UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS

ÉCOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

THÈSE

pour l'obtention du grade de

Docteur en Sciences

de l'Université de Nice-Sophia Antipolis

Mention : INFORMATIQUE

Présentée et soutenue par

Emmanuelle CHAPOULIE

Gestes et Manipulation Directe pour la Réalité Virtuelle Immersive

Thèse dirigée par George DRETTAKIS

préparée à l'INRIA Sophia Antipolis, Équipe REVES

soutenue le 30 Juin 2014

Jury:

Rapporteurs :	Bernd FRÖHLICH	-	Bauhaus-Universität Weimar
	Pascal GUITTON	-	Université de Bordeaux - INRIA
Directeur :	George DRETTAKIS	-	REVES / INRIA Sophia Antipolis
Président :	Sabine COQUILLART	-	PRIMA / INRIA Rhône-Alpes
Examinateurs :	Maud MARCHAL	-	HYBRID / INRIA Rennes / IRISA
	Martin HACHET	-	POTIOC / INRIA Bordeaux
	Sabine COQUILLART	-	PRIMA / INRIA Rhône-Alpes

UNIVERSITY OF NICE - SOPHIA ANTIPOLIS

DOCTORAL SCHOOL STIC

SCIENCES AND TECHNOLOGIES OF INFORMATION AND COMMUNICATION

P H D T H E S I S

to obtain the title of

PhD of Science

of the University of Nice - Sophia Antipolis

Specialty : COMPUTER SCIENCE

Defended by

Emmanuelle CHAPOULIE

Gestures and Direct Manipulation for Immersive Virtual Reality

Thesis Advisor: George DRETTAKIS

prepared at INRIA Sophia Antipolis, REVES Team

defended on June 30th, 2014

Jury:

Reviewers :	Bernd FRÖHLICH	-	Bauhaus-Universität Weimar
	Pascal GUITTON	-	University of Bordeaux - INRIA
Advisor :	George DRETTAKIS	-	REVES / INRIA Sophia Antipolis
President :	Sabine COQUILLART	-	PRIMA / INRIA Rhône-Alpes
Examiners :	Maud MARCHAL	-	HYBRID / INRIA Rennes / IRISA
	Martin HACHET	-	POTIOC / INRIA Bordeaux
	Sabine COQUILLART	-	PRIMA / INRIA Rhône-Alpes

Acknowledgments

First of all, I would like to thank my advisor, George Drettakis, who made this thesis possible, and who has always provided me with sound advice and many encouragements in moments of doubts. I also wish to thank him for his comprehensive guidance and the project opportunities he gave me. Then, I would like to thank the members of my jury, my reviewers Bernd Fröhlich and Pascal Guitton, who gave me detailed and insightful feedback which greatly improved this thesis, and my examiners Maud Marchal, Martin Hachet, and Sabine Coquillart, for their valuable comments, and all of them for their support.

I am also very grateful to all the bright people I have had the opportunity to work with: it has been a pleasure and honor to be part of one of Adrien Bousseau's projects and I also thank him for the great conversations we shared; I also enjoyed working with Maria Roussou, Evanthia Dimara, Jean-Christophe Lombardo, Philippe Robert, Pierre-David Petit, Ravi Ramamoorthi and Maneesh Agrawala; and I am particularly grateful to Maud Marchal, Theophanis Tsandilas, Wendy Mackay and Lora Oehlberg for their considerable help with statistics, their very insightful comments and advice, and their availability and encouragements. This work has greatly benefited from their help.

Then, I would like to thank all the wonderful people I have met during my stay at Inria Sophia Antipolis: my friends, my patient and enthusiast pilots and participants whom I have "tortured" during lengthy experimental sessions, and my wondrous team. I wish to thank Pierre-Yves for his unfailing good mood and communicative enthusiasm, Rodrigo for his kind words, for rebooting my computer many times during the writing of this thesis and for the mandarins, Rachid for helping me so often with photos and any kind of practical and organizational issue, Gaurav for his availability with every technical issue I encountered, Emmanuel and Stefan for their dedicated help as pilots, Christian for his encouragements and help with stereo-tests, Jorge for driving me back and forth during bus strikes, Clément for the stupid Internet links we shared, Sylvain for his instructive conversations, Adrien D. and Laurent for the very funny moments we had, and also Marcio, Ares, Peter, Charles and Carles... and all of them for their heartwarming kindness. I am also deeply grateful to my colleagues and dear friends in the Dream team (SED) for their availability and help with the immersive space hardware, but mostly for their kindness and the wonderful moments and laughters we shared: David Rey, Jean-Luc Szpyrka, Julien Wintz, Thibaud Kloczko, Erwan Demairy, Nicolas Chleq, Hannah Carbonnier, and Jérôme Esnault whom I bothered so many times with Git issues! A special thank goes to my dear friend Julia who has been very supportive, even when I was conspicuously silent; she probably does not realize how much she helped me! Thanks also to my course and bus buddy Nadia for her support and our passionate conversations.

I also wish to thank the people who have been an important support during my time here: first of all, Sophie Honnorat for her kindness and dedicated help with all the administrative work; then the support and technical services for their availability and their good will.

A very special thank goes to Pascal Masson, my former teacher and my friend, who gave me the opportunity to teach practical work during my thesis. It has been a tremendous and unforgettable experience, and working with him has been a great honor and pleasure.

Lastly but most importantly, I would like to deeply thank my family, Mom and brothers, who have always been loving and supportive during this whole curriculum, and I dedicate this work to them. I probably wouldn't have achieved such goals without their constant help and encouragements, and I know it has been a hard and painful time for them to stand by a stressed and bad-tempered daughter and sister for all those years; I am very grateful for their patience!

To conclude, I would like to thank the Regional Council of Provence Alpes-Côte d'Azur and the EU project VERVE (http://www.verveconsortium.eu) which supported this work.

Texture resources used in the virtual environments

http://archivetextures.net
http://www.freestockphotos.biz
http://www.flickr.com
http://www.defcon-x.de
http://www.public-domain-image.com
http://images.meredith.com
http://onlyhdwallpapers.com
http://www.merveilles-russie.com
http://www.zastavki.com

Contents

CHAPTER 1

Introduction

For more than fifty years, virtual reality has gained more and more popularity, and continues seeing its hardware progress and its application possibilities diversify. It has become the crossroads of many fields and now is beginning to see widespread usage. It provides realistic experiences to users through the configurable combination of high quality stereo display, surround sound, tracking system and force feedback machinery. This quest for realism leads application designers to opt more and more often for gesture-based interfaces due to their well-acknowledged naturalness, and for fully immersive setups as they are known to provide a higher sense of presence within the virtual environment. In this thesis, we are thus interested by the challenging context of the use of gestures in fully immersive spaces.

Since its early days, virtual reality has been extensively used for training in military or flight simulators, and it has been proven that natural interfaces, because they are inherently intuitive, can help transfer virtual training to the real world [Bowman 2012]. Hardware advancements also made it possible for new applications such as industrial design. Virtual reality testing allows the reduction of the number of design iterations and thus prototyping costs and time-to-market.

But nowadays, we are witnessing a boom in virtual reality. A first reason for this is its widespread availability to the general public through the highly popular field of video games. Important technological progress has made new interfaces available to the masses. First, in 2010, the Kinect[1] allowed users to control video game inputs with their own body. The device, initially designed for large interaction spaces, now has a second version designed for PC users since 2012, further extending its usage. Then, in 2012, the Leap Motion controller[2] allowed users to interact with desktop applications directly with their bare hands. Another innovation concerns the display hardware. The very recent Oculus Rift[3], a head mounted display (HMD), should be commercially available in 2015 and aims at being affordable to a majority of gamers; it is already supported by several video games.

In what follows, we present several areas which motivated the research in this thesis.

[1]http://www.xbox.com/fr-FR/Kinect
[2]https://www.leapmotion.com/
[3]http://www.oculusvr.com/rift/

1.1 Virtual reality and rehabilitation

A major reason for the boom of virtual reality is that it is considered by many as the future of clinical rehabilitation in general, and of brain damage treatment in particular [McGee 2000, Rose 2005, Weiss 2006], becoming an integral part of cognitive, motor and functional assessment and rehabilitation. Studies have proven virtual reality to be highly effective for active learning, to provide powerful, safe and ecologically valid environments, to be highly adaptable to the patient's needs, to considerably motivate patients, to record objective measures of performance, and to provide alternate modes of feedback [Weiss 2006]. Such flexibility, immersion capacity and interaction control make it highly suitable for rehabilitation. Indeed, virtual reality has proven effective with exposure therapy and is becoming increasingly used in combination with traditional techniques for the treatment of a wide range of psychological disorders such as phobias and other neurological disorders such as post traumatic stress disorder, spatial neglect or autism spectrum disorders. More recently, these technologies have been used extensively for memory rehabilitation, notably in the context of Alzheimer's disease treatment where virtual reality training has a significant impact on age-related disorders. Virtual reality has been used to assess prospective, episodic and visuo-spatial memories, and presents the advantage of being replicable across laboratories and patients. In general, virtual reality gets high acceptance and tolerance from both healthy and impaired subjects. However, the use of virtual reality with elderly people requires specific considerations in the design of the system since they are more prone to fatigue and arthritis limits their movement abilities; tolerance has to be assessed.

1.2 Presence and immersion

Such a positive response to virtual reality treatment is justified by the sense of presence patients experience when facing virtual environments. [Slater 2009a] presents this notion as the combined effect of *place illusion*, which is the sensation of being in a real place, and *plausibility illusion*, which is the illusion that the situation and events are actually happening. However, such a sensation is highly dependent on the technical limitations of the virtual reality system. Several studies report the limitations of the use of virtual reality. In the context of clinical assessment, the cost of such systems make it difficult to afford for individual clinicians. Moreover, despite the realism and control provided, the involvement of clinicians in the process remains crucial, as the design and results of experiments suffer from biases and interpretation.

The key element behind all the proposed solutions is technological progress, as the compromise between limitations and needs defines the approach used. Since the seminal work by [Sutherland 1968] who designed the first HMD, display devices have evolved and diversified a lot, and so have interaction devices, allowing the creation of richer and more complex systems.

1.3 Displays

Concerning displays, immersion has been a major point of interest. While a monitor may be enough for simple operations, most applications will use stereo displays and combine manipulation and viewing spaces through this perception of depth. Head tracking is also a critical feature, as it allows the users to see from their own point of view. Early solutions are tabletop systems, such as the Responsive Workbench [Krüger 1994], and wall-sized displays. These are projection-based displays combined with embedded sound and tracking system. However, such setups only provide a restricted immersion as the user obviously sees the surrounding real world. To overcome this issue, CAVE-like systems [Cruz-Neira 1993] combined several of these wall displays to build up a large high level interaction workspace. Another very early solution for full immersion is the use of head mounted displays [Sutherland 1968]. They allow large workspaces and prevent the user from being distracted by the real world. They have been extensively used for training and rehabilitation applications.

1.4 Interaction devices

In what concerns interaction devices, the recent combined progress of physics simulation and gesture-based input hardware lead to a significant interest in this type of interaction, both for virtual and augmented reality. Data gloves are the most common input device to get data on the user's hand posture and motion as they preserve human manual dexterity. Virtual object manipulation (such as pointing, grabbing, translating, rotating and releasing) has become more accurate and intuitive. Indeed, in the 1980s, the available hardware was not fast enough to support interactive application, as reported by [Sturman 1989]. Simpler approaches were thus adopted. Vocabulary-based interfaces defined a set of gestures, describing a language, for each manipulation operation. They were efficient, but a) the number of gestures had to be restricted due to the learning overload, and b) the gestures had to be well designed to take into account the hand functioning and fatigue. Another approach, relatively natural and powerful, consisted in combining depictive gestures with speech to create intuitive multimodal interactions [Bolt 1980, Koons 1994, Latoschik 1998]. However, the most intuitive and natural interaction model is direct manipulation, where users behave as they do in real life. Such approaches get a lot of interest since software and hardware progress make them more realizable, and they have been proven more suitable for active learning [Bowman 2012]. Until the development of widely available physics engines, designers opted for heuristic approaches, using logic operations to rule the interaction. Physics simulators have brought virtual reality to a whole new level, supporting natural feedback and thus intuitive interaction with the environment. It also led to more complex systems as the virtual representation of the hands has to be thoroughly handled to provide realistic external forces onto the virtual objects in the scene. Physics simulations allow for more realistic interaction, thus enforcing the plausibility illusion. As a result, it is highly recommended in applications requiring an important sense of presence such as rehabilitation. Finally, a deeper level of immersion can be reached through the use of tactile feedback. It has been reported by [Sturman 1989] that visual feedback alone cannot provide enough cues for the users to experience direct presence, and the visual quality cannot compensate for such a shortcoming. However, currently available haptic devices are restricted to relatively small workspaces and are not compatible with large fully immersive spaces. They can be used in combination with HMDs, but physical navigation remains restrained. An exception to this is the scalable SPIDAR [Buoguila 2001], which simulates force feedback within CAVE-like spaces. However, the presence of cables hinders free navigation within the workspace.

1.5 Usability evaluation

Providing new interaction solutions is good to urge new ideas and thus innovative interfaces on. However, evaluating the pros and cons of those solutions is also important to derive directions for future research. Such evaluations assess the user's experience in terms of immersion and usability, and the effectiveness of the interface in terms of design and reachable performance. They combine objective and subjective measurements to provide rich information about the interface legitimacy. While natural interactions provide better spatial understanding and precision, are more intuitive and ensure the transfer of training to the real world [Bowman 2012], finger tracking in fully immersive spaces is challenging, due to the necessity of calibrating the devices and the occurrence of noise in the signal because of occlusions of the markers by the user's body or by the walls. On the contrary, flystick devices are well-established (designed to avoid occlusions, reliable signal), and such hyper-natural interfaces extend human capacities and thus interaction, easing users' actions [Bowman 2012].

