complications.

A.1. Cameras, their calibration and usage

At the core of mixed reality is the ability to obtain and understand visual information in relationship with the space. To this end, we need to be able to correctly model light paths, and the relationship between optical devices.

A.1.1. The pinhole camera model

Optical devices, such as cameras and projectors, can be described using an extended pin camera model. This model combines spatial properties (*extrinsics*, position and orientation in space) and lens properties (*intrinsics* and *distortion*). As a result, it is possible to determine the relationship between a pixel in the image plane with its associated ray of light.

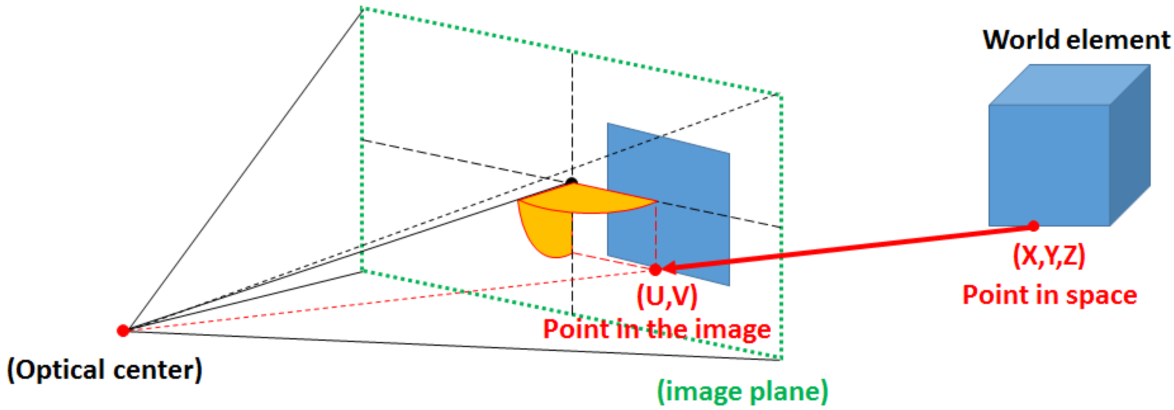


Figure 73: the Pinhole camera model

The pinhole camera model can be expressed in matrix form, as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = CameraCalibration * \begin{bmatrix} x \\ y \\ z \end{bmatrix} = [Intrinsics][Extrinsics] \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Each of these matrices is explored in this section, including its contents and how can be obtained through calibration.

* * * *

A QUICK NOTE IN LINEAR ALGEBRA[8]

The remaining of the section will use matrix expressions to describe the behaviour and calibration of optical devices. For the readers not well versed in linear algebra, the required knowledge is presented below (prioritizing simplicity over rigour):

1. MATRICES CAN EXPRESS TRANSFORMATIONS BETWEEN SPACES

A Matrix groups linear operations in N-Dimensions. The operation we are interested in is the transformation of vectors (for us, points in space), through multiplication:

$$\text{Matrix}_{3x3} * \text{Vector}_{3D} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} m_{11}x + m_{12}y + m_{13}z \\ m_{21}x + m_{22}y + m_{23}z \\ m_{31}x + m_{32}y + m_{33}z \end{bmatrix}$$

The trivial operation that gives the same input as output is called the **Identity** matrix (I):

$$\text{Vector}_{3D} = I * \text{Vector}_{3D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

More interestingly, we can use matrices to ***scale*** and ***rotate*** a point in 3D (among other uses not relevant for our work):

$$Scale_{xyz}(a, b, c) * \text{Vector}_{3D} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$Rotate_x(\theta) * \text{Vector}_{3D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & -\sin(\theta) \\ 0 & sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$Rotate_y(\theta) * \text{Vector}_{3D} = \begin{bmatrix} cos(\theta) & 0 & sin(\theta) \\ 0 & 1 & 0 \\ -sin(\theta) & 0 & cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$Rotate_z(\theta) * \text{Vector}_{3D} = \begin{bmatrix} cos(\theta) & -sin(\theta) & 0 \\ sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Since matrices express linear operations, that means we cannot have independent factors (that is, constant translations). Here is where affine transforms come into play.

2. AFFINE TRANSFORMS ALLOW US TO ALSO PERFORM TRANSLATIONS

Affine transforms have an additional column (displacement) and an additional row of zeroes, with its bottom-right corner equal to 1.

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}x + m_{12}y + m_{13}z + m_{14} \\ m_{21}x + m_{22}y + m_{23}z + m_{24} \\ m_{31}x + m_{32}y + m_{33}z + m_{24} \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}x + m_{12}y + m_{13}z \\ m_{21}x + m_{22}y + m_{23}z \\ m_{31}x + m_{32}y + m_{33}z \\ 1 \end{bmatrix} + \begin{bmatrix} m_{14} \\ m_{24} \\ m_{34} \\ 0 \end{bmatrix}$$

A direct consequence of this is that, in addition to the previously supported operations, we can now perform ***translations***:

$$Translate_{xyz}(a, b, c) * \text{Vector}_{3D} = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \\ 0 \end{bmatrix}$$

Note: in order to be able to perform affine transforms on a vector, an additional component w must be added (a process coincidentally called "augmentation"). If this last component w equals 1, the translation will be applied, but If it w equals zero the translation is ignored (which comes handy when transforming directions).

[8] http://www.c-jump.com/bcc/common/Talk3/Math/Matrices/Matrices.html

3. OPERATIONS CAN BE STACKED

The same way we can multiply a matrix with a vector to apply the contained operation, we can multiply matrices to stack their operations.

$$\text{Matrix}_2 * (\text{Matrix}_1 * \text{Point}) = (\text{Matrix}_2 * \text{Matrix}_1) * \text{Point}$$

The result of multiplying two matrices A and B is another matrix C, each cell containing a linear combination of A's rows and B's columns. For 3x3 matrices, the result is:

$$C_{3x3} = A_{3x3} * B_{3x3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{bmatrix}$$

When performing matrix multiplication, it is critical to remember that:

- matrix multiplication is not usually commutative, so order is important
- it is associative; in order to preserve the order of operations, we will multiply right to left
- the amount of rows on the right matrix must match the amount of columns of the left matrix (we will mostly use square matrices)

4. AFFINE TRANSFORMS ARE PERFECT TO REPRESENT RIGID TRANSFORMS

Affine transforms allow us to represent rotation, translations, and scaling, and they can be stacked. As a result, they are perfect to represent rigid transforms (as in, transforms that *do not deform* the space). Based in the operations previously presented, we can then define new transforms:

$$Rotate_{xyz}(\alpha, \beta, \gamma) = Rotate_x(\alpha) * Rotate_y(\beta) * Rotate_z(\gamma)$$

$$Transform = Translate_{xyz}(a, b, c) * Rotate_{xyz}(\alpha, \beta, \gamma) * Scale_{xyz}(d, e, f)$$

The order of the operations contained in the transform matrix does not have to be the one presented here, and multiple partial or complete transforms can also be stacked.

5. (SOME) MATRICES CAN BE INVERTED

The inverse of a matrix represents "undoing" the stored operation. That is not always possible, as not all operations are reversible. Luckily, it is the case for rigid transforms.

$$A * A^{-1} = A^{-1} * A = \mathrm{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

6. THE AFFINE NOTATION CAN IMPLY A HIDDEN DIVISION

With the matrices presented so far, their multiplication with a vector or another matrix will lead to a 1 in the bottom right corner. If this is not the case, then the resulting vector or matrix must be divided by that cell, in order to obtain the normalized value.

$$\begin{bmatrix} x'' \\ y'' \\ z'' \\ 1 \end{bmatrix} = SomeMatrix \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ w \end{bmatrix} = \begin{bmatrix} x'/w \\ y'/w \\ z'/w \\ 1 \end{bmatrix}$$

For instance, This is the technique used to create a perspective effect.

This is all the required knowledge to be able to understand the following section.

A.1.2. Intrinsics: Lens properties

The lens properties are stored in the $Intrinsics$[9] matrix, along with non-linear distortion coefficients. This information describes how the different pixels in the image plane relate with their light paths.

The $Intrinsics$ matrix describes the relationship between the optical centre (where all light paths converge) and the window those paths go through. Concretely, the matrix contains the information of *focal length* (sensor size:f_x,f_y) *principal point* (image centre:c_x, c_y), and optional *skewness factor* (s, zero by default). In matrix form:

$$Intrinsics = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

This way, it is possible to match a point in space with a point in the image plane:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

When the multiplication is applied, we can see that z (the distance between the point in space and the lens) divides the other components, creating the perspective effect:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x * f_x + y * s + z * c_x \\ y * f_y + z * c_y \\ z \end{bmatrix} = \begin{bmatrix} f_x * {}^{x}/_{z} + s * {}^{y}/_{z} + c_x \\ f_y * {}^{y}/_{z} + c_y \\ 1 \end{bmatrix}$$

A.1.3. Distortion

Since optical lenses are not perfect, distortion is usually generated.

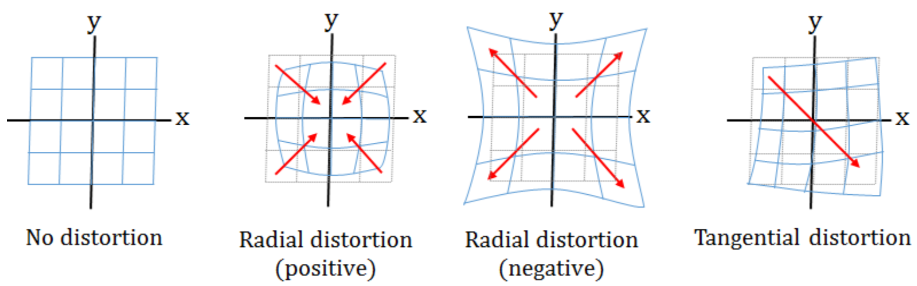


Figure 74: Lens curvature creates radial distortion, while the sensor tilt creates tangential distortions.

The distortion can be taken into account using distortion coefficients (k_1, k_2: radial distortion; p_1, p_2: tangential distortion). They can be used to compute the correct intersection with the image plane:

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} x'(1 + k_1 r^2 + k_2 r^4) + 2p_1 x'y' + p_2(r^2 + 2x'^2) \\ y'(1 + k_1 r^2 + k_2 r^4) + 2p_2 x'y' + p_1(r^2 + 2y'^2) \end{bmatrix} \; with \; \begin{cases} x' = x/z \\ y' = y/z \\ r^2 = x'^2 + y'^2 \end{cases}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'' \\ y'' \\ 1 \end{bmatrix}$$

[9] For the readers that would like to know more about *Intrinsics*, I recommend to visit http://ksimek.github.io/2013/08/13/intrinsic/. It includes an interactive version that allows the user to fiddle with the intrinsic parameters.

A.1.4. Extrinsics: Camera transform in space

So far, the intrinsics matrix only provide information using the camera referential (the origin of coordinates at the optical centre, looking towards Z). In order to change this, we need to know the camera transform in relationship with any given referential. The $Extrinsics$[10] matrix contains this transform. It is important to note that the extrinsic calibration describes a rigid transformation, and as such can be easily recomputed and updated.

$$Extrinsics = \begin{bmatrix} RotationAndScale_{camera} & Translation_{camera} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}$$

As a result, the optical device can be described by the following equation (not including distortion as it requires non-linear operations):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

A.2. Calibration

Camera Calibration is the process of determining the properties of the optical device. The process is conceptually simple: each of the elements contained in the matrices discussed so far is an unknown. By providing pairs of 3D points in world coordinates (x,y,z) and their corresponding 2D points in the image plane (u,v), it is possible to create an equation system to obtain the unknown coefficients.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} ? & ? & ? & 0 \\ 0 & ? & ? & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Luckily for us, solving this equation system is a built-in functionality of OpenCV[11]. Our only concern is then find the best way to obtain the matching 2D and 3D points.

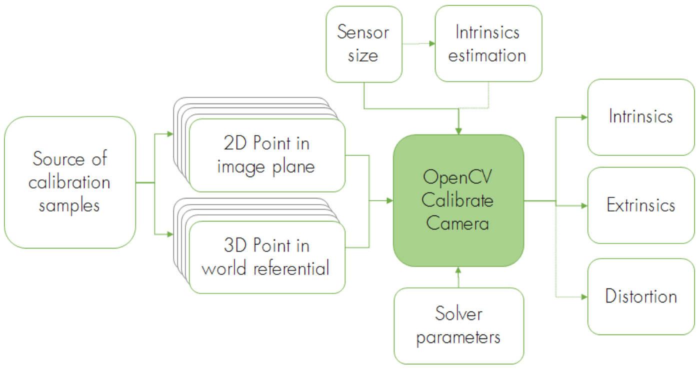


Figure 75: the calibration pipeline.

[10] For the readers that prefer a practical approach, or simply want to obtain more detailed information, I recommend them to visit http://ksimek.github.io/2012/08/22/extrinsic/, where they can fiddle with the extrinsics and intrinsics. Note that this website discusses their usage in the context of rendering. This is covered at the Section 3.4

[11] OpenCV Calibrate Camera Functionality: http://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration/camera_calibration.html

A.2.1. Camera calibration

In order to obtain the samples, we use easy-to-identity features on a known object. The best known example of this is a checkerboard: each of the internal corners is highly distinctive (easy to identify in the image plane), and when the size of a cell is known, then it is possible to use it as a flat 3D object (thus being able to know the corner's position in the referential of the checkerboard). It is worth nothing that the same principle is behind marker-based tracking such as ARToolKit [216].

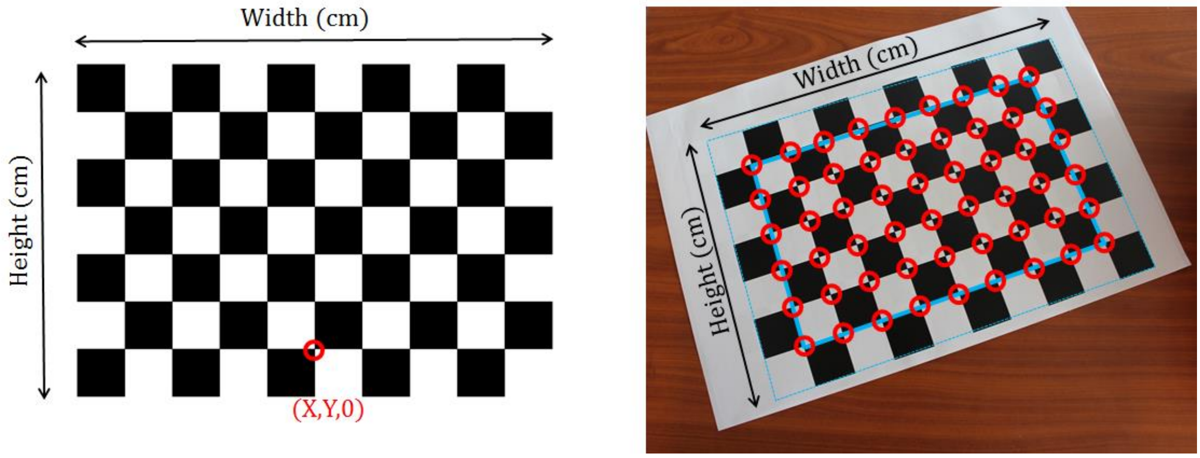


Figure 76: A plane with distinctive features and known dimensions can be used as a referent to obtain pairs of 2D points in the image plane and 3D points in the object referent.

The obtained calibration can be either:

- **2D calibration (Homography):** a homography is a solution to the equation system that is limited to a plane in space. For the case of pinhole cameras, this restricts the calibration to a planar 3D surface. Homographies that do not take distortion into account can be computed with only 4 points (corners), and then applied using bilinear interpolation.
- **3D calibration:** 3D calibration provides an estimation of both the location and orientation of the projector, along with the projector frustum. It is important to know that such calibration is very sensitive to distortion.