1.6 Overview

In this thesis, our goal is to evaluate gesture-based interaction in the context of fully immersive virtual reality spaces. Our intuition is that the users' performance and experience are directly linked to the movements required by the tasks, due to the functioning of the human limbs and the interaction tools. More specifically, we will investigate how natural gestures compare to flystick-based interaction for direct manipulation. We thus propose experimental designs that involve the main everyday movements (Chapter 4), and develop devices to assess the influence of movements on tasks by decomposing them in a controlled setting (Chapter 5). In both cases, we compare virtual settings to a real world equivalent, which provides insight into issues concerning experimental and interface design. We also put a simple gesture-based interface to the test of its use with elderly people in the context of reminiscence therapy (Chapter 6).

We thus consider the following to be contributions of this thesis:

- **Evaluation of direct manipulation using finger tracking for complex tasks in an immersive cube:** We propose a solution to manipulate 3D virtual objects in a close-to-natural manner for general purpose tasks within a cubic immersive space. Our solution couples finger tracking with a real-time physics engine, combined with a heuristic approach for manual interaction, which is robust to tracker noise and simulation instabilities. We evaluate our interface through relatively complex manipulations, such as balancing objects while walking in the cube. Performance using finger tracking was compared to performance with a six degree of freedom flystick, and the analyzed tasks were also carried out in the real world. We also required our participants to perform a free task to observe their perceived level of presence in the scene. To our knowledge, this is the first time that such a system is evaluated in the context of general purpose tasks within fully immersive virtual spaces.
- **Controlled evaluation of direct manipulation through the decomposition of motion:** We propose a framework for the careful evaluation of direct manipulation of 3D virtual objects within fully immersive spaces. We specifically address the impact of motion type on the overall performance. To do so, we design devices that allow the decomposition of motion into movement nature (rotation and translation), and further into single and multiple degrees of freedom. We evaluate direct manipulation through pointing and orienting tasks, and compare to the performance obtained with a six degree of freedom flystick. The devices designed for decomposing the movement were also replicated in the real world. To our knowledge, this is the first time that the apprehension of decomposed movement for direct manipulation is evaluated within fully immersive spaces.
- **A feasibility study about the use of virtual reality for reminiscence therapy:** We present a novel virtual reality solution for reminiscence therapy, which is a popular intervention in dementia care. Our immersive system allows easy presentation of highly realistic familiar environments using the innovative image-based rendering technology by [Chaurasia 2014], and provides natural interaction with virtual objects within those environments. We evaluate the effectiveness of our system through a user study. Healthy elderly participants underwent a verbal autobiographical fluency protocol adapted for virtual reality in which they had to generate memories based on images they were shown. The system was tested for an unknown and a familiar environment to assess if it can convey familiarity of a given scene. To our knowledge, this is the first time that the use of virtual reality with image-based rendering is evaluated for reminiscence therapy.

The rest of this thesis is structured as follows:

- In Chapter 2, we discuss previous work in virtual reality in terms of interface technologies, evaluation and applications.

- Chapter 3 presents the hardware and software framework we use in our experiments, as well as the challenges presented by gesture-based interaction in the specific context of fully immersive spaces.
- Chapter 4 evaluates a direct manipulation interface for relatively complex general purpose tasks. The interface combines a heuristic approach to physics simulation.
- Chapter 5 assesses the effect of movement in 3D manipulation tasks through the decomposition of motion.
- Chapter 6 combines highly realistic rendering and natural interaction to evaluate the usability of virtual reality in reminiscence therapy for elderly adults.
- In Chapter 7, we summarize the results of this thesis and propose directions for future work.

CHAPTER 2

Previous work

Virtual reality (VR) is an active and increasing area of research as its applications spread across fields, from automotive design [Jacobs 2012] to military training or clinical rehabilitation [Legault 2013]. From the first head mounted display by [Sutherland 1968], virtual reality is now deployed in a diverse set of workspaces and interfaces. Its complex realization involves the combination of many fields such as electrical engineering to build ever more immersive hardware, highly realistic computer graphics and sound, physics simulation for intuitive interaction, or mechanical engineering to provide force feedback.

In his early work, [Sutherland 1968] designed a device allowing users to see a 3D scene from their own point of view, with the displayed images being updated when they moved their head. In this project which focused on display, head motion was the essential interaction with the environment. However, interaction has been enriched with the development of input devices; gesture-based interaction has received significant interest in virtual or augmented reality research [Sturman 1989, O'Hagan 2002, Buchmann 2004, Cabral 2005]. Over the last ten years, both physics simulation solutions and gesture-based input hardware have progressed immensely, providing the ability for much more accurate simulation and intuitive interaction.

In this thesis, our goal is to evaluate gesture-based interaction compared to flystick-based interaction in the context of direct manipulation within fully immersive virtual reality spaces. Specifically, we try to identify and explain the cases when one interface is better suited than the other to complete a task. We evaluate usability through criteria such as performance in terms of speed and accuracy, and presence in the virtual environments. As we show in the later sections, fully immersive spaces are technically quite challenging, and the study of usability of the interfaces in this context has not received much attention. Our intuition is that the users' performance and experience are directly linked to the movements required by the tasks, since the movements are dependent on the functioning and limitations of the human limbs and the interaction tools. We thus propose experimental designs that involve the main everyday movements such as grasping, releasing, translating and rotating, and develop devices to assess the influence of movements by decomposing them into individual and multiple degrees of freedom. We also put a simple gesture-based interface to the test of its use with elderly people, as virtual reality gains popularity in clinical treatments such as rehabilitation.

We are thus interested in the nature/technical aspect of the interfaces, their evaluation, and their possible applications.

2.1 Interfaces in virtual reality

2.1.1 Interface classification

In their thorough review of natural gestures for virtual reality, [Bowman 2012] classify user interfaces into three main categories: *traditional 2D interfaces* such as classic mice and keyboards, *natural interfaces* where the user behaves as in daily life, and *hyper-natural interfaces* which extend the human capacities or provide guidance. The two latter are grouped under the term *3D user interfaces* as they benefit from spatial input. *Hyper-natural interfaces* enhance interaction by making it more powerful, using a virtual or physical tool. For example, users are allowed to walk while staying in place or to point and select objects from afar. On the contrary, *natural interfaces* provide interaction as close as possible to the real world through the use of the least invasive input devices possible. The authors underline the many positive features of *natural*, but also discuss the utility of *hyper-natural* interfaces. Indeed, this type of interaction may be more convenient than *natural interfaces* as it eases the actions, e.g. by removing the traveling action that users also tend to avoid in real life through the use of remote controls, thus improving performance on task completion. On the other hand, *natural interfaces* are mentioned as one of the important future directions for 3D user interfaces as they provide better spatial understanding when traveling, and better precision due to the scaling of movement. They also come easily for novice users as they are inherently intuitive, and ensure that training will transfer to the real world.

We are interested in 3D user interfaces, but we will group them into *tool-based* and *gesture-based* interfaces. *Tool-based* interfaces rely on physical input devices which do not reproduce daily life manual interaction such as flysticks (or "wands"); they are necessarily *hyper-natural* interfaces. However, they can simulate daily-life tool-based interaction, such as the use of a fork. *Gesture-based* interfaces may be *natural interfaces* in the case of direct manipulation, or *hyper-natural interfaces* when gestures are used to implement virtual tools. *Tangible interfaces* lie at the intersection between tool-based and gesture-based interfaces; the simulation of the use of real tools also increases the sense of presence and naturalness. However, we won't study this type of interaction in this thesis.

2.1.1.1 Tool-based interfaces

Many 3D input devices have been designed to overcome the limitations of traditional 2D devices for 3D manipulations. In his thesis, [Zhai 1995] studies 6 degree of freedom input devices in terms of human performance.

One of the first mouse-derived hardware solutions presented is the isometric Spaceball^{TM}, which provides a ball featuring pressure sensors between the sphere itself and the supporting surface that detect the direction and magnitude of the applied force, allowing for 6 degree of freedom interaction. Elastic devices were also developed, such as the SpaceMaster^{TM} or the SpaceMouse^{TM}, which, however, have a restricted range of movement. Similarly, [Zhai 1993] designed a prototype of an elastic 6 degree of freedom input device, the EGG (Elastic General-purpose Grip), and compared its use with the isometric Spaceball^{TM}. This type of device is useful for desktop applications, but is not usable in the context of immersive spaces as it needs to be set on a desk.

For such cases, glove-like and wand-like input devices are more appropriate as they allow the user to move in space, at least within a certain range. However, only glove-like devices are designed for gesture-based interfaces. Hand-held devices simply combine a tracking sensor with a handle, and feature buttons and/or a joystick. This is the case of the Bat[1], of the Cricket^{TM}, and of the ART wand[2] we're using (see Sec. 3.1). This kind of interface is extensively used in fully immersive applications since it eases interaction (e.g., joystick-based interfaces are reported as being the most precise for 3D steering tasks because of rate-control) [Bowman 2012].

2.1.1.2 Gesture-based interfaces

Vocabulary-based interfaces In many cases, gestures are used to define a language, even if the number of gestures is often limited. Specific gestures and hand motions are used to point, grab, translate, rotate and release 3D virtual objects, to change the camera zoom and position or to navigate in the virtual world [Sturman 1989, O'Hagan 2002, Buchmann 2004]. In some cases, they are also used for menu item selection [Sturman 1989] and button triggering [Buchmann 2004]. However, this kind of interface can create an overload for the user who must remember the meaning of each gesture.

[Sturman 1989] describe a hand motion taxonomy which takes advantage of the richness of expressions of a whole-hand glove. They map hand and finger pose and motion to specific actions through lookup tables. However, they report that the tolerance range needed for recognition robustness (e.g. for finger flex angles) limits the number of unique postures.

On the contrary, [O'Hagan 2002] recognize a reduced number of hand poses and movement sequences with statistical classifiers through a computer vision-based interface.They use the action context to reduce the number of distinct gestures, and thus the memory overload. They outline that the hand functioning must be clearly understood and the gesture set carefully chosen to reduce the user's fatigue when manipulating.

In their urban planning system, [Buchmann 2004] simplify the interface by only using fingertips for interaction through image processing software, and glove-based finger and hand markers. To modify city blocks, draw streets or switch between immersive and map

[1] designed by C. Ware in 1990

[2] http://www.ar-tracking.com/products/interaction-devices/flystick2/

views, they implement simple gesture recognition through measurements of the fingertip position, relatively to the objects, to the scene ground plane, or to each other.

In our work, we prefer to avoid the use of a vocabulary, and concentrate on simple, intuitive gestures.

Multi-modal interactions An early solution providing relatively natural and powerful interfaces involved the use of multimodal interactions, combining depictive gestures with speech.

In his early work, [Bolt 1980] develops the “Put-That-There” system with which the user can place simple 2D shapes using combined speech and gestures. It only uses a pointing gesture to indicate the desired position of the object, and the rest of the manipulation (creating, moving, resizing, deleting or even naming shapes) is dialog-driven. The author observes that pointing allows the reduction of speech complexity using pronouns, whereas speech improves gesture precision when referencing.

[Koons 1994] propose a prototype interface named “Iconic” where users set the layout of 3D scenes, placing and orienting objects. Their approach simultaneously interprets information from speech and depictive gestures to build the final command. As the latter do not have a defined form, spatial relations are used to map the clues together and apply the final transformation to the object. Speech and gestures are inseparable to build the instructions.

On the contrary, [Latoschik 1998] propose an interface where speech supports a gesture-based interface to provide more precision and to strengthen some hypotheses. Their system implements a virtual construction scenario where assembly constraints and system knowledge are exploited to ease the interpretation. They use deictic gestures such as pointing for selection of objects and locations, and mimetic gestures for manipulation. They also use a keyword spotting approach to make the analysis of the speech input faster.

Direct manipulation In this thesis, we are more interested in the case of direct manipulation, i.e., users interacting with the environment in a natural manner with their hands. Such approaches attract more interest as the hardware progresses, making these interfaces more realizable. They are preferred because they provide a simulation which is closer to a real experiment, and thus are better suited for applications such as training [Bowman 2012].

In [Moehring 2010] and [Moehring 2011], the authors develop a glove-based interface for functional aspect validation in automotive design. Users naturally interact with constrained objects with their phalanges. They can adjust an interior car mirror, press buttons and turn knobs in fully immersive settings. The system provides a robust and realistic heuristic-based interaction. The authors report that virtual reality allows the evaluation of generated models earlier in the design process, thus considerably reducing the number of design iterations and required hardware mockups.

[Hilliges 2012] present an interactive system using a see-through display to provide the illusion of direct manipulation. Users interact within a "manipulation space" between the desktop and a half-silvered glass display. Their interface allows general purpose direct manipulation of virtual objects with bare hands, or with physical props. It also handles occlusions of virtual and real objects, providing a seamless mix. Hands are tracked with a Kinect^{TM}, and interaction is physics-based, allowing for a highly realistic interaction.

2.1.2 Realism of the simulation

Over the last ten years both physics simulation solutions and input hardware for gesture-based interaction have progressed immensely, providing the ability for much more accurate simulation.

2.1.2.1 Heuristic approaches

Initial work, before physics simulation became widely available, focused on the definition of appropriate heuristics for certain kinds of operations.

In their early work, [Cutler 1997] rely on the two-hand human manipulation framework described by [Guiard 1987] to implement one- and two-handed virtual tools for 3D object manipulation. The authors define some heuristics to control the transitions between those tools, the transitions being more or less implicit to the user. Each 3D manipulation, such as grabbing, scaling, translating and rotating with or without constraints, is represented by a tool and thus by a specific gesture. The interface also offers visualization tools such as cutting planes or opacity adjustment.