A.2.2. Projector Calibration

Even when projectors can be modelled as pinhole cameras, they are effectively "blind", as they cannot see what is in front of them. This makes the calibration process more difficult. Still, many solutions have been explored in the literature, and follow one of three categories:

- **Manual techniques**, where 2D image points and/or 3D world positions are indicated manually. The quality of the results can greatly vary. This technique is well suited when the devices involve have a poor resolution.
- **Semi-automatic techniques** combine computer vision with manual displacement of a known object, such as [7].
- **Automatic techniques** involve the projector actively providing information (structured light). This reduces the amount of manual work, yet it is are highly dependent of the paired sensor and the ambient illumination. Alternatives include pre-calibrated embedded light sensors [125,188]), and projector-camera systems. The latter are either commercially available or free, for both LDSR camera calibration systems (as *Rulr*) and depth sensors (Microsoft Kinect v2 [98]).

A.3. 3D Perception

As can be seen in the pinhole camera model, it is not possible to determine the location in space with only its matching location on the image plane; instead, a ray is obtained. There are several techniques that are used to navigate this problem, the most common ones are:

- **Stereo capture:** for two (or more) calibrated cameras, it is possible to compute the relative transform between them. Then, for a pair of image points, it possible to compute the intersection of the independent rays, giving as a result a unique point in space.
- **Depth from movement:** instead of using two cameras simultaneously, it is possible to use a single camera at a different moment in time and space (i.e., a moving camera). For this technique to work, the moment needs to be correctly computed, which can be performed using both a visual estimation of the movement, or embedded sensors [39].
- **Structured light:** structured light works similarly to stereo capture, but instead of using two light captors, one of them is a light emitter. In order to synchronize captor and emitter, many alternatives exist. For instance, Microsoft Kinect v1 uses a dot based pattern that is projected over the scene, and two stereo cameras reconstruct a point cloud from it. Other devices use either binary encoding of the space (projecting over time a binary pattern of an amount of bits), or sequential scanning (as with HTC Vive and [125]). Sequential scanning is well suited for embedded captors given their low latency compared to cameras.
- **Time of flight:** is a LIDAR (**LI**ght ra**DAR**) technique that computes distance based on the speed of light, by sending light into the scene and computing the time it takes to return to the sensor.

The first two alternatives use similar approaches to the human vision. The last two alternatives illuminate the scene in a non-continuous way. When using visible light, this can interfere with the user experience unless high frequencies are used, and it is usually performed offline (i.e., before the usage of the system). Even if infrared light is used, interference can be created when two or more devices are used simultaneously. Professional grade hardware allows to either change the light wavelength, or to synchronize the different devices, yet this is not possible with most consumer grade hardware (such as Microsoft Kinect v1 or v2, RealSense or LeapMotion).

* * * * *

A.4. From Pinhole Cameras to Virtual Cameras

So far we studied how pinhole cameras relate a point in space to a point in the image plane. When creating mixed reality applications, our interest is not only to see the physical world, but to also add coherent digital information.

In order to align digital information with the physical scene, a parallel virtual scene must be processed and rendered. Such a scene can be obtained either by manually modelling or 3D scanning, and contains the elements that are required for the augmentation, and in most cases these representations are both incomplete an imperfect.

This virtual scene is then looked at through a virtual camera (i.e., rendered). In order to correctly align both spaces, the virtual camera properties need to match the physical device (camera or projector).

A.4.1. What is a Virtual Camera?

It is important to notice that the properties of a pinhole camera and the virtual equivalent used in 3D rendering differ slightly, particularly on the intrinsic matrix. In rendering (APIs such as OpenGL or DirectX) the intrinsic matrix is referred as *projection* or *perspective* matrix, and contains not only the field of view of the camera, but also determine which elements are *inside* or *outside* of the rendering space. For that, two additional parameters are used: *near plane* (too close to the camera) and *far plane* (too far to be rendered) are included. Regarding the $Extrinsics$ matrix, its equivalent in rendering is called the $View$ matrix. The virtual camera then describes the following Frustum:

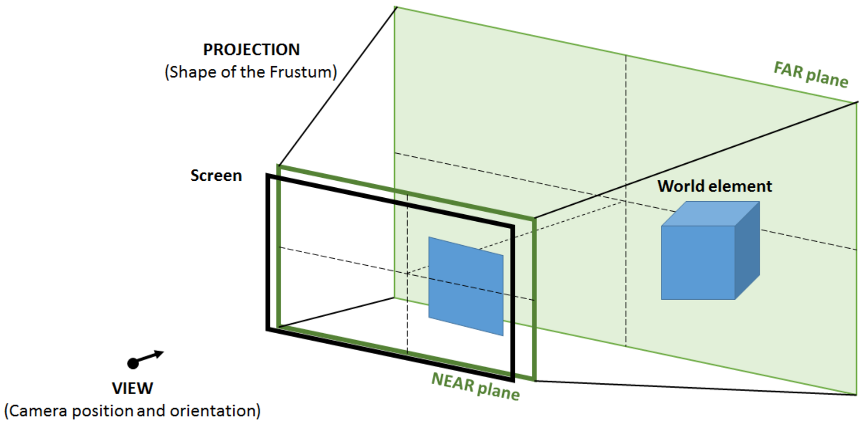


Figure 77: Camera model using for rendering, it is a pinhole camera with near and far clipping planes

A virtual camera is then described by:

$$[u \quad v \quad d] = Projection * View * Point_{xyz}$$

A.4.2. Projection and View: Intrinsics and Extrinsics for rendering

As was previously mention, the $Projection$ matrix is the equivalent of the $Intrinsics$ matrix, while it also includes $near$ and far planes, and it is resolution independent.

$$Projection_{perspective} = \begin{bmatrix} size_x & 0 & center_x & 0 \\ 0 & size_y & center_y & 0 \\ 0 & 0 & -\dfrac{far + near}{far - near} & -\dfrac{2 * far * near}{far - near} \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

When no relationship with a physical camera is needed, the projection matrix is computed by looking at an arbitrary rectangle, using its corners (namely, right, left, bottom and top) and clipping planes.

$$Projection_{lookatquad} = \begin{bmatrix} \frac{2*near}{right-left} & 0 & \frac{right+left}{right-left} & 0 \\ 0 & \frac{2*near}{top-bottom} & \frac{top+bottom}{top-bottom} & 0 \\ 0 & 0 & -\frac{far+near}{far-near} & -\frac{2*far*near}{far-near} \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

There are some minor adaptations required to use the Intrinsic information obtained during calibration in order to define a virtual camera. First, instead of using pixel coordinates (from 0 to width horizontally, and from 0 to height vertically), it uses normalized coordinates (-1,1). Then, near and far planes need to be defined, based on the properties of the scene. Distortion coefficients need to be addressed manually, via post processing or custom rendering techniques. When $Intrinsics$ are used to define a $Projection$ matrix, the normalization gives the following result:

$$Projection_{intrinsics} = \begin{bmatrix} \frac{f_x}{width} & 0 & 2*(c_x - \frac{width}{2}) & 0 \\ 0 & \frac{f_y}{height} & 2*(c_y - \frac{height}{2}) & 0 \\ 0 & 0 & -\frac{far+near}{far-near} & -\frac{2*far*near}{far-near} \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

The projection matrix works by transforming a point in the space to a normalized value. A point is inside the rendering volume when it possess values between -1 to 1 for the image plane (u,v), and a value between 0 and 1 for the distance (d) to the camera. Otherwise, the point is not visible.

$$\begin{bmatrix} u \\ v \\ d \\ 1 \end{bmatrix} = Projection \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, render\ if \begin{cases} -1 \le u \le 1\ and \\ -1 \le v \le 1\ and \\ 0 \le d \le 1 \end{cases}$$

Note that even when the distance is stored in a normalized way, it can be reconstructed to world units by using the near and far planes.

The $Extrinsics$ matrix is called $View$ matrix, and it contains the transform from world coordinates to camera coordinates (which is the inverse of the camera transform in world coordinates):

$$View = Extrinsics = \begin{bmatrix} Rotation_{camera} & Translation_{camera} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}$$

A.4.3. Objects in Space

Digital scenes are not assorted list of points in space, but instead hierarchical collections of objects. A given point can be part of an object or group, which in term can be part of a bigger structure, and so on. For practical reasons, objects are usually described in reference to an arbitrary local centre, and then *placed* in space. The information regarding the relationship between local coordinates and world coordinates is contained in the $Model$ matrix (sometimes called *Model-To-World* transform).The $Model$ matrix is computed as the recursive multiplication of the current space transform with its parent transform. To complete the equation constructed so far, a point in a given local space relates with the image plane by:

$$[u \quad v \quad d] = Projection * View * Model * Point_{xyz}$$

A.5. Using Physical and Virtual Cameras

So far, the equations presented describe the relation between an image and the space they represent. Taking the liberty to express points for the pinhole camera as belonging to an object, we get similar equations for pinhole and virtual cameras:

$$[u \quad v \quad d]_{virtual} = Projection * View * Model * Point_{xyz}$$

$$[u \quad v \quad d?]_{camera} = Intrinsics * Extrinsics * (Object * Point_{local\ xyz})$$

Using these equations, it is possible to go from 3D spaces into 2D images, and vice versa (yet I remind the reader that one static camera on itself is not enough to obtain 3D information, as mentioned in **Section A.2**). This enables us to create images out of 3D models (Render), while also allowing us to create 3D models out of images (scan), or to detect and follow a given object moving across a scene (detect and track).

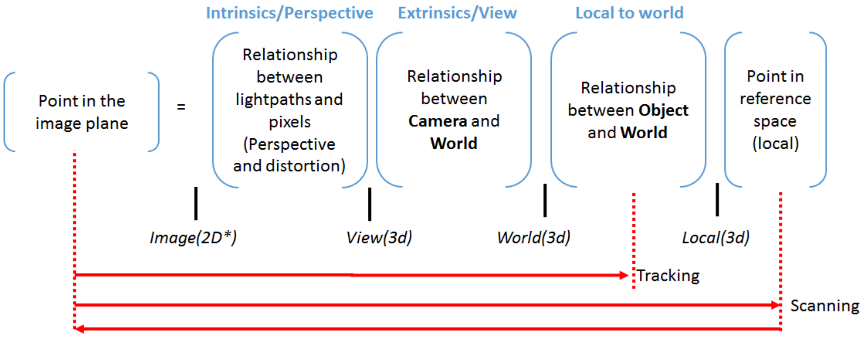


Figure 78: Different usages of these matrices

These are by no means the only approaches to do Rendering and Computer Vision, but this Chapter provides a high level introduction to the techniques used to implement the systems described in this manuscript. Other important issues include involve structure and appearance, yet we refer the reader to their graphics book of preference to know more about these subjects.

* * * * *

A.6. Implementing MR illusions

In Chapter 3, the different possible ways of create augmentation are introduced, yet he technicalities are not covered keeping the explanations at a conceptual level. This section goes a little more into detail (yet still does not cover all the aspects).

A.6.1. Augmented surfaces through projection mapping

Surface augmentation implies changing the aspect of physical objects, and in our case it is implemented using projectors. If the virtual equivalent of a physical scene is rendered from a calibrated projector's perspective, each pixel will fall on its corresponding physical location.

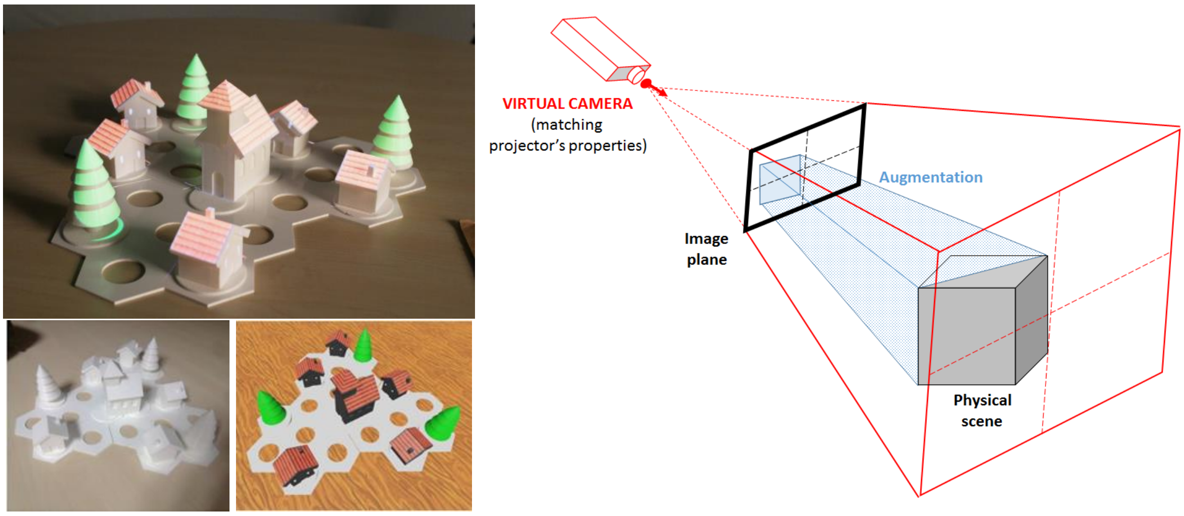


Figure 79: Texture mapping (left) is the result of combining physical and digital (left-bottom) counterparts with the same geometry. This is achieved by rendering a virtual scene with the same geometry as the physical one, using a virtual camera that shares all its properties with the projector used (right).

This technique requires:

1. For each projector, a virtual camera matching its properties
2. Knowledge of the geometrical properties (shape, position, orientation) of the physical elements to augment
3. Desired appearance to apply to the physical objects
4. Ideally, knowledge of the material properties (colour, reflectance, opacity) of the physical surfaces to augment
5. Ideally, knowledge about any occluding elements (to prevent accidental augmentation)
6. Optionally, knowledge about user(s) head position, to create non-diffuse illumination effects

Besides the projector calibration and the requirement to know the physical geometry, this technique does not require any additional challenging steps.

* * * * *

A.6.2. Windows

A window provides a view onto a scene through a frame. There are two cases: fixed-perspective and perspective-corrected windows.

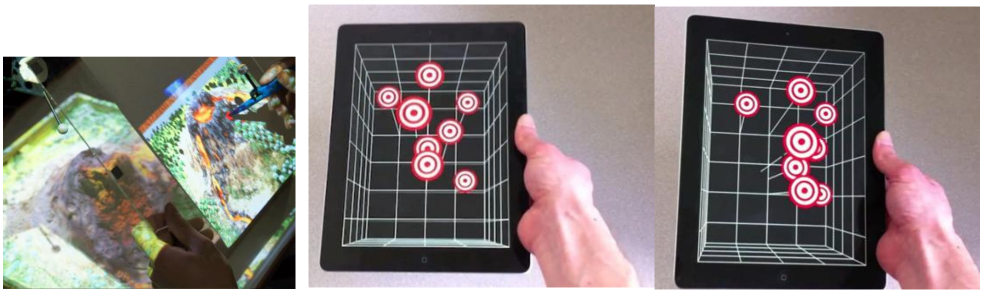


Figure 80: left: see-through augmented reality using a fixed-perspective. Right: Fish tank effect, even when the screen is flat, the perspective correction gives the illusion of depth.

In both cases, they are implemented as a *Camera* facing a *Window* (Figure 82). The difference between fixed perspective and perspective corrected windows is how the camera *Frustum* is constructed (Figure 82):

- **Fixed-perspective windows** possess a static relationship between the *Camera* and *Window* locations, moving together across space; this reason the shape of the Frustum stays fixed. This is the case of traditional see-through Augmented Reality (Figure 14-top).
- **Perspective-corrected windows** allow the *Camera* to move independently of the *Window*, and for this reason the shape of the Frustum can change. The independent movement is commonly addressed by making the *Camera* look at the *Window* centre. The most known example of this is the fish-tank technique [219] (Figure 14-right), where the Camera transform is associated with the user's head position, creating a similar effect of looking through an actual window.

Fixed-perspective windows can use an arbitrary relationship between camera and window, yet it remains fixed. When used for video see-through technology (as in the case of classic mobile AR), the simplest solution is to use a virtual camera matching the properties of the physical camera. If no depth information is available for the camera, the video feed can be displayed a full screen quad at some arbitrary distance from the camera (thus being able to occlude elements if desired).