[Ullmann 2000] focus on a heuristic approach for grabbing virtual objects with data gloves to provide realistic and intuitive grasping gestures. Their algorithm relies on object geometry at contact points to determine if a grip is possible. They define grasping conditions on angle and length between contacts for one-handed grasp. For the two types of two-handed grasps, they identify wrapping around the object, and combining two one-handed grasps. The object transformation follows the transformation of a frame set at the *contact midpoint* (middle point between contacts). To compensate for the lack of touch and force feedback, the authors focus on visual feedback, providing position correction of the hand representation during collisions and when the hands move apart while the grip should be maintained.

In their validation system for car interior design, [Moehring 2010] provide intuitive, plausible and robust grasping heuristics, introducing the notion of "Normal Proxies" to improve grasp detection and stability. They extend objects with appropriate normals to create a pseudo-physical metaphor where physically correct friction is approximated by a geometrical representation of contacts called friction cones. Their system implements various types of constrained objects, involving mounting, functional and location constraints.

Objects are decomposed into parts that can be grasped, and others that can be pushed. Grasping is triggered by collisions between the user's phalanges and the virtual objects.

In [Moehring 2011], the authors report that using rigid body simulations is the preferred solution to implement realistic hand-object interaction in car interiors, but that current physics engines are not robust enough to correctly handle complex geometries and constraints.

2.1.2.2 Haptic feedback

[Sturman 1989] report that visual feedback alone cannot provide sufficient cues for gesture-based interaction, preventing the users to experience direct presence in the virtual environments. Similarly, in their system, [Ullmann 2000] focus on the visual feedback quality to compensate as much as possible for the lack of haptic (force and touch) feedback.

In this perspective, [Hirota 2003] conceived an algorithm for real-time haptic rendering with stable motion to simulate quite complex dexterous object manipulation. The user's hand and fingers are represented by a set of interface points, and interaction force is computed for each of those. Friction can also be simulated. They observed from their evaluation study that force and tactile feedbacks are required for dexterous manipulations to reduce colliding volume and object slipping.

[Ortega 2007] propose a generalization to 6 degrees of freedom of the "God object" method (also called "kinematic objects", they can apply forces to the other dynamic objects being simulated, but are not affected by other objects) for haptic interaction between rigid bodies. They provide high quality haptic interaction through continuous collision detection and constraint-based quasi-statics. Their method prevents interpenetration between objects and allows precise contact and sliding over surfaces. They report that haptics improve interaction because it provides the ability to feel the detailed geometry of the virtual objects.

Haptic devices are mechanical robots involving motors, such as the Phantom [Massie 1994], the Pantograph [Ramstein 1994] or the HapticMaster [Van der Linde 2002]. Thus, even if they greatly improve the interaction with the virtual environment, they are restricted to relatively small workspaces, such as desktop applications. This lack of embeddedness limits full immersion to the use of head mounted displays.

Integrating haptics in multi-screen immersive spaces is hard and requires specialized expertise. As a result, we do not use haptics in this thesis.

2.1.2.3 Physics engine integration

The addition of physics simulation to direct manipulation provides natural feedback in the environment and thus truly intuitive interaction with the objects in the scene. Such approaches for interface design have received much interest in recent years, often linked with tabletop systems. A major difficulty is how to handle objects controlled by the users hands (often called “God objects”; see “kinematic objects” defined above and in Sec. 3.3.2) with respect to the simulation of the rest of the environment, i.e., correctly providing external forces from the hands.

[Borst 2005b] use a spring model coupled with a commercially-available physics engine to simulate grasping and manipulation of objects of varying complexity; they avoid hand-object interpenetrations and provide force computation for haptic feedback. Grasping and interpenetration were also the focus of [Prachyabrued 2012]. [Agarawala 2006] propose a new and more physically realistic organizational structure for virtual desktops based on piling instead of filing. [Wilson 2008] combine sensing surfaces and advanced game physics engines to provide highly realistic interaction with virtual objects.

One remarkable early result is that of [Fröhlich 2000], which demonstrates the use of a fast physics solver and hand-based interaction, in the context of a workbench environment, for complex assembly tasks and multiple users and hands. The physics solver presented was one of the first providing sufficiently fast simulation to allow realistic interaction.

More recent work has concentrated on developing appropriate soft models of fingers [Jacobs 2011] and efficient solvers to avoid interpenetration of God objects and other objects in the scene [Jacobs 2012], mainly in the context of automotive project review. [Hilliges 2012] use physics particles and depth-aware optical flow to approximate 3D shapes of interacting rigid and non-rigid physical objects.

A simpler approach was proposed by [Holz 2008], where flexible grasping is simulated without complex physics, though allowing manipulation of multiple objects with multiple fingers. We follow a similar approach of simplified physics simulation.

2.1.3 Hardware setup

Since the seminal work by [Sutherland 1968] introducing the first head-mounted display, workspaces have diversified a lot, and so have interface devices, providing an important variety of virtual experiences.

2.1.3.1 Workspace

Sometimes, desktop-sized workspaces are sufficient for the purpose of the desired experiment, albeit reducing the interaction space and immersive characteristic. One may directly use the monitor as a display for relatively simple manipulation tasks [Sturman 1989], when others, more recent, build systems combining the viewing and interaction spaces, such as the HoloDesk [Hilliges 2012] which uses a see-through stereo display to yield a seamless mix between real and virtual for direct manipulation, creating a manipulation space between the desktop and the half-silvered screen.

Larger setups include tabletop systems, such as the Responsive Workbench [Krüger 1994] and the ImmersaDesk [Czernuszenko 1997]. The Responsive Workbench is a powerful tabletop projection-based stereo display featuring head tracking, and allowing for a large interaction space. The ImmersaDesk is a similar setup, only the rear-projected screen shows a 45 degree angle. Sound is also embedded.

Many applications requiring full immersion benefit from the use of head mounted displays (HMD), helmets providing screens in front of the user's eyes. Depending on the rest of the setup, applications using HMD allow the users to move physically in a larger workspace than, e.g., desktop applications. HMD are typically used for military training. They also have been used extensively in rehabilitation systems [Pair 2006, Bruce 2009, Corbett-Davies 2013, Bekele 2013, Banville 2012]. The Oculus Rift[3] is a very recent example of HMD; its low price promises to significantly spread the use of HMDs, making virtual reality setups affordable to small companies and individual clinicians.

Full immersion with high level of interaction (multi-modal feedback, multiple interaction metaphors, user tracking etc.) is also provided by CAVE-like systems. The original CAVE (CAVE Automatic Virtual Environment) [Cruz-Neira 1993] comprises three rear-projected stereo screens for walls and a projection from above for the floor, and also features a surround sound system as well as head and hand tracking through electromagnetic sensors. The DiVE is a recent 6-sided CAVE-like system[4].

2.1.3.2 Gesture tracking

In [Sturman 1989], the authors originally planned to implement a complete physical simulation of the hand to manipulate objects in the virtual world. However, at the time, the available hardware was not fast enough to support an interactive application. Since then, much progress has been made and an important variety of input devices for direct manipulation are available.

Gloves are the most common input device to get data about the user's hand position and motion as they preserve the human manual dexterity. [Sturman 1994] and [Dipietro 2008] provide detailed surveys of electronic glove-based input devices. They present the different tracking technologies: optical systems, magnetic, and acoustic tracking, and discuss the characteristics of the existing gloves as well as their applicability and applications.

Optical systems are based on vision and some do not require the user to wear any device, allowing completely free hand movement and poses. The Kinect[5] and the very recent Leap Motion Controller[6] use infrared-based tracking. The Kinect is designed for a

[3]http://www.oculusvr.com/rift/
[4]http://virtualreality.duke.edu/about/
[5]http://www.xbox.com/fr-FR/Kinect
[6]https://www.leapmotion.com/

relatively large interaction space in front of the device and tracks the skeleton of the visible user's body. The Leap Motion controller is a device designed for desktop applications which provides highly accurate tracking of both hands. On the contrary, ART[7] tracking system, also based on infrared signals, detects the position of markers that the user is wearing. This system is designed for large interaction spaces.

[Wang 2009] present an intermediate and innovative solution called "Color glove" which combines computer vision and a multi-colored plastic glove. Their system avoids the price and complexity of tracking gloves; their use is restricted to desktop applications. Through a carefully designed color pattern associated with a look-up table of natural hand poses, they infer the current hand pose in real-time from the rasterization of a single frame, providing efficient and reasonably accurate tracking.

Most of these solutions operate in a limited workspace and are thus unsuitable for our immersive setting.

2.2 Interfaces evaluation

Many new virtual reality interfaces are proposed to improve numerous aspects of our current society, be it by providing professional training to soldiers or pilots, by shortening industrial design and thus time-to-market in our ever demanding consumer society, or by providing clinicians with powerful assessment tests and treatment procedures to help impaired patients. However, when proposing a virtual reality system, its legitimacy has to be addressed, and this is usually done through system evaluation.

Two major aspects must be considered in such evaluations: the user's experience, and the effectiveness of the interface. The first relies on a) the user's sensation of immersion within the virtual world, as this is a key characteristic of virtual reality, and b) on usability, i.e., if users can efficiently achieve their goal while using the system. The latter concerns the attainable performance using the interface, and thus its design. While the first is more application-oriented (i.e., can we assess what we want to assess?), the second is more interface-driven (i.e., does the interface technically and physically allow the user to efficiently perform the task?).

2.2.1 Experience evaluation

2.2.1.1 Presence

In [Slater 2009a], Mel Slater justifies the realistic behavior of users facing an immersive virtual reality environment through the notion of presence, i.e., their sensation of immersion. He decomposes presence into two major components: *place illusion*, which is the

[7] http://www.ar-tracking.com

sensation of being in a real place, and *plausibility illusion*, which is the illusion that the situation and events are actually happening. These are *illusions* since the user *feels* that the experience is real, even if she *knows* it isn't. However, the author mentions that these illusions are conditioned by the virtual reality system, in terms of sensorimotor contingencies concerning the place illusion, and in terms of credibility of events to the user concerning the plausibility illusion.

In [Slater 1995], the authors hypothesize that the match between proprioceptive information and sensory feedback enhances presence in virtual environments. They study the influence of walking in place and present an interactive technique for moving within virtual environments in the direction of the user's gaze whenever a walking behavior is detected. Their study shows that subjective ratings of presence significantly increase with their walking metaphor. In [Slater 2009b] and the follow-up work [Yu 2012], the authors propose experiments to assess if visual realism in immersive virtual environments enhances presence. In their first experiment [Slater 2009b], they compare ray tracing, including shadows and reflections, to ray casting; both rendering techniques used texture mapping and the Phong lighting model. Participants have to face a pit room featuring a mirror for a few minutes before answering a questionnaire; the experiment is replicated twice, one per rendering technique; physiological measures, such as heart rate, are recorded. The use of ray tracing implied higher stress within subjects, confirming that higher realism improves presence. In their follow-up study [Yu 2012], disambiguate the effects of lighting quality from the dynamic effects of real-time shadow and reflections. Participants explore a virtual library and are instructed to read book titles; meanwhile, stressful events sequentially occur, such as books suddenly falling or a boy appearing and floating around the room. The experiment is replicated three times, once per rendering technique which are: a baseline condition with basic interpolated shading, the same basic shading with dynamic shadows and reflections, and global illumination. Results show that the quality of illumination did not impact place illusion, implying that the positive response in their previous study was due to the dynamic shadows and reflections. However, they observed that global illumination improved plausibility illusion.

In such experiments, presence can be evaluated through physiological responses. This is the case of the studies by [Meehan 2001], where the users face a pit room from which they can see a living room below. The author hypothesizes that an increased presence will evoke greater physiological responses, and that this level of presence is related to the realism of the virtual environment. He found in particular that heart rate proved to be the most sensitive measurement and correlated best to subjective presence. On the contrary, [Hoffman 2000] use virtual reality to distract burnt patients during physical therapy. The level of presence within the virtual environment is used as a distraction from other physiological feelings such as pain. They show that the magnitude of pain reported (no physiological measurement) from motion exercises was significantly lower when patients were distracted during treatment, making virtual reality a feasible non-pharmacological pain reduction technique.

However, in less anxiety-provoking experiments, presence is mainly evaluated through questionnaires. In the work presented in this thesis, we use the questionnaire[8] presented in [Schubert 2001]. It provides subjective rating scales of spatial presence, involvement, and experienced realism.

2.2.1.2 Usability

When evaluating interfaces, two complementary kinds of measurements can be recorded:

a) Objective measurements address the possibility for the users to achieve their goal by evaluating their performance, in terms of completion time, accuracy or number of errors. In their dexterous object manipulation project, [Hirota 2003] evaluate computation time of simulation cycle and sensor update loop interval, and thus interactivity.

b) Subjective measurements address the users' experience, evaluating their satisfaction, fatigue, perceived ease of use and cybersickness (sickness that may arise when using virtual reality technology). For example, when evaluating their urban planning system, [Buchmann 2004] reported user fatigue and frustration due to tracking problems. Similarly, [O'Hagan 2002] warn about the importance of the chosen gesture set on the users' fatigue when designing vocabulary-based interfaces.
   Cybersickness is a critical factor of feasibility. In [Viaud-Delmon 2000], the authors propose an evaluation questionnaire consisting of a 22-item scale assessing the immediate level of discomfort. It combines symptoms associated with autonomic arousal, vestibular and respiratory symptoms, as well as signs of somatisation. In the various experiments of this thesis, we use the questionnaire presented in this paper.