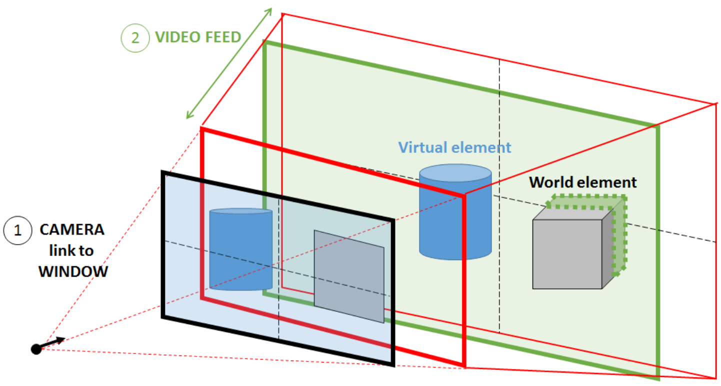


Figure 81: For fixed perspective windows, the camera position relationship with the window does not change. For video-see through applications, the virtual camera simply match the physical camera.

Perspective-corrected windows require to compute both view and perspective matrices, based in the relationship between *Camera position* and *Window transform* (position, size, and orientation in space).

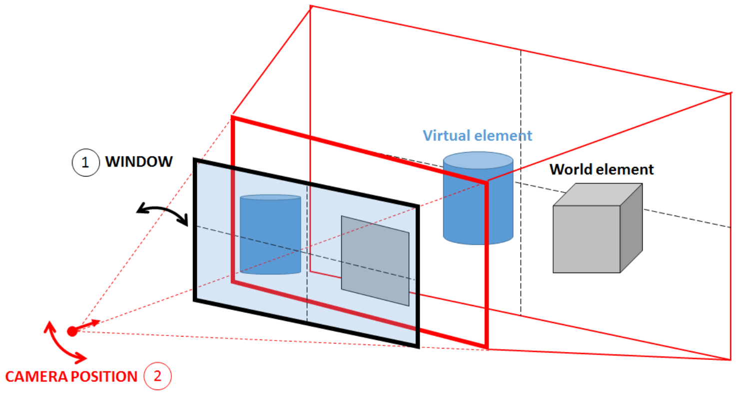


Figure 82: The window (1) can move and rotate freely if desired, while the camera position can have either a fixed relationship with the window, or move independently; either way, the orientation of the camera is computed based on the relation between (1) and (2). It is important to mention that the virtual plane is not required to be closer to the camera than the near plane, being then possible to display elements both in front and behind the window.

The view matrix is computed by placing the camera at its correct location, and then orienting it towards the window:

$$View_{lookatwindow} = \begin{bmatrix} RotationAndScale_{window} & Translation_{camera} \\ 0 & 1 \end{bmatrix}^{-1}$$

The perspective matrix is computed by first finding the camera location in the referential of the window, and then looking at the window from that position. Sadly this step is not trivial, and I recommend the reader to refer to the literature.[12]

In order to implement augmentation using windows is then necessary to know:

1. Position and shape of the *Window*
2. Desired virtual information to display
3. For *perspective-corrected windows*, knowledge about user's head position
4. If video feed is involved, *intrinsic* and *extrinsic* camera properties, and
5. Ideally, knowledge about any occlusions in the scene (i.e., depth information)

When windows are implemented using projection, a two pass render is used. On a first pass, the window view is rendered, and on a second pass the resulting image of the first pass is applied as a texture to an opaque rectangle.

* * * * *

[12] http://csc.lsu.edu/~kooima/articles/genperspective/

A.6.3. Arbitrary surfaces

The idea behind Windows can be extended to arbitrary surfaces. A classic example of this are CAVEs, room size spaces, composed of juxtaposed flat or curved surfaces resulting on an immersive space around the users. Originally, the idea of Spatial Augmented Reality was based on the extension of CAVEs to everyday surfaces.

Figure 83: Spatial immersive augmentation: Square CAVE (left), dynamic perspective-corrected projection over an arbitrary environment (right, the box and the bluish hexagon are projected)

When considering arbitrary SAR, it is best understood by CAVEs. Traditional CAVEs have a limited amount of flat walls, each of them behaving as a perspective-corrected window. A more complex surface can be seen as a juxtaposition of many smaller windows, reducing their size up to the point that each window involves a single light path.

From an implementation point of view, this requires to know (1) what should be the final scene from the user's perspective, and then to determine (2) where lays the physical geometry that will support the projection. The technique involved is very similar to projective shadows: imagine the user as a source of light, a given virtual object laying at some location on the scene will cast a shadow over the first surface behind them. The same approach used to implement shadows can come at hand at the moment of implementing arbitrary SAR augmentation. First, the user's viewpoint is rendered and stored as an image (texture). Then, the physical geometry is rendered from the projector perspective. When a point is rendered, it is possible to know where it lays from the users POV, and this information is used to query the previously rendered image.

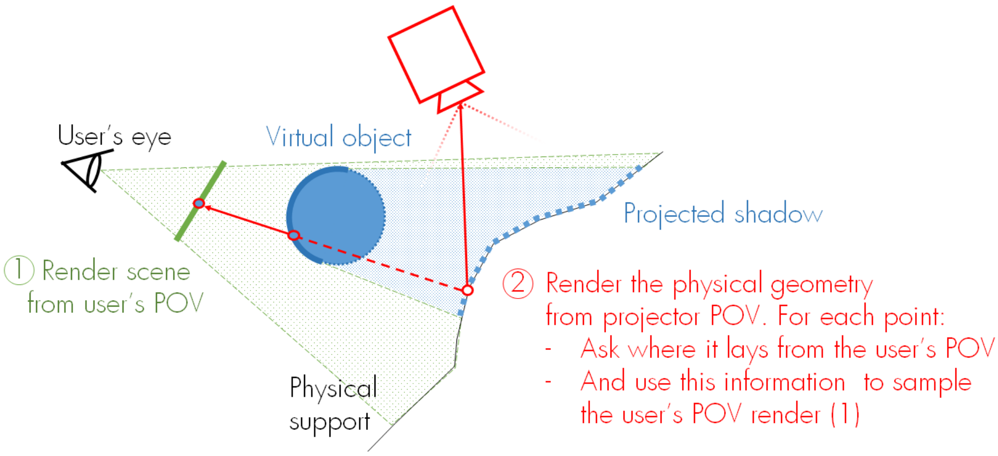


Figure 84: In order to project arbitrary virtual elements over non-flat surfaces, it is necessary to determine (1) where a given virtual elements is in relationship with the user's field of view, and (2) where is the supporting geometry. This process is similar to projective-shadow rendering.

A.7. Stereoscopy

Perspective-corrected windows allow the users to estimate depth from movement, yet humans not only use movement to estimate 3D locations, but also stereo-perception. The described techniques can be easily extended to stereoscopy, by having not a single *Camera* position but instead one per eye. The main difficulty of stereography is not how to create the multiple viewpoints, but how to display different information to each eye. There are two main approaches, which involve either Auto-stereoscopy (materials that send light at different angles), or eyewear that provide eye-specific information. The latter can involve either filtering glasses (by either passively filtering light wavelength or polarization) or screens directly placed at the eye level (such as in the case of head mounted displays).

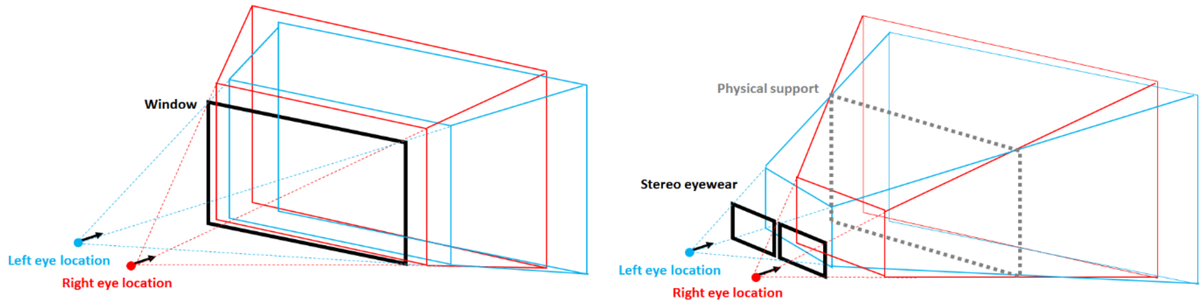


Figure 85: stereo displays show different information to each eye. The information can be provided by a single window in space, or by different windows, one for each eye. How the information is placed on these windows is not of relevance for now.

A.7.1. Issues caused by Differences between Physical and Digital Geometries

When looking at pictures and illustrations from this Chapter, the illusions created seem to work particularly well, yet this is not always the case: these illusion are fragile, and can cause discomfort. This happens because the perceived image and the screen displaying it lay at different distances from the user (Figure 86). Stereoscopy causes discomfort at the eye level, since rapid muscular accommodations happen constantly (called *vergence-accommodation* conflict)[151], while 2D displays cause discomfort at a perception level, given the contradictory information. These effects are even more noticeable for non-flat displays and at short distances. It is important to mention that surface augmentation does not suffer from these conflicts.

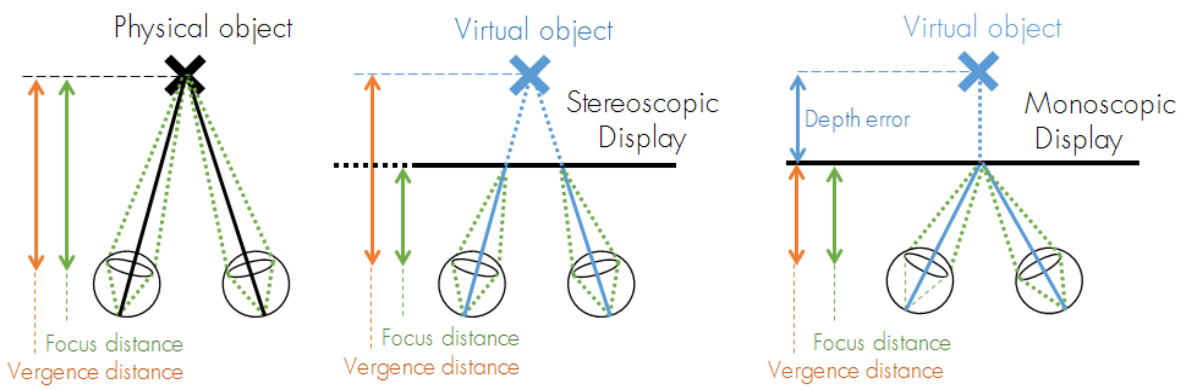


Figure 86: Sources of discomfort when support and virtual object differ in location. No conflict (left), vergence-accommodation conflict (middle), and lack of depth information (right).

Appendices: Technical and Personal

B. the *actual* path

"Life is mostly outtakes."
- Chris Hadfield

The narrative presented in this thesis is the result of the PhD process. This means that the orders of the chapters are sorted for clarity. For other students like myself, I would like to share a brief faithful representation of the process, including the actual order the research was performed, and the failed attempts and dead ends along the way.

B.1. Struggling with Spatial Augmented Reality

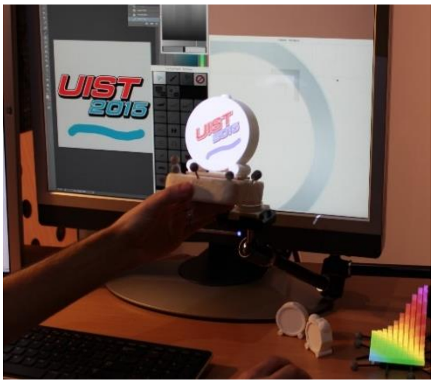
The main topic at the beginning of my thesis was the rather broad question of *"how can we interact with (projection-based) Spatial Augmented Reality?"* My strength was not HCI in particular, since I come from software engineering with some basic knowledge on computer vision and graphics.

Given the technical background, I started by trying to understand the underlying technology. My first impression was quite discouraging ***(Section 3.6, Projection-based Augmentation)*** since the usage of projection is a very challenging (and unrewarding) task. Commercial projectors are designed to create a flat screen on a specific surface and a controlled environment (e.g., dim ambient light). As a consequence, their usage to project over tracked moving objects is far from ideal. An important part of my first year involved then trying to tackle calibration **(Appendix A)**, which I tried to document as clear as possible in this manuscript because the literature was not clear enough in my case. The struggles with calibration lasted for a large part of the PhD, only becoming clear towards the end ***(Chapter 3, Mixed Reality).***

The main concern since the beginning was the rationale behind using SAR (I myself asked *"why?"*). Slowly, the notion of the physical world become of importance. Bear in mind, I was not particularly aware of the benefits of tangibility at the time, as I had a 3D interaction perspective.

B.1.1. The first paper (and first failure)

Since the beginning, ***Renaud Gervais*** (a recurrent name in this manuscript) was there sharing his knowledge and experiences regarding SAR and tangibility, and he accepted to collaborate on Tangible Viewports. Being his original topic to work in stereo screens and having worked on pointing on objects [68], the idea of placing objects in front of the screen made much sense. I helped him finish a proof-of-concept, resulting on a paper submission for ***UIST'15***: it got bluntly rejected, because there was no interaction beyond cursor based, so there was no point in using a *physical object as a display*.



We made the mistake of shooting the whole video using UIST'15 logo, so we had to reshoot it in order to resubmit to TEI.

This was an opportunity to extend the interaction to the physical world, and we included the supported interaction space **(Chapter 3.2)**. With this extension, the paper got accepted half a year after at ***TEI'16*** **[TEI16b] (Chapter 4).**

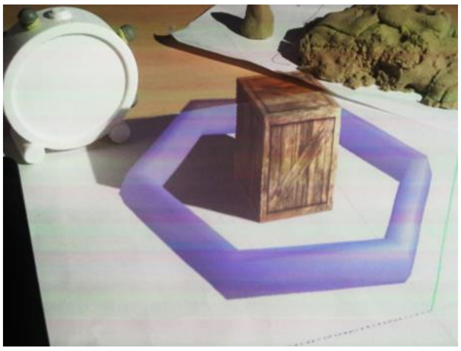
B.1.2. Experimentation with Microsoft Kinect

Going back to the calibration fight, at the 6 month mark of my PhD I was able to use existing tools, but could barely understand what was happening underneath. This was infuriating, as I was totally unable to solve problems, in particularly the small -- yet always present – projection error. Since I wanted to reduce the calibration time (a haphazard process sometimes taking hours), I wanted to use an automatic technique: structured light. For this, Kinect seemed to be the best solution since it has 3D sensing capabilities. This took me the good part of 3 months. The source of the main bug was that I was wrongly aligning colour and depth cameras. Once it worked, the calibration error was as bad as with manual calibration. But now I knew how to do it (that is what I believed, and *oh boy! I was wrong*).

B.1.3. Anamorphosis

The main motivation behind the usage of Kinect was the desire to extend Tangible Viewports to the desk (or any arbitrary surface). It made sense at the time: the technology was working, and we were trying to move the interaction outside computers (but we were still thinking in widgets).

The rationale was then: if we use head tracking to point, we can then create perspective dependant renderings (these are called *anamorphic illusions* or *anamorphosis*). Another important aspect of the use of scanning was to replace 3D printing with *hand-sculpted objects* (you can see in the picture of the left a big stack of polymeric sand). After some struggle, the anamorphosis worked, but never with scanned surfaces (as they were too jittery). Either way, it was an acceptable addition to Tangible Viewports, and was presented at a doctoral symposium **[IHM15s]**. In this manuscript, this time-period is only reflected on this picture and the last drawings of the Chapter for Tangible Viewports **(Section 4.7)**.



This is one of the few pictures I have of the anamorphosis project (box and hexagon are projected). It shows the transition between tangible viewports (clock on the left side), and Inner Garden (polymeric sand on the right side).

B.2. Getting a grip of tangibility

Slowly, thanks to long discussions with colleagues, and recurring elements from the literature pointed towards a clear advantage of SAR: it can be transparent, as a way to support the interaction with the world around us. This opened a door that was completely new for me: tangibility (beyond the notion of *"digital objects that you can grab"*).