[Zhai 1998] propose a new measure to quantify coordination in multiple degrees of freedom based on movement efficiency, and apply it to the evaluation of two devices: a free-moving position-control device and a desktop elastic rate-controlled hand controller. They conclude that more direct devices may lead to fatigue, coarseness of the control action (in terms of coordinated motion) and anatomical limitations of the human limb whereas they take shorter time to learn; on the contrary, less-direct tool-like devices may take more time to learn but may be more efficient.

In their car design interface, [Moehring 2011] compared their virtual simulator to a real car interior, and validated their interaction metaphor through an expert review. The expert committee included ergonomists, simulation experts and automotive virtual reality system developers who interacted with all of the objects in the interface before interviews. The review concluded that the system provides intuitive and reliable assessment of object functionalities.

[8] http://www.igroup.org/pq/ipq/index.php

In [Nancel 2011], the authors study the influence of three key factors on task performance and on users' preferences, in the context of mid-air pan-and-zoom techniques for wall-sized displays. Those factors are uni- vs. bi-manual interaction, linear vs. circular movements, and level of guidance (restriction of movement degree of freedom). They proved that free space gestures are generally less efficient and more prone to fatigue than device-based interaction, that guidance increases accuracy, and that linear gestures are generally more efficient than circular ones. In addition to that, they also found that bimanual techniques perform very well, whereas unimanual techniques may be considered when tools must be held in one hand for specific actions.

We participated in [Cirio 2012] which proposes new navigation metaphors in Cave-like environments to encourage real walking while keeping the user safe from reaching boundaries. These techniques are evaluated through a user study with travel-to-target and path following tasks. Results showed that speed and accuracy were favored by traditional controller interfaces, while some new paradigms were better suited for physical walking.

2.2.2 Performance evaluation

2.2.2.1 Fitts's law

In [Fitts 1954], Paul Fitts ran pointing task experiments and derived a relation between speed, amplitude and tolerance in perceptual-motor activities also known as Fitts's law. The movement mean time (MT) is a linear function of the index of difficulty (ID) of the task:

$$MT = a + b.ID \tag{2.1}$$

The index of difficulty is defined by a relation linking the target distance D and the target width W:

$$ID = log_2(1 + \frac{D}{W}) \tag{2.2}$$

He also finds that performance is relatively constant over a central range of amplitude and accuracy conditions, which means that time and difficulty are proportional (Eq. 2.1), and falls off outside these limits (i.e., the linear equation is no longer valid). He proves the validity of this relation for a variety of linear pointing task experiments such as reciprocal tapping, disc transfer and pin transfer tasks.

[Guiard 1999] apply Fitts's model to navigation in WIMP (Windows, Icons, Menus, and Pointing) interfaces since they define these as a form of multi-scale pointing task. They show through a preliminary experiment that users can handle higher levels of task difficulty with two-scale rather than traditional one-scale pointing control. They also prove that for very high precision movements, styluses outperform mice.

[Stoelen 2010] and [Nguyen 2014] apply Fitts's model to combined rotational and translational movements. [Stoelen 2010] designed three experiments, respectively with pure 1D translations, pure 1D rotations, and a combination of these movements. They show that performance times for combined movements are, on average, equal to the sum of each component individual performance times, and they propose a Fitts's model equivalent for combined movements. They also found a strong degree of coordination of the rotational and translational components by the central nervous system generating parallel execution. [Nguyen 2014] study the influence of orientation in combined movements on multitouch screens using a Fitts's law setup. Participants had to combine left or right translations with clockwise or counterclockwise rotations to move objects to a target position using two fingers. They show that right-oriented tasks are performed faster and more easily, and that combinations of differently oriented movements cause more fatigue and resulted in more strategy switches.

2.2.2.2 Steering law

In [Accot 1997], the authors study the existence of robust regularities in trajectory-based tasks, leading them to identify the existence of a steering law. While Fitts' law rules pointing tasks, they report that human-computer interface device evaluation needs further testing, and especially for trajectory-based tasks. The authors demonstrated that the logarithmic relationship between movement time and tangential width of target in a tapping task also exists between movement time and normal width of the target in a "goal passing" task and confirmed that there is a simple linear relationship between movement time and the "tunnel" width in steering tasks. They also describe a generic approach defining a global law that applies to a variety of trajectories, such as classic, narrowing, or spiral tunnels and, more generally, to curves.

In [Accot 1999], the same authors further apply their model to the evaluation of input devices for trajectory-based tasks. They argue that most evaluation tasks consist in target acquisition whereas computer interfaces often involve trajectory-based tasks. While the former are accurately modeled by Fitts's law, steering law applies to the latter. Thus, steering law evaluation can complement Fitts's law. They test five computer input devices for a linear and a circular steering task. They could group devices with respect to their performance, and prove the applicability of the steering law for all tested devices.

While being an extension of Fitts's law, the steering law does not apply to our context of general purpose direct manipulation. Indeed, general purpose direct manipulation tasks focus on the final position and orientation of 3D virtual objects (i.e., target acquisition), and, most often, not on the trajectory itself.

2.3 Applications

In Chapter 6, we will study the feasibility of using virtual reality for reminiscence therapy. We thus review the clinical use of virtual reality for the assessment and rehabilitation of brain disorders.

2.3.1 General use

As one can see in Sec. 2.1, virtual reality is extensively used in automotive design [Moehring 2010, Moehring 2011, Jacobs 2011, Jacobs 2012]. It is also a privileged tool for training (flight, military or surgical simulators) as there is evidence that skill improvement will transfer to the real world [Bowman 2012]. More recent reviews of the field [Rose 2005, Weiss 2006] predict the dramatic expansion of the use of virtual reality, in particular for brain damage treatment and rehabilitation in general.

[Rizzo 2005] propose a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis of virtual reality rehabilitation and treatment. They report successful “proof of concept” systems and applications being effectively used in clinical practice, despite the fact that the use of such technology for rehabilitation is still recent. They encourage multidisciplinary collaboration for the development of applications, which must remain user-centered, and foresee a “significant positive impact on the rehabilitation sciences”.

[Weiss 2006] claim that virtual reality-based therapy presents attributes making these technologies highly suitable for the evaluation and intervention for cognitive, motor and functional rehabilitation. Indeed, studies have proven virtual reality to be highly effective for active learning, to provide safe and ecologically valid environments, to be highly adaptable to the patient’s needs, to considerably motivate patients, to record objective measures of performance, and to provide alternate modes of feedback. However, they object that the development of virtual reality rehabilitation is conditioned by technological and financial limitations.

[Rose 2005] focus their review on brain damage rehabilitation, describing studies using virtual reality for the assessment and rehabilitation of resulting disabilities, such as executive dysfunction, memory impairments, spatial ability impairments, attention deficits, and unilateral visual neglect. Some other studies address rehabilitation training strategies. The authors foresee that the use of virtual reality for brain damage rehabilitation will become an integral part of cognitive assessment and rehabilitation, due to its immersive and interactive benefits.

2.3.2 Rehabilitation

Early work in virtual reality for cognitive assessment and rehabilitation is described in [Riva 1997]. The authors underline the advantages of assessment tools based on virtual reality as the use of virtual environments provides powerful, highly immersive, flexible and programmable tools featuring numerous stimuli and a high level of interaction control. Clinicians can measure a wide variety of responses. Environments and interaction can be configured for the specific need of each patient. However, experiments and results are prone to theoretical biases and interpretation, so the involvement of diagnostic clinicians in the development and data analysis of the experiments remains crucial.

[Rizzo 1998] discuss the various theoretical and practical issues to be considered for the use of virtual reality for rehabilitation, underlying the importance of having graphics indistinguishable from the real world. They also determine the utility of virtual reality in this context as this technology allows for complex and dynamic stimuli, and for precise measurements within the virtual environment, providing a wide range of cognitive and functional scenarios.

[Pugnetti 1998] also address the utility of virtual reality-based tests which overcome limitations of traditional tests in probing incidental memory and executive functions. Virtual reality provides a considerable range of opportunities to monitor and measure behavior, and is at least as sensitive to target cognitive impairments as traditional paper-and-pencil tests. Virtual reality memory tests also present the advantage of being replicable accross laboratories and subject samples, and benefit from high acceptance and tolerance from healthy and impaired subjects.

[Weiss 2004] use basic gestures in the context of rehabilitation. They use a video-capture virtual reality platform where the users see themselves embedded within simulated environments displayed on a large screen. Their own natural movements, tracked without markers by a camera, are the input as they entirely control the interaction with the graphic elements, yielding a complete engagement of the users in the simulated task. The authors report video-capture virtual reality as a useful tool for rehabilitation intervention as it presents almost no side effects while improving presence, enthusiasm and satisfaction.

[McGee 2000] foresee that the advancements in the clinical options offered by virtual reality could enhance the study of human cognitive and functional processes, and improve the assessment and treatment of impairments found in persons with central nervous system dysfunction. More recently, [Bruce 2009] report that virtual reality exposure therapy is becoming an increasing commonplace technique for the treatment of a wide range of psychological disorders, as a supplement for conventional therapy techniques.

[Klinger 2013] develop a tool to provide therapists and patients with customized rehabilitation sessions. The AGATHE tool allows real-time monitoring of patients activity and creates adapted virtual interactive environments. To develop their prototype, they focused on patients suffering from acquired brain injury such as Traumatic Brain Injury and stroke. Therapists can test the patient's abilities before designing the virtual therapeutic scenarios and then supervising and analyzing the patient's activity. Pilots tests have been performed, and the efficacy of the prototype is being assessed among various populations of patients.

2.3.2.1 Phobias

Recently, there has been an increasing interest in investigating the use of virtual reality to treat phobias, where exposure therapy has proved effective.

[Bruce 2009] propose a prototype system for virtual reality exposure therapy to claustrophobia (i.e., the fear of small or closed spaces). Users are invited to explore an apartment comprising four interconnected rooms with an increasing number of claustrophobic cues (such as the room size decreasing, the light diminishing, the furniture being sparser and the windows getting removed). Once the user enters a room, the door locks behind her. The study presents a treatment scenario with non-patient participants who had to explore the virtual environment to report as much detail as possible. The experiment focuses on presence and thus effectiveness, as well as on affordability and robustness for actual clinical use. Presence is evaluated through the use of a HMD for half the participants, and a large screen for the other half.

[Taffou 2012] study the efficiency of auditory-visual environments in generating presence and emotion to investigate the efficacy of virtual reality exposure treatment for cynophobia (i.e., the fear of dogs). The experiment involved healthy participants sensitive to cynophobia, within a CAVE-like fully immersive space featuring spatialized surround sound to account for the important auditory aspect in this phobia. Users navigate within two virtual environments, a garden and a hangar, with increasing anxiety-producing clues, such as the progressive aggressiveness of dogs or changes in lighting conditions (lights switching off or appearance of fog). A behavioral assessment test, consisting of a simple encounter with a virtual dog, is performed before and after the exposure to the virtual environments. Results showed that manipulating auditory-visual integration could modulate affective reactions, making the system a promising tool for the treatment of cynophobia.

[Corbett-Davies 2013] present an interaction framework for augmented reality and demonstrate its applicability in an interactive exposure treatment system for arachnophobia (i.e., the fear of spiders). Their tabletop application tracks and models real world objects through the use of a Kinect for realistic interaction, allowing the use of almost any object in the scene. Users see the real scene (desktop, objects, and user's arms) and virtual spiders are displayed on top. The system allows interaction with the user's own body, and handles occlusions, giving the illusion that spiders are randomly walking up, around or behind real objects through a simple set of behaviors.

2.3.2.2 Neurological disorders

Virtual reality as a tool to study some neurological disorders has also been the subject of many studies in recent years.

[Pair 2006] propose a Post Traumatic Stress Disorder (PTSD) virtual reality therapy application for Iraq war soldiers. PTSD are caused by "traumatic events outside of the range of usual human experiences". They propose a fully immersive experience through the use of a HMD together with hardware providing auditory, olfactory and tactile stimuli, as well as vibrations. The prototype has been tested with non-PTSD soldiers who had returned from an Iraq tour of duty. It presents highly customizable scenarios (time, weather, lighting, company...) and proposes several environments with relevant context for exposure (e.g., a rural village, a desert base, or a desert road convoy). The setting determination depends on which environment most closely matches the patient's needs. Clinicians can also trigger auditory and visual triggers, such as weapons fire, explosions, movement of people and vehicle or wounded people and remains. [Yeh 2009] present a follow-up of the previous study with active duty participants who had engaged in previous PTSD traditional treatments without benefit. Their results showed significant reduction of PTSD after virtual reality exposure treatment.

[Tsirlin 2010] study unilateral spatial neglect, a post-stroke neurological disorder that results in failure to respond to stimuli presented contralaterally to the damaged hemisphere. Through their application, they evaluate the hypothesis that patients avoid the affected space since their sensorimotor experience is decorrelated in this space. They use a large scale visuo-haptic system to induce spatial biases in healthy participants. Writing is simulated without occlusion thanks to a stylus featuring a haptic device for perturbations and a virtual ray. In several sessions, users have to trace a curved line appearing gradually on a surface before it disappears. Perturbated sensorimotor stimulation is provided through variable haptic perturbations, randomized in magnitude, type, and duration. Their results support their hypothesis that spatial biases in unilateral spatial neglect result from avoidance of perturbed sensorimotor experience in the contralesional hemispace.

[Bekele 2013] address Autism Spectrum Disorders (ASD) in adolescent populations. They are characterized by atypical patterns of behaviors and impairments in social communication. One of the main issues is to recognize and respond to facial expressions. The authors developed a system for facial emotional expression presentation featuring eye-tracking and the monitoring of physiological signals related to emotion identification. Virtual characters narrated a story without expression, only showing the corresponding expression afterwards. A set of emotional expressions with different levels was provided. A usability study has been performed with adolescent with and without ASD, all wearing a HMD. Their system proved to be able to suggest emotions as some ASD patients even recognized expressions with greater accuracy than the control group.

These studies show that virtual reality proved effective for an important variety of neurological disorders, making it a recommendable tool for clinical rehabilitation in general.

2.3.2.3 Memory rehabilitation

Virtual reality has been used extensively for memory rehabilitation; the review in [Brooks 2003] demonstrates the positive effect of this technology. Recently, virtual reality has been used to assess prospective memory using a driving simulator with a large screen [Gonneaud 2012], or to detect disorders through virtual visits with a HMD [Banville 2012]. Similarly, age effect on episodic memory was studied in a driving simulator [Plancher 2010], while a later study was performed for patients with amnestic mild cognitive impairments [Plancher 2013]. A follow-up study [Plancher 2012] demonstrates that *interacting* with the virtual environment can enhance performance for visuo-spatial memories.

[Brooks 2003] explain that virtual reality training promotes procedural learning in patients with memory impairment as it provides comprehensive, ecologically valid and controlled evaluations of prospective, incidental and spatial memory better than with traditional tests. This learning has also been found to transfer to improved real world performance.

[Gonneaud 2012] use virtual reality to study prospective memory (i.e., the ability to remember to execute an intention at an appropriate time in the future). Young subjects explore virtual cities by car and have to recognize pictures of the elements present in the city. In a second immersion, they have to remember to perform actions on the way, intentions having been learned beforehand. The authors report that when the link between the prospective and retrospective components of prospective memory is lacking, the execution of delayed intentions is harder.

[Banville 2012] demonstrate the relevance of a virtual reality-based assessment test to detect posttraumatic prospective memory disorders as a complement to traditional assessment tasks. Users who sustained a traumatic brain injury as well as control participants had to visit two virtual apartments and perform three prospective tasks. They also had to perform two prospective tasks for the traditional assessment. Results show that impaired participants are less efficient, yielding a correct classification of 75 percent of the participants.

[Plancher 2010] investigate the effects of normal aging on the main aspects of episodic memory (i.e., what, where and when). Young adults and healthy older participants drive a car in a virtual city comprising specific areas. Half of the users have to memorize the environment (intentional encoding); all of the participants are asked to recall as many elements as possible. Results show that older participants had less spatio-temporal recollection than the younger with intentional encoding, but similar recollection with incidental encoding. In their follow-up study, [Plancher 2013] include healthy older adults, patients with amnestic mild cognitive impairment and patients with early to moderate Alzheimer's disease. Users explore a city as the car driver or as a passenger, and are instructed to encode all the elements of the environment as well as the associated spatio-temporal context. They report a higher correlation between memory complaints and performance with virtual reality than with traditional tests. Also, active exploration improves the patients' episodic memory performance.

[Plancher 2012] suggest that the effectiveness of learning depends on the degree of interaction with the environment and freedom in the planning of actions. They isolate these two factors to investigate their effect on factual and spatial memory. They use a driving simulator in which the user can be passive, have an itinerary planning role, or drive the car through a given itinerary. They report that both factors enhanced spatial memory, while interaction impairs factual memory.

2.3.3 Virtual reality and elderly people

The use of virtual reality with the elderly imposes specific considerations. These have been studied early on (e.g., [McGee 2000]), using virtual environments with stereo display and head/hand tracking. More recent results show that memory function can be improved with immersive audio-visual virtual reality [Optale 2010]. Most recently, the work in [Legault 2013] shows that training with virtual reality can significantly reduce age-related effects. Our approach in Chapt. 6 follows this line of work, in proposing novel tools to improve reminiscence therapy for the elderly.

[McGee 2000] study assessment and possible rehabilitation of visuospatial skills in healthy older adults. They use 3D virtual environments projected on an ImmersaDesk type display [Czernuszenko 1997] to develop visuospatial tasks assessing skills such as visual field-specific reaction time, depth perception, 3D field dependency, static and dynamic manual target tracking in 3D space, and spatial rotation. Through this study, they address the motivation, rationale, and relevant issues in using virtual reality with elderly people.

[Optale 2010] propose a system for virtual reality memory training. They try to reduce cognitive decline and improve memory functions through auditory stimulation and path finding experiences, and compare to more traditional face-to-face training sessions using music therapy. Participants were residents of a rest care facility and were impaired on the Verbal Story Recall Test; they underwent neuropsychological and functional evaluations. Results show that the new approach may improve memory functions by enhancing focused attention.

[Legault 2013] use a dynamic virtual environment to assess the capacity of elderly people to improve their tracking speed, a skill that is clearly affected by healthy aging. They propose 3D multiple object tracking tasks to younger and older participants in a fully immersive CAVE setup, and compare the speed thresholds of the observers. Their results show that perceptual-cognitive training can significantly reduce age-related effects in elderly adults as learning in healthy older persons is maintained for processing complex dynamic scenes.

2.4 Conclusion

In this chapter, we have reviewed virtual reality interfaces, in terms of hardware display and input devices, and in terms of software approaches. We have also reviewed methods of evaluating such interfaces, and the applicability of virtual reality for clinical neurological disorders.

Recently, gesture-based interaction has received more attention for the naturalness it provides to the experience, and for the dexterous manipulations it allows. Also, virtual reality appears to be a major future direction for clinical assessment and rehabilitation of brain impairments. We also notice that presence is a critical element for such treatments to be efficient. We thus choose to study gesture-based interaction in the challenging context of fully immersive spaces and propose the following contributions: in Chapter 4 we propose a gesture-based interaction system based on physics simulation and heuristics and use it for relatively complex general purpose tasks; we evaluate such an interface in Chapter 5 through the careful decomposition of movements; finally, we apply it to the context of reminiscence therapy in Chapter 6 with a feasibility study involving high quality rendering, manipulation and navigation within familiar and unknown environments.

CHAPTER 3

Apparatus

In this chapter, we present the hardware setup used in all of our experiments, as well as our software framework. The projects in this thesis are placed in the challenging context of fully immersive spaces and use interfaces such as traditional 6 degree of freedom flysticks and finger tracking gloves. This involves specific equipment and technical constraints. We thus also discuss the solutions adopted to overcome those limitations.

3.1 Hardware

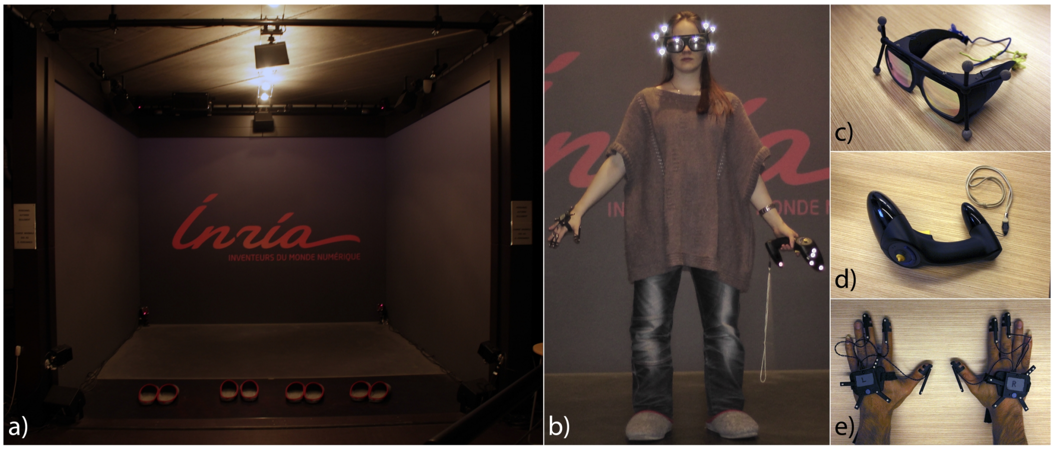


Figure 3.1: *Experimental setup. The four-sided immersive cube is shown in a) with eight infrared cameras and surround sound. In the middle, b) a user wears the tracked input devices which are: c) Infitec glasses with frame, d) ART flystick and e) ART finger tracking gloves.*

In this thesis, we use a four-sided immersive cube (the Barco iSpace[1], comprising three walls and a floor), which has four rear-projected "black" screens. The front and side screens are 3.2m wide x 2.4m high, and the floor is 3.2m wide x 2.4m long. All screens have a resolution of 1600 x 1200 pixels. Users have to wear slippers inside the cube to avoid damaging the bottom screen. Stereo is provided using Infitec technology[2].

[1] http://www.barco.com/en/products-solutions/visual-display-systems/3d-video-walls/multi-walled-stereoscopic-environment.aspx

[2] http://www.infitec.net/

For tracking, we use an ART[3] infrared optical system, with eight cameras (see Fig. 3.1 left). We track the head with a frame mounted on passive Infitec glasses and topped with small reflective spheres (passive device); the images displayed are then computed for the user's point of view as the eyes correspond to the virtual camera. The tracked devices provided by ART are the wand and the finger tracking system to track the palm, thumb, index and middle finger of each hand (see Fig. 3.1 right). The wand is also a passive device using the same kind of frame as for the glasses under a translucent casing. The finger tracking system tracks the tips of the fingers giving the relative positions of these frames with respect to a 6 degree of freedom tracker which is on the back of the hand. This device is active: infrared electroluminescent diodes emit signals from the fingertips and the back of the hand.

Normally, six cameras are sufficient for head tracking. However, to allow better quality finger tracking and reduce “shadow" regions (see Sec. 3.2.2), two additional cameras were added. As these cameras must be placed in the bottom front corners of the immersive cube and the cables connecting them to the system are visible in certain areas of the side walls, this can affect the user's immersive experience. We thus designed our scenarios to avoid eye gaze on these areas as much as possible, and we also designed our models to hide the cables and the cameras (for example, by adding very dark baseboards to the room in the environments of Chapter 4).

The screens are driven by five HP Z800 computers in a master-slave setup. The slaves, one per screen, are equipped with a Nvidia FX 5800 graphics card each. The master receives the tracking data, broadcasts this information to the slaves and synchronizes them. Each slave handles the rendering of its allocated screen.

Lastly, the cube is equipped with a 6.1 surround sound system.

3.1.1 Stereo-blindness test

A fraction of the human population has impaired stereo vision. This inability to see in 3D using stereo vision results in an inability to perceive stereoscopic depth, by combining and comparing images from the two eyes [Whitman 1970, Barry 2009]. Each participant is tested for stereo-blindness before an experiment actually starts. A random-dot stereogram is then displayed on the front screen. It consists of random dot images which, when viewed with stereo glasses, produce a sensation of depth, with objects appearing to be in front of or behind the display level [Julesz 1971]. After the participant confirms that depth is correctly perceived, she is considered as not stereo-blind. The experiment continues only if the participant is not stereo-blind.

[3] http://www.ar-tracking.com

3.2 Finger tracking

In each experiment, the user is presented with a representation of the palm, thumb, index and middle fingertips in the form of small colored shapes, providing visual feedback of hand position and orientation (see Fig. 3.2). In our first experiment (see Chapter 4), we chose to use cubes for computational efficiency; they were mostly hidden by physical fingers, and thus did not interfere with interaction. However, we did realize when designing our second experiment (see Chapter 5) that the comparison with wand and real conditions was unfair. Indeed, using cubes implies that the collision configuration changes with respect to the fingertip orientation, whereas this is not the case with handles representing wand selection (also called "wand selector handles") which use spheres. Moreover, the simulation would be closer to natural with spheres as real fingertips are closer to spheres than to cubes. In our last experiment (see Chapter 6), because our participants were elderly people, we chose to use cylinders so that the fingertip representation was more visible and they could more comfortably rely on it.

Figure 3.2: *Visual representation of the palm and three fingertips in the virtual environment (shapes are shown away from the fingers for clarity of illustration). Left: cubes from Chapter 4. Middle: spheres from Chapter 5. Right: cylinders from Chapter 6.*

3.2.1 Device calibration

The finger tracking system recognizes a user's hands by identifying a pattern specific to the user's morphology (finger length and radius). Hence, the finger tracking devices are calibrated for each user so that we have more reliable signals. Both devices are calibrated separately to avoid interference, using the procedure defined by the manufacturer. The calibration consists of two steps lasting a few seconds each: first, the user has to keep her hand flat and steady; second, she has to pinch in a regular movement with all her fingers staying straight. During the entire calibration phase, the user has to hold the back of her hand oriented towards the ceiling.

3.2.2 Signal filtering

Even though we use a high-end system, finger tracking in the immersive cube is challenging because the signal for the fingers is prone to noise (“trembling”) and interruptions (“jumps”) (see Fig. 3.3 (e)). Most losses of signal are due to visibility problems: a freely walking and moving user can often be in, or close to, “shadow” regions of trackers where the signal is deteriorated. This is due to occlusion from the user herself or when the device is close to the screens: the cameras are not able to see the markers and recover the signal. Tracking “shadow” regions occur in the zone 10-20 cm away from the screens. Hence, we designed our tasks so that the hands are always visible to a maximum of cameras (at least 3 or 4) by avoiding activities where the user’s body blocks the tracking cameras or involves objects in the “shadow” regions. Other signal artifacts are due to the reliability of the finger markers: while the palm is well identified and tracked (many markers), each finger has only two markers resulting in unstable captured data.

We conducted a preliminary noise study (see Fig. 3.3) where a user had to follow a spherical guide along a path. We measured the position of her hand and of the wand compared to the guide. We observed that the noise pattern seems to be similar for the wand and palm; however, we can see that hand-tracking presents additional noise for the fingers, whose order of magnitude is in centimeters. The user kept her hand steady while following the guide. As the position of the fingers are relative to that of the palm, we should theoretically obtain a constant. To account for the unsteadiness of the user, some variations obviously appear. But we can also observe jumps and plateau regions in the signal.