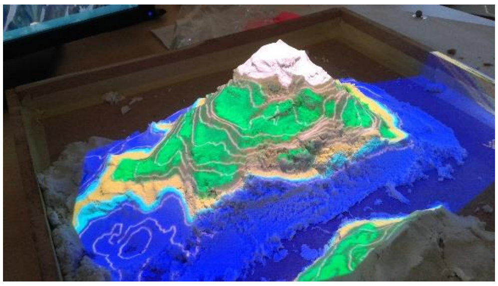
I have to admit my initial scepticism about this topic, as the first papers I read were very puzzling: Implementation was not at the core, there was no code present, nor algorithms. Nevertheless, there was a game changing element in these works. They had a higher perspective, there was a vision. Besides the vision of a better future, there was a clear influence of design, and it was good **(Chapter 2.2).** From the notions of *embodiment* and *metaphor*, to the materials and experience design. Before this moment, I barely ever considered that emotions and Aesthetics were variables of interfaces. After going deeper in the field, it was clear that Interaction, design and art were intertwined. As a follow up to the notion of tangible interaction, I stumbled into ***Bret Victor*** and ***Sebastian Deterding***, which pointed towards more *humane* interfaces with explicitly considered values.

Since then, physicality (in its broad sense) became an important part of my work, yet I wish I had more time to go deeper into fabrication (in part thanks to the passionate explanations of ***Thibault Laine***).

B.3. A sandbox

At this point I was reaching the end of the first year of my PhD, and I had nothing concrete yet. My only results were fragmented understandings of calibration, rendering, and that I wanted to move the interaction towards the environment. I had the polymeric sand from the failed *anamorphosis* project, and I decided to try to reproduce a project I was part (in no significant way) during my undergrad studies: EfectoMariposa (a sandbox, much like *illuminating clay* [154], where the terrain takes the shape of an tropical island). My main objective at this time was to try to reproduce a *god game* [129], where you could shape a terrain and a civilization will grow by itself.

Next to my desk there were ***Renaud Gervais*** and ***Jérémy Frey*** working on physiological computing. They had built TEEGI the year before (a puppet/avatar showing EEG information) [63] and were at the time working on TOBE (a customizable avatar/toolkit to work with various physiological signals)[67]. With their help, we added physiology to the island. Then, it was clear, this looked very much like a Zen garden, and could be used for the same ends. Using technology to relax, and connect with your inner state. Literally, *calm computing*. This idea got accepted as a poster (and a demo) at TEI'16 **[TEI16a]**.



This is the first version of Inner Garden, which had level curves, rivers and some living creatures that for some reason built grids by eating the trees. These creatures built campfires during the night.

B.3.1. The Absurd

While I was working on a meditation sandbox, I was working crazy hours and was very stressed with some personal issues. Being my first international conference and traveling only for a poster, I applied for the student volunteer program (and got accepted). Excited and stressed, I sprained my foot the day before we left for the conference. The demo at TEI'16 was barely a demo, because the Kinect calibration did not work. This technology for well-being was not helping my health.

From left to right: Renaud, Jérémy, me, and the sandbox.

B.3.2. The follow-up

With all these problems, the reception for *Inner Garden* was great. As we contacted people we thought they will be interested, we progressively improved the system. We sent a full paper to ***UIST'16*** describing the process (without too much emphasis in the interviews): it got rejected. This time, the reviews were encouraging, and with some restructuring of the document (taking a more design approach), it got accepted at ***CHI'17*** **[CHI17]** ***(Chapter 8, Introspectibles and Inner Garden).*** What was supposed to be a side project lasted over a year, as studies, demos and outreach events chained with each other (the system not always worked). In retrospective, *Inner Garden* looks nothing like I envisioned, but I am very happy with the results.

B.4. A final iteration on calibration, and a book

In between the Inner Garden studies, I stated working with ***Julia Chatain*** on mobile projection. This was a challenging task, and it required a clear understanding of what we were doing (in contrast with previous projects that shared the same tools). Trying to share my knowledge forced me to formalize something that was more a set of intuitions than anything else, and at that point *I finally understood what I was doing.* We managed to achieve the mobile calibration, yet the scope of their project finally changed and our work was not required. On a complete different topic, ***Julia*** convinced me to read Diamond Age [194], from where I took the inspiration for an augmented book **(Chapter 9.2)**.

B.5. An HMD at hand

After we did the first version of Inner Garden, we got an Oculus HMD (for our colleagues doing VR). Going back to the idea of doing god-games, I wanted to be able to use VR as a way to remove the surrounding environment for an immersive experience, and then travel to the surface of the island. Around the same time, ***Jérémy*** did a *hackaton* in Canada where he explored meditation in VR using biofeedback. We then decided to add it to the Inner Garden.

B.5.1. Combining SAR and VR

At this point I knew how to track objects, to perform rough surface scanning, and to combine SAR and VR. One of the nice features of VR was that it did not share all the limitations of SAR: the graphics were great, it supported information mid-air, and to override reality in general. That said, SAR kept the user focused onto the physical environment, and by now I had seen the undeniable benefits. So, the idea of use them to complement each other was the logical next step. A work-in-progress with this premise got accepted at ***3DUI'17*** **[3DUI17]**, including the usage of the VR HMD to change the viewpoint of the physical scene (digitalized using a combination of Kinect scanning and manual 3D modelling).

B.5.2. An inspirational bug

I still remember that at the time there were some problems with the Microsoft Kinect adapter, and the Kinect froze after 90 seconds of usage. As a result, your avatar would freeze at a random moment, and you will see a froze-in-time representation of your own body, creating a very unsettling out-of-body experience *(a virtual death).* This brought to my attention *time*, and in combination with the sensitivities created by the *Introspectibles* project, steered me towards the idea for ***Chapter 9.1.***

B.5.3. What is *real?*

The feeling of "virtually dying" created a very real discomfort, so I started asking myself: *what is real?* With SAR, the reason of failure of the illusions was a perceptual one (the augmented scene had conflicting information). Desynchronizing your mind and body involved a much deeper understanding of the space; it was not coincidental that VR was used as a mean to better understand the human mind. Around this time (and perhaps the reason this was my main concern), the team presented a rotation of members: ***Renaud*** and ***Jérémy*** finished their PhDs, and a new wave of students arrived, most of them with a background in Cognitive Science (***Philippe Giraudeau***, ***Pierre-Antoine Cinquin*** and ***Léa Pillette****)*. Then, the *mind* became topic of discussion in my free time.

From this point on, my main concern was about the construction of cognitively coherent spaces (more or less the topic of the final version of ***Chapter 2*** and the narrative presented in ***Chapter 1****).* A full system and the notion of incremental augmentation was submitted to *UIST'17* ***(Chapter 5, One Reality)***, and got accepted (it took 3 trials to get accepted at UIST) **[UIST17]**. The main critic we got for this submission was the lack of evaluation: to avoid providing a weak one, we provided none.

B.5.4. A proper evaluation

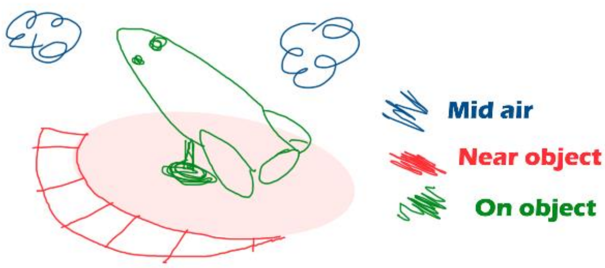
At the moment this manuscript is being written, we are waiting for an answer for the dedicated study submitted to CHI’18 **[**

Indexed publications

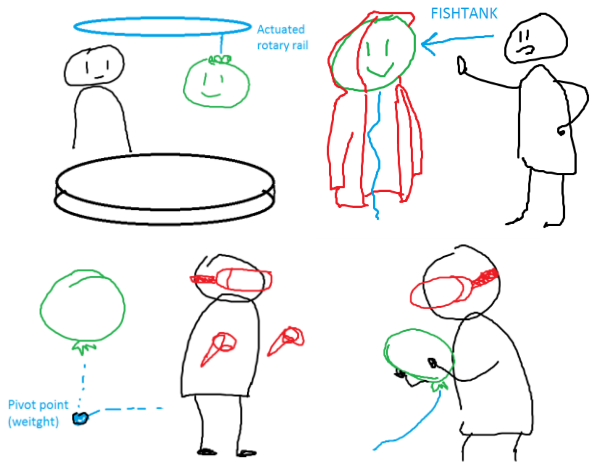
CHI18] *(Chapter 6)*. This study is the result of the discussion with my colleagues from Cognitive Science, with a more rigorous approach to the evaluation process (materialized thanks to the help of ***Pierre-Antoine*** and ***Jean Basset***). The evaluation is only the first step towards the validation of hybrid mixed reality spaces, yet as far as I am aware, an important step from the HCI standpoint. This study is not only beneficial for HCI because it takes inspiration from Cognitive Science, but could also help Cognitive Science: if VR can tell us a lot about the brain, perhaps Mixed Reality can also help. Further iterations of the protocol could lead to a formally robust methodology to evaluate both human *spatial representation capabilities* and *interactive systems.*

B.5.5. Another casual collaboration

Rewinding a little: while I was working on for 3DUI’17, I needed the infrastructure to place all the cameras, so I moved to the room where ***Damien Clergeaud*** (a friend and colleague) was working. ***Damien*** is a PhD student working in immersive collaboration in VR for Airbus. As we discussed casually about frustrations and progress, it become clear that the technology I was working on (and mixed reality in general) could be of great use for remote collaboration. More generally, hybrid MR was a nice approach to link remote locations (either physical or digital).



This is my first sketch of how we could collaborate. The result was finally quite different, but our discussions helped me a lot to clarify what ended up being One Reality, Chapter 5 of this manuscript.



Balloons are floating tangibles. They offer a support for projection (top), follow people at a fixed distance *when attached to them (bottom-left), and can serve as a tangible prop (bottom-right).*

As it was usually the case during my PhD, the original project (combining SAR and VR) got delayed, and it was not before 8 months after we started discussing with ***Damien*** that we finally implemented our ideas, and what was originally a broad framework (around the construction of spaces) was narrowed to concrete instances because of time constraints ***(Chapter 7, Asymmetric Collaboration in Mixed Reality)***, which was accepted at ***VRST'17*** **[VRST17]**.

This is one of the things I regret the most, not having the time to push further our collaboration, as it included ideas such as theatrical interfaces (providing physical objects with virtual/magical properties by using pulleys) [33], and the use balloons as floating tangible props for SAR and VR. Over a year before, ***Renaud*** used to mention balloons once in a while *("they are slow, they float, and they provide a slow friendly tangible support for projection")*, but my mind was somewhere else at the time. The idea of using balloons reappeared because there was a celebration at our building and there were helium balloons floating around.

B.6. The inspiration along the way

I hope that it is clear by now that the ideas presented in this manuscript are not only mine, but the result of the very fruitful interactions with others.

Outside of the people around me, I personally find the talks of ***Bret Victor*** [13] and ***Sebastian Deterding***[14] to be the ones to influence the most. Our concrete results might differ from what they propose, but they were my main sources of inspiration. They are great public speakers, and they talk to both the technician and the game enthusiast in me, while addressing the modern Human condition. It is fair to say that ***Chapters 4, 5 and 7*** very much align with ***Victor's*** "seeing spaces" [214], ***Chapter 8*** is my personal attempt to use ***Deterding's*** playful design [44], and more generally the vision presented is the result of being exposed to theirs (which in terms drives inspiration from others). I would have loved to be able to explicitly explore more of their dissertation topics in these pages, yet I find their influence on this manuscript is undeniable.

B.7. An overview

This thesis is then the result of a simple, yet broad question: *"how do we interact with spatial augmented reality?"* as a by-product of trying to answer this question, I got exposed to many technologies and theoretical concepts completely novel to me; more importantly, I got exposed to a change in paradigm triggered by the ability to interact with the physical environment. The narrative presented in this document reflects my current vision (at the moment of writing the manuscript), but it is the result of an exploratory approach -- in most cases driven by curiosity or technical challenge -- rather than based in a well-defined path.

---

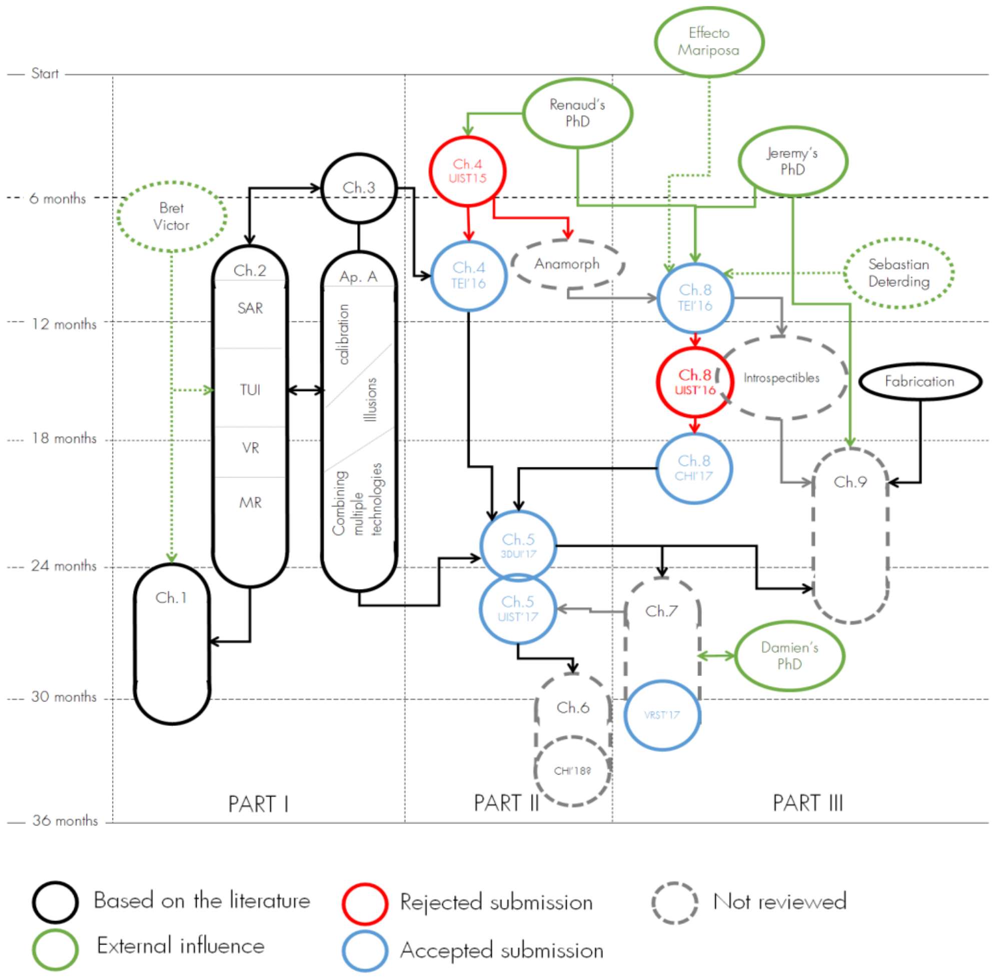
[13] I strongly recommend to watch Bret Victor's *"The Humane Representation of Thought" (*Closing keynote at UIST'14), as it was one of the strongest inspirations behind this thesis. It is available at Victor's official website: http://worrydream.com/#!/TheHumaneRepresentationOfThoughtTalk

[14] I recommend Sebastian Deterding's "Designing against Productivity". An exploration of how technology carries values, and that we need to embrace (and reconsider) our design decisions. It is available at MIT's website: https://www.media.mit.edu/videos/wellbeing-2015-11-24-2/

The objective of this Chapter is to show that even the failures sometimes served as stepping stones, allowing us to reach contributions (even by simply buying a type of sand that works well as modelling clay). Here is where I must recognize the advantages of my work environment: at *Inria-Potioc* I had both the access to materials when needed, and people with complementary profiles to discuss potential ideas. Without these, many of the presented projects would not have been possible.

The idea behind this Chapter comes from Warren Bergers' book [18], which explores how innovation comes not from the thin air, but instead of people (that might call themselves designers) combining the knowledge and needs from a field with their own subjective perspective and insights developed over time, by working in heterogeneous projects. Knowledge, even when formal, changes the way we look at the world. We should not be blinded by our subjectivity, but perhaps we should sometimes embrace it.

I want to finish this Appendix with a graphical representation of the path (next page):



This image tries to overview the core influences and relationships between projects.