To overcome this problem, we apply a double filter. First, we track the variance of the signal over a sliding window. If variance is above a threshold we test if there is a plateau following in the signal, in which case we identify this as a loss of signal and do not apply the motion to the object. In the other case, the movement is considered too fast. Second, a Kalman filter is applied to each finger in time to smooth the signal (reducing “trembling”).

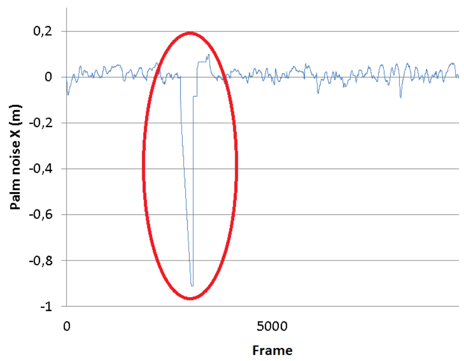
Filtering is applied on the master side when receiving the data from the tracking system so that the position and orientation of the palm and fingers are coherent between screens.

3.3 Software

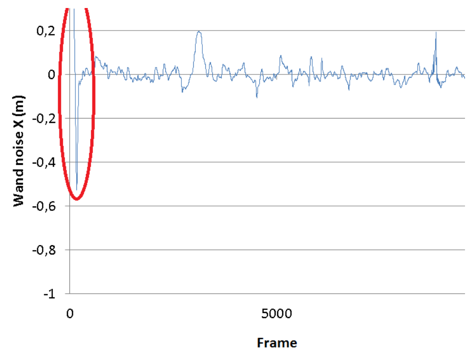
3.3.1 Framework

Our experiments are built on top of our in-house virtual reality software system, which is based on OpenSceneGraph[4]. As we chose *Bullet* for our physics simulation, we also use osgBullet[5] and osgWorks[6].

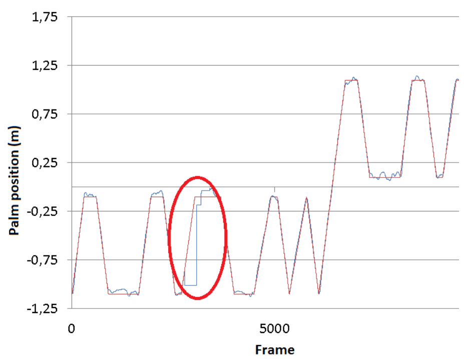
[4]http://www.openscenegraph.org/
[5]https://code.google.com/p/osgbullet/
[6]https://code.google.com/p/osgworks/



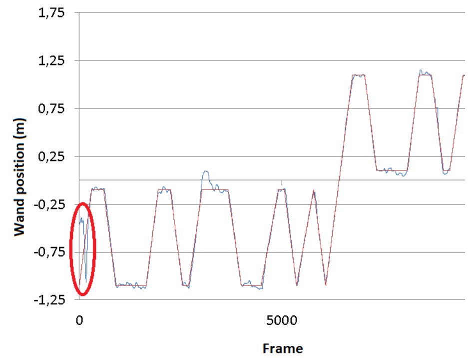
(a) *Difference between palm and guide positions, x axis.*



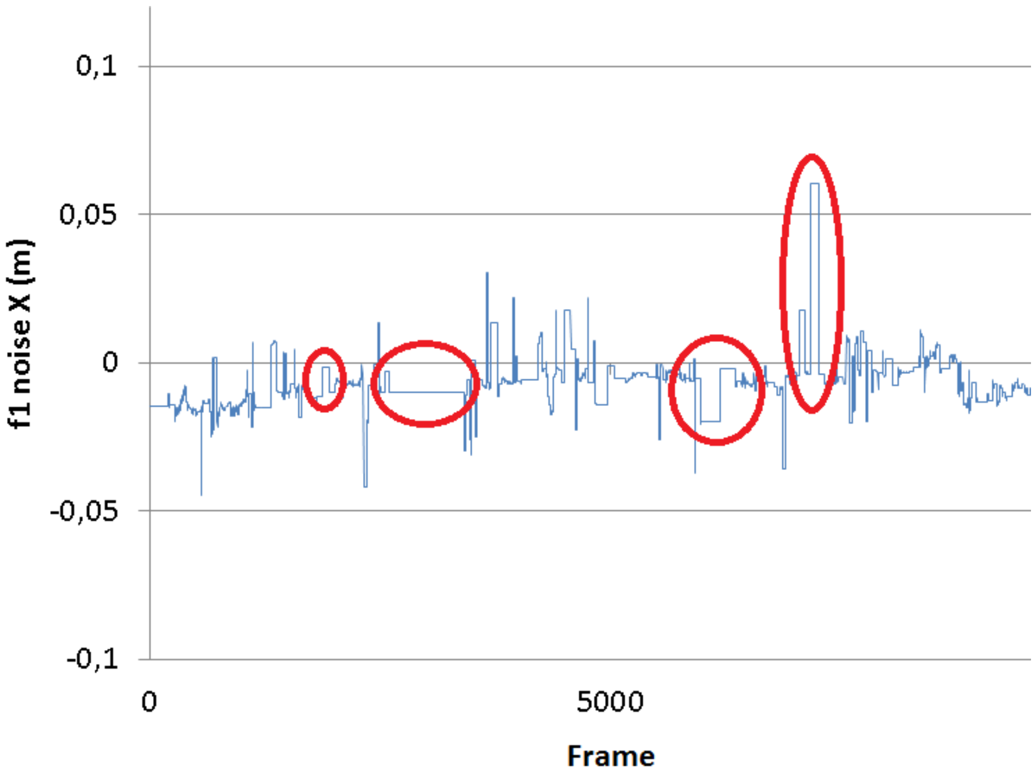
(b) *Difference between wand and guide positions, x axis.*



(c) *Palm and guide positions, x axis.*



(d) *Wand and guide positions, x axis.*



(e) *Thumb relative position, x axis.*

Figure 3.3: *Tracking noise. Red circles highlight examples of the tracking signal being blocked, resulting in steady positions and peaks in noise. Graphs c) and d) represent the superimposed signals of the palm or wand (blue) and guide (red). The lower graph shows that finger tracking has large additional noise with an order of magnitude of centimeters.*

Our in-house middleware gathers the tracking data and handles the synchronization between the master and slave machines. It provides a framework for distributed applications.

3.3.2 Bullet physics integration

In the different experiments we designed, our tasks required the use of a fast physics engine; we chose to use *Bullet*[7], which is fast enough to handle quite complex scenes in real-time.

As we work in real-time (involving physics computation approximations) and *Bullet* uses an impact-based simulation approach, objects tend to bounce between the fingers instead of being grasped, and the contacts are unstable. Many recent solutions improve physics simulation, most notably to avoid inter-penetration of the hand and other objects in the scene (e.g., [Jacobs 2012, Prachyabrued 2012]), while other approaches [Ullmann 2000, Holz 2008] avoid the need for complex – and often expensive – precise simulation using specific algorithms. In contrast, we use a fast but simple physics simulation, combining it with an approach for basic direct manipulation such as grasping and releasing, and a finite state machine to handle sequences of manipulations such as moving from one-hand to two-hand grasping (see Sec. 4.1 for details on the approach). More specifically, we adapted the *Bullet* open source physics engine to communicate the data needed by our heuristic approach through extra object properties, as we enforce the transformations of the handled objects coherently with those controlled by the physics simulation.

This implies that the engine constraint solver is applied only when no kinematic object is involved in the collision set of a given dynamic object. By definition, “kinematic objects” can apply forces to the other dynamic objects being simulated but are not affected by other objects. Fingers and wand handles are thus kinematic objects. Otherwise, we store the list of kinematic contacts (with extra data such as collision normals and kinematic objects transforms) on each dynamic object to derive the number of fingers applied onto it, as well as the trajectory it should follow and its physics properties (such as speed). We also store the contacts and collision normals with other static and dynamic objects to limit the enforced trajectories of handled objects coherently with the expected collisions. Due to the “trembling” signal mentioned in Sec. 3.2.2, objects could be dropped during the manipulation. Hence, we store the aforementioned contact lists for the previous and current frames, and check if kinematic contacts need to be maintained over time.

3.4 Conclusion

In this chapter, we have presented the hardware and software setup used in the following experiments. We have also discussed the solutions adopted to deal with the fully immersive and real-time constraints.

[7]http://www.bulletphysics.org

The limitations of complete immersion required precautions in the design of the experiments to avoid tracking “shadow” regions, as well as a specific double filter to overcome the diverse signal instabilities. We also modified the physics library so that the simulation did not suffer too much from the tracking limitations.

The real-time constraint implied a compromise between the speed and precision of the physics simulation, which led us to design a finite state machine for simplicity and efficiency.

CHAPTER 4

Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube

In this chapter, we will study the tradeoffs between direct manipulation with physics and flysticks in the context of large immersive spaces. Thus, we present a solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube). Rather than using a traditional 6 degree of freedom device, such as a flystick, users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks (not specific to a field).

Figure 4.1: *A user in the four-sided cube holding a tray with two hands.*

Interaction in immersive virtual reality systems (e.g., CAVEs[1]) has always been challenging, especially for novice users. In many systems, 6 degree of freedom devices are used for navigation as well as selection and manipulation tasks. Such devices are well-

[1]Cave Automatic Virtual Environment

established and can be very powerful since they allow users to perform actions which cannot be performed naturally, such as picking objects from afar, or navigating while physically staying in the same place. However, one goal of fully immersive systems is to enhance presence and immersion [Slater 2009a]. In such a context, flysticks can potentially degrade the realism and naturalness of the virtual environment. To avoid this shortcoming, we investigate the use of direct manipulation using finger tracking, resulting in an experience that is as close as possible to the manipulation used in the real world [Moehring 2011, Wexelblat 1995, Jacobs 2012, Hilliges 2012].

With the advent of hand and finger tracking solutions, there has been recent interest in using direct manipulation and gestures in immersive systems to achieve specific tasks, such as automotive design [Jacobs 2012, Moehring 2011], focusing on carefully handling the physics of collisions between hands and virtual objects. However, little has been done to investigate the usability of direct manipulation in a fully immersive setting for relatively complex, general purpose tasks. Moreover, the effect of direct manipulation on presence, or the similarity with real world manipulations has not been sufficiently researched. These questions are central to the research in this chapter.

We present and evaluate a solution which incorporates direct manipulation for grasping and moving objects, using both one and two hands, based on finger tracking with a glove-like device. Our solution operates in a four-sided immersive cube as mentionned in Sec. 3.1, providing a high sense of presence, and couples finger tracking with a real-time physics engine (see Sec. 3.3.2 for details), allowing close-to-natural manipulation of objects. We present a heuristic approach for direct hand manipulation which is robust to tracker noise and instabilities of the physics simulation (see Sec. 3.2.2 for details). We performed an exploratory study to evaluate our interface. We also calibrated the 3D position of the eyes for each participant at the beginning of the experiment to provide a more immersive and comfortable experience to the user (see Sec. 4.2.2.1).

The goals of our study are to evaluate:

a) whether direct manipulation is a feasible alternative to traditional immersive cube interfaces such as a wand for moderately complex, general purpose tasks,

b) the effect of using direct manipulation on presence,

c) the similarity to real world manipulation.

To provide a meaningful evaluation of feasibility, we purposely choose a task that involves quite complex translations and rotations while the user walks in the cube, and at the same time balances objects on a tray (see Sec. 4.2). We compare direct manipulation to traditional wand-based interaction, and both the wand and direct manipulation to a real world task through a study focused on the sense of immersion: the virtual world used is a replica of a real space in which the same tasks are performed (Fig. 4.13). We record both objective measurements (time to completion and errors or precision) and subjective judgments through the use of a questionnaire. Users are also asked to perform a free task, allowing us to observe their perceived level of presence in the scene.

Our hypotheses are:

a) using the wand will be more precise and faster than finger tracking based manipulation and

b) using finger tracking based direct manipulation will be more natural and will provide a higher sense of presence.

This chapter provides the two following contributions: a fully immersive solution which allows one- and two-handed interaction in a four-sided rear-projected cube, and, more importantly, an exploratory study focused on the sense of immersion comparing our direct manipulation interface both with the traditional wand and with a real world task. We will explain our heuristic approach for hand manipulation before detailing our experimental procedure and discussing the results of our study.

4.1 Heuristic approach for contacts

There are several difficulties in developing a direct manipulation interface in an immersive cube-like environment. In contrast to systems with a restricted workspace, in which the user is sitting [Prachyabrued 2012, Hilliges 2012] and thus the workspace is small, we target a room-sized environment, and allow the user to walk around the scene while manipulating objects. We identify four main difficulties:

- First, finger tracking in the immersive cube is challenging and needs specific treatment as explained in Sec. 3.2.2.
- Second, given our goal of assessing feasibility of direct manipulation, we ask users to perform relatively complex tasks: grasping and translating objects, including balancing objects one on top of the other while walking. In previous work, users often manipulate a single dynamic object per hand, with simple constraints, e.g., collision detection with static objects. In contrast, our tasks involve several dynamic objects, and multiple indirect constraints from physics-based interactions between objects.
- Third, our tasks require the integration of a fast physics engine as detailed in Sec. 3.3.2.
- Fourth, the physics simulation requires a careful synchronization of the displays. The simulation is hence run on one machine, and the transformation information of all dynamic objects is propagated to all slaves at each frame.

In the following, only the thumb and index of each hand are used; we call these *active fingers*; the middle finger is ignored in our current implementation. We made this choice since *active fingers* proved to be largely sufficient to provide a natural-feeling direct manipulation interface (see Sec. 4.4).

We also introduce the notion of *main contact* (see Fig. 4.2). When grasping with one hand, the *main contacts* are the actual contacts of the thumb and index with the object. When grasping with two hands, there is one *main contact* per hand. This is the actual contact between the finger and the object if only one finger is applied, or a "mean contact" if two fingers are applied with this hand. The "mean contact" is set at the midpoint between the two actual finger contacts and is useful for the grasp/release finite state machine described next.

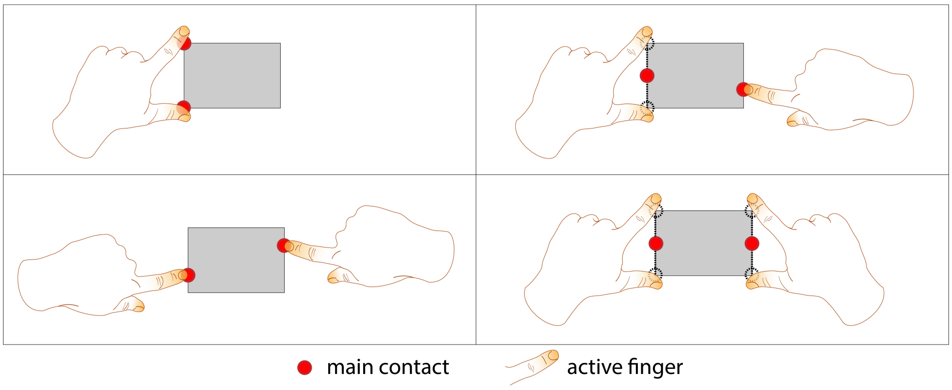


Figure 4.2: *Active fingers and main contacts when grabbing with one and two hands.*

4.1.1 Grasp/release heuristics

We directly use the physics engine to control and derive the manipulations in the environment. As explained in Sec. 3.2, cubic kinematic[2] objects are attached to the fingertips and the top of the palm. For each dynamic object, we track contact events with the fingers provided by the physics simulation. Once two kinematic contacts on a given object are identified, we mark the object as "grasped", until the contacts are released. We remove grasped objects from the physics simulation and apply the transformation and speed of the trackers to them so that the simulation of objects in contact is still correct.

When grasping, we store specific data (e.g. position and orientation of the selected object), called *grasping data* (see Fig. 4.3), which is reset each time the number of contacts on the object changes. These data are used to compute the transformations of the selected object, depending on the fingers movements.

[2] See Sec. 3.3.2 for a definition

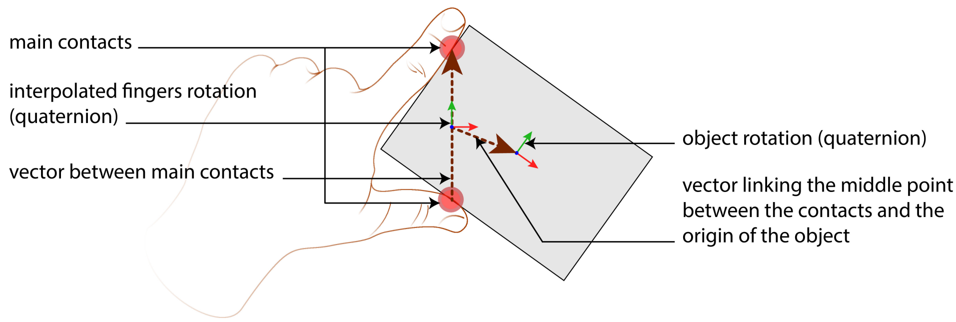


Figure 4.3: *Data stored when a user grasps an object or when the number of contacts onto the object changes.*

If we simply use the information from the tracker and the physics engine, users "drop" objects very quickly, or objects may move in an unstable manner. We make the selection more robust by preserving the kinematic contacts list and marking an object as "released" only when the distance between the contacts is 10% longer than the length of the vector stored in the *grasping data* (see Fig. 4.3). It is important to understand that contacts are determined by the physics engine in a natural manner when treated with our approach; they are not "pinch"-like gestures as the selection is determined by the actual contacts and distance between contacts and not by the movement (see Fig. 4.5). If a hand participates in holding an object by applying a single finger, it releases as soon as the distance between the two main contacts is 10% longer than that stored in the *grasping data*. We chose these thresholds by pilot trial-and-error tests in several different settings. While releasing, the object being handled follows the translation of the midpoint between the two *main contacts*. Accurately placing objects thus requires a short learning period (see Fig. 4.4 for an example). However, users did not complain about this limitation.

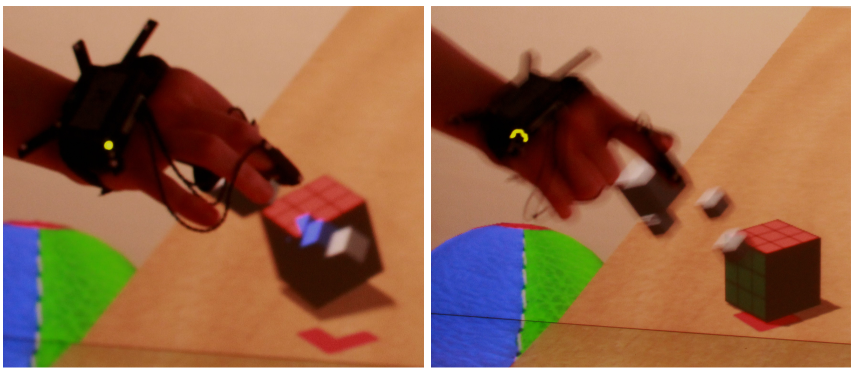


Figure 4.4: *Accurately placing a cube in a corner.*

The threshold for release is the main restriction of our approach in terms of "naturalness". If a user grasps a "thick" object with one hand, it is not possible for her to directly release it in the current implementation if she cannot open her hand by at least 10% of the current gap. To release the selected object, the user would need to switch from a one-handed grasp to a two-handed grasp to reset the *grasping data* and then release the object. This scenario did not occur in our tasks.

As mentioned before, selected objects are not handled by the physics simulation. The translation applied is that of the midpoint between the main contacts, and the rotation applied is that of the segment defined by the main contacts. When released, the objects are re-inserted into the physics simulation. If an object is released while the fingers are still inside it, it will automatically be re-grasped.

Additionally, color coding is applied to the representations of fingertips. By default, the cubes are white. When only one active finger is touching a dynamic object, the corresponding cube becomes green. When two active fingers of the same hand grasp an object, the corresponding cubes turn blue. When the object is grasped with both hands, the cubes in contact with the object become red (see Fig. 4.6). A similar coding was applied to the wand handles, besides the fact that only the white and blue cases were used. This provides visual feedback to the users to which they can refer when learning how to use the interfaces.

Figure 4.5: *A user spreads her arms to release a large object without the need to further open her hands.*

4.1.2 Finite state machine for object manipulation

When performing complex tasks, users naturally grasp and release objects, and often mix one- and two-handed manipulations. To treat such transitions, we design a finite state machine approach. This approach is similar to the one described by [Cutler 1997]. They define transitions between the use of one- and two- handed tools to complete tasks on a Responsive Workbench [Krüger 1994]. Depending on the default tools used, transitions may seem more or less implicit to the user. As we are concerned by transitioning between grasp states – which is a natural way of handling 3D objects – we place ourselves in the context of implicit transitions, where the transition mechanism is transparent to the user, because it is intuitive. A more physically accurate implementation would use friction cones [Moehring 2010] to determine grasping conditions, but we left this for future work.

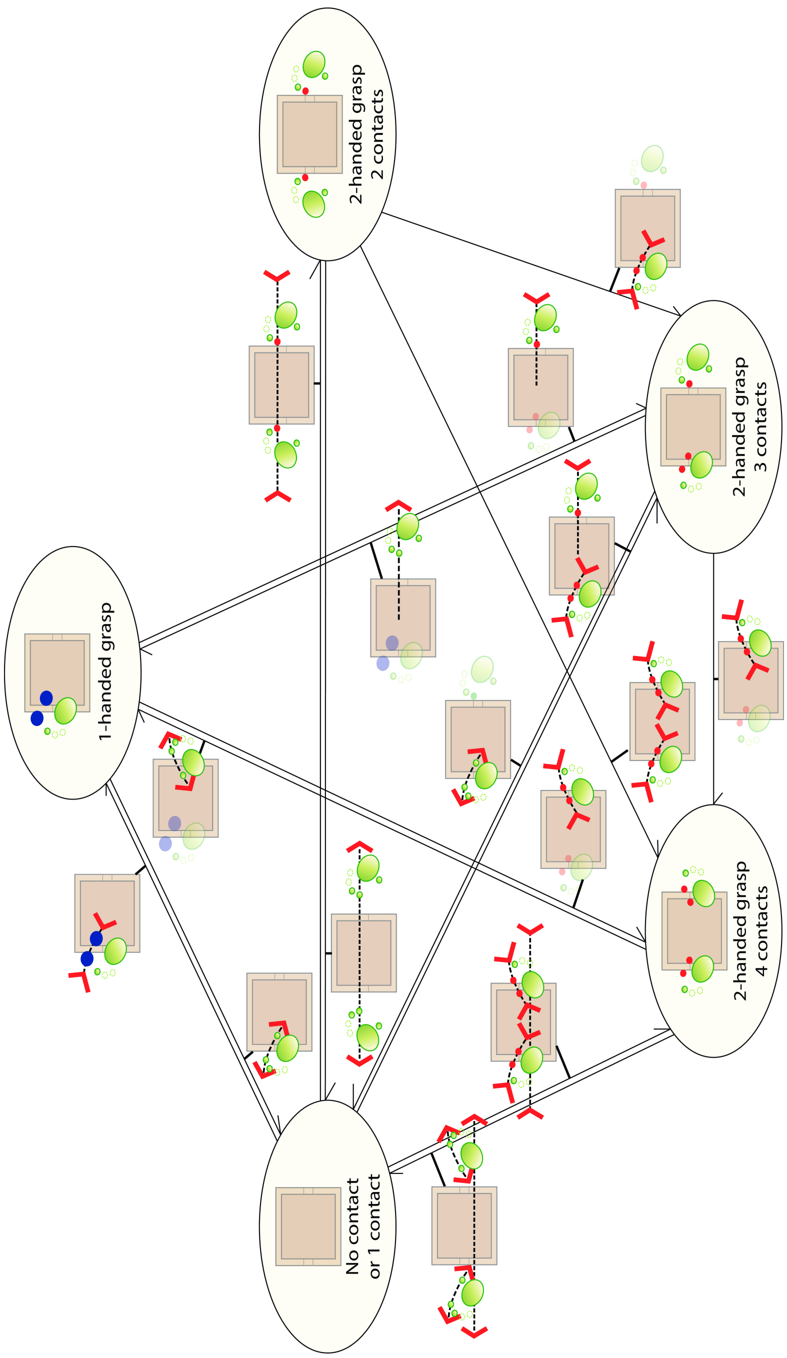


Figure 4.6: *Finite state machine graph detailing transitions between grasping and releasing states. A representation of the hand is shown in green; dots represent fingers. When only one active finger is touching a dynamic object, the corresponding dot becomes plain green. When two active fingers of the same hand are grasping an object, the corresponding dots turn blue. When the object is grasped with both hands, the dots in contact with the object become red.*

We next describe the different possible transitions for each manipulation.

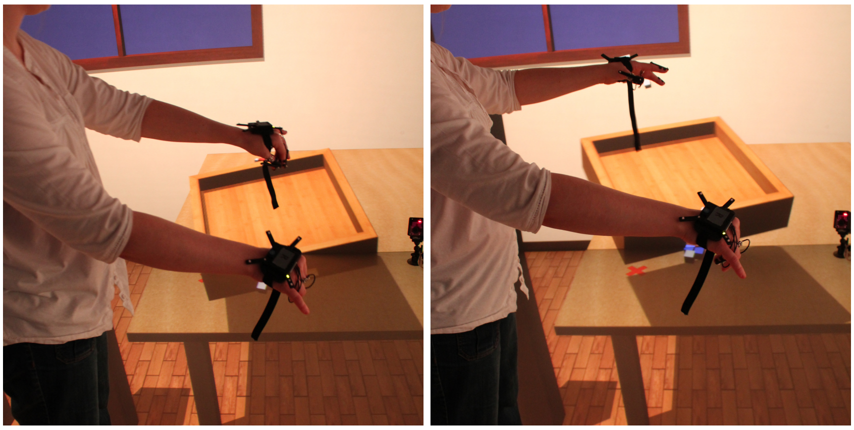


Figure 4.7: *Left: two-handed grasp. Right: two-handed release.*

Grasp An object is grasped when the user applies at least two fingers on it, with only one hand or both hands (see Fig. 4.7, left). The user can switch from a one-handed grasp to a two-handed grasp by touching the object with one or both active fingers of the second hand.

There is no alignment test on contacts, so users could lift an object with two contacts on the same side of a cube. However, this would be unnatural, and no user attempted this manipulation in our tests.

Release An object is released when there is at most one finger in contact with the object. For a grasping hand to release the object, the user simply has to open her fingers (see Fig. 4.7, right). If one or both hands are grasping just by applying one of their active fingers, the user simply has to remove this (these) hand(s) from the object. If both hands are grasping, the user can switch to a one-handed grasp by opening one hand. Finally, to completely release an object grasped with two hands, the user must spread her arms to ensure that the fingers are no longer in contact with the object.

The full list of possible transitions is provided in Fig. 4.6.

Translate The grasped object is translated when the fingers in contact translate. It follows the translation of the midpoint between the contacts.

Rotate The grasped object is rotated as soon as the fingers in contact rotate. We currently limit the rotation with two hands to simplify the implementation, avoiding problems with the combined effect of tracker noise on each hand: the user can rotate one hand with respect to the other (e.g., tipping out balls from a tray).

The user can switch from a one-handed grasp to a two-handed grasp and vice versa; two one-handed grasps can be used at the same time to manipulate two distinct objects (Fig. 4.8, right).

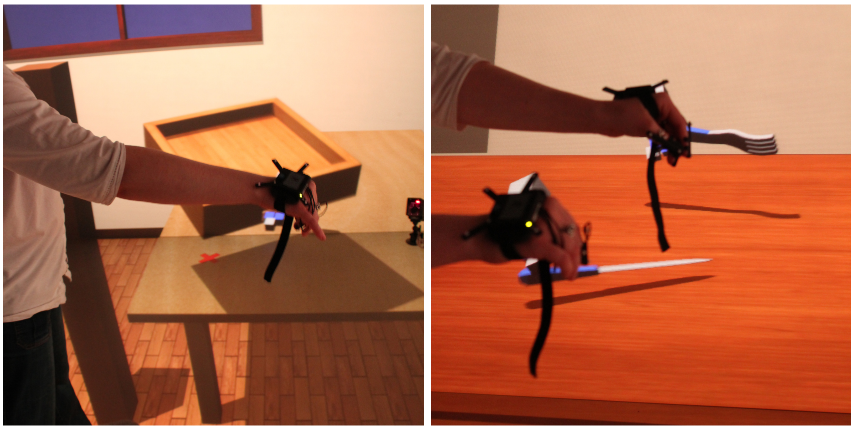


Figure 4.8: *Left: one-handed grasp. Right: two hands grasping two objects.*

4.1.3 Wand interaction

For the wand, we use a standard thin virtual ray emanating from the wand. To grasp an object with the wand, the user points the wand towards it and presses a trigger button when the ray traverses the object. Small blue spheres are displayed at contact points between the ray and the object when grasping (Fig. 4.9). The trigger is kept pressed during manipulation. To release an object with the wand, the user simply has to release the trigger.

We implemented the ray to be thin to limit occlusion, and to be long enough (1m) allowing sufficient coverage of the 3.2mx3.2mx2.4m immersive cube.

4.2 Experimental procedure

The experiment lasts 1h30 on average. It consists of an eye calibration step, a training session, a usability test, a "free form" user experience and the completion of a questionnaire. The usability evaluation has three conditions: using the wand, finger-based direct manipulation, and the real world condition. Hence, we had six groups of three participants to test all the possible orders of conditions, which were randomized. For direct manipulation and real, the users always performed the tasks first with two hands, and then with only one hand. We decided not to vary this order within the conditions using hands to permit the users to learn progressively. Specifically, when handling larger objects such as a tray, using

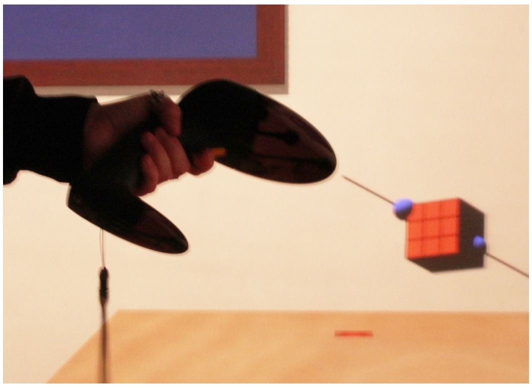


Figure 4.9: *Wand selection used for comparison.*

two hands provides better control and balance overall. This is because the main contacts are split onto two opposite sides of the object so the control vector (i.e. the segment between the main contacts) is large and almost encloses the gravity center. In contrast, using one hand involves grasping only one border of the virtual tray so the control vector is very short and far from the gravity center. As a result, a slight motion of one finger with respect to the other can result in a large motion of the tray, making it harder to control.

In the design of all tasks, we tried to minimize the cases of the hands incorrectly occluding virtual objects (i.e. when the object should be displayed above the hand, which is not possible with rear-projected displays). Evidently this is not always possible, but we did not receive any negative feedback about this from participants.

The training session and the free-form user experience session are not performed in the real condition, but the order of the other conditions (i.e., wand, one hand and two hands) is the same as in the usability task. Concerning the direct manipulation, the training session and the usability task are performed with two hands and one hand separately whereas, in the user experience, the user can freely manipulate objects with one or both hands.

4.2.1 Population

We ran the experiment with 18 participants, 10 men and 8 women aged between 24 and 59 years old. The mean age was 32.5 years, with a standard deviation of 10.7 years. Most had no experience with virtual reality (13 out of 18); 5 had experienced virtual reality demonstrations before. Four participants had previously used a wand, including three who had previously manipulated virtual objects.

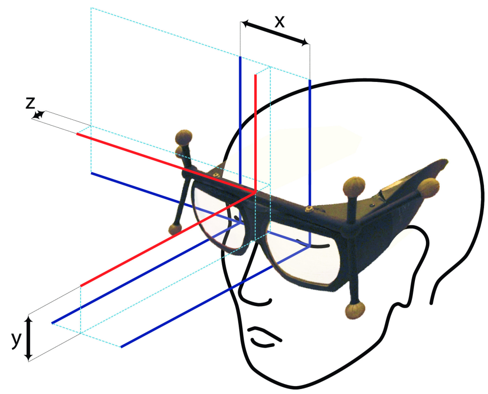


Figure 4.10: *The position of the user's eyes is calibrated in the three dimensions.*

4.2.2 Environments

4.2.2.1 Eye calibration

We performed a pilot study where we simply evaluated and set the interocular distance: this proved to be insufficient since the environment still displayed a significant "swimming" effect (perceived movement of static objects when moving in the immersive cube). To overcome this, we also calibrate the Y-Z positions of the eyes for each participant at the beginning of the experiment (see Fig. 4.10). This provides a much better immersive experience in the cube by improving the projection from the user's point of view (Sec. 3.1). This is also more comfortable for the user as it is expected to reduce the risk of cybersickness.

In a recent study, [Ponto 2013] came to the same conclusion and proposed a protocol to set the 3D eye position. However, in their method, they derive a function linking the interocular distance and the depth offset using perceived depth alignment. This observation is quite difficult as the virtual object is hidden by the real one, and the bias is mainly evaluated through the alignment of side borders which does not change much considering the allowed range for the interocular distance. We thus tried to come up with a simpler approach.

Our calibration is done using a simple scene, i.e. a floor and a stool (see Fig. 4.11). The experimenter progressively modifies the coordinates of the eye positions until the cubes representing the fingers are closer to the fingertips and the user feels comfortable with the projection. A subsequent step consists in modifying values to minimize the "swimming" effect.

4.2.2.2 Training session

The goal of the training session is to familiarize the participant with the interaction techniques. The experimenter first explains the color coding of the cubes and how to use the

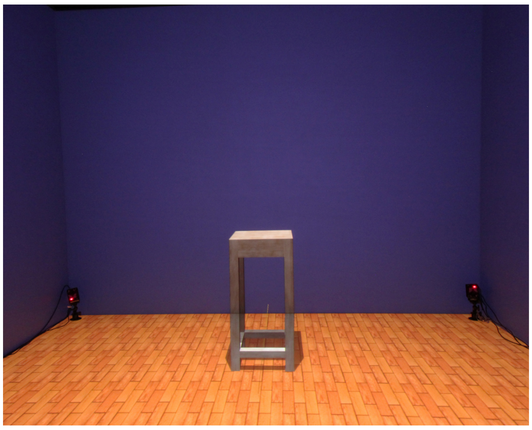


Figure 4.11: *Calibration Scene.*

techniques to grasp, release and move the objects.

To maximize immersion, the virtual scene consists of a closed room exactly the size of the actual immersive cube, with two tables, and three cardboard posts in between (see Fig. 4.12, left). On the left table, the user is provided with a colored cube and a tray containing two balls. There are red crosses marked on the tables which serve as targets: one under the cube on the left table, and two on the right table (see Fig. 4.12, right).

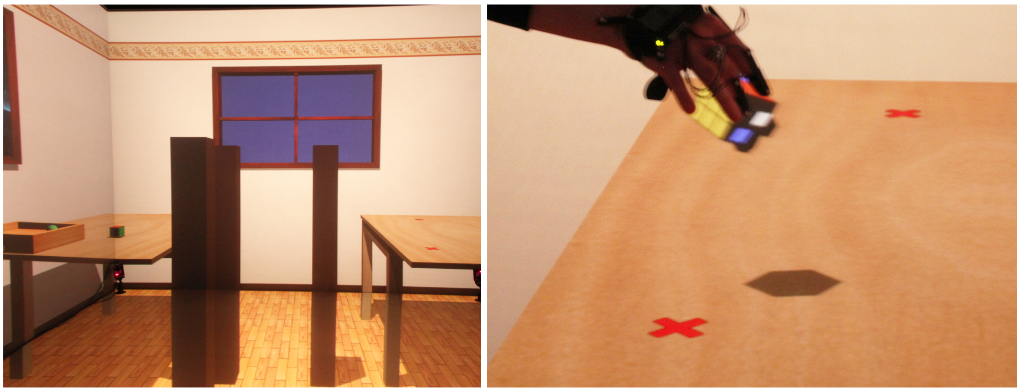


Figure 4.12: *Left: Training scene. Right: In the training session, target crosses are used to guide the participants.*

The experimenter demonstrates the movements in the immersive cube before the training session starts, explains that an alarm sounds when the tray or hands hit the posts, and asks the user to test this. The experimenter then explains that the goal is to keep as many balls as possible on the tray, and that the participant should avoid hitting the posts with the tray or hands. The participant starts the training by lifting the tray off the table. The tray must then be rotated by 90 degrees, passed between the two first posts, rotated again by 90 degrees and passed between the two second posts before being released on one of the tables (preferably the one on the right). The user can then repeat these steps and try other movements until she feels comfortable with manipulating the tray. Time and trials are not restricted during the training session.

Once this is done, the user ends the session by placing the colored cube onto the red crosses in a specific order. The experimenter explains that the sound heard when the cube touches the marks means that the subtask is validated, and that the same sound will be used when validating a subtask in the following session.

Once the session is complete, the scene is reset so that the user can train with the other technique (wand or direct manipulation), and when it is complete for the second time, the next session is automatically loaded.

The tray and the cube are manipulated with the wand and one hand; the tray is also manipulated with two hands.

4.2.2.3 Usability task

We evaluate usability in a single task which tests all the manipulations (i.e., grasping, releasing, translating and rotating).

The main evaluation scenario takes place in the same virtual room as in the training session but with different equipment. This room contains a stool on the left, the same table on the right, the same three cardboard posts in between, and a cupboard in the back. Again, the virtual scene has the exact size of our four-sided immersive cube. We place a tray with nine balls on the stool and a bowl on the front half of the table (see Fig. 4.13, left).

This session is also performed in the real condition: the same scene has been built in the vicinity of the immersive cube (see Fig. 4.13, right).

Figure 4.13: *Left: Virtual Usability Environment. Right: Real Usability Space.*

The experimenter first explains the entire task to the user by mimicking the required operations at the beginning of each condition, insisting on the order of the steps. The beginning and end of each subtask is determined automatically by the system. The task is composed as follows (see Fig. 4.14):

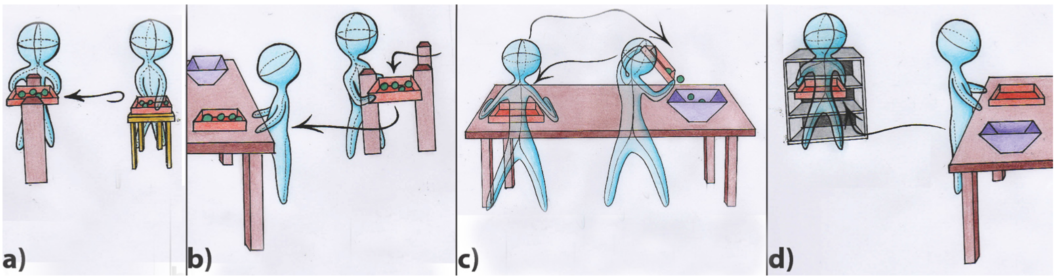


Figure 4.14: *Subtasks of the Usability Task (see text).*

- *Subtask 1:* The user grasps the tray and passes between the two sets of posts after rotating the tray by 90 degrees each time, and then releases it on the table. The participant is instructed to avoid dropping balls as well as to avoid touching the posts with the tray or the hands (see Fig. 4.14 (a) and (b) and also Fig. 4.15).
- *Subtask 2:* The user lifts the tray off the table and empties it into the bowl before releasing it on the table again. The user is instructed to keep as many balls as possible into the bowl (Fig. 4.14 (c)). If the tray is empty at the end of subtask 1, this subtask is skipped.
- *Subtask 3:* The user picks up the empty tray again and places it inside the cupboard. (Fig. 4.14 (d)).

Figure 4.15: *The user passes the tray between the posts.*

This task is quite challenging, even in the real world. Notice that the first subtask corresponds to what the user already had to do in the training session, as this is the most challenging part of the task.

4.2.2.4 User experience

The free-form user experience session corresponds to a qualitative observational evaluation, to see what the user will do with almost no instruction.

The scene is a dining room, with a cupboard, shelves and a table. The table is empty, and plates, glasses, forks and knives for six people are placed in the cupboard on the left and on the shelves in the back (see Fig. 4.16). The virtual scene has also the exact size of the immersive cube.

Figure 4.16: *The user experience scene. Left, at the outset the table is clear. Right, at the end the table is set with plates, cups and cutlery.*

The user is introduced to the scene and receives no other instructions than “you have 4 minutes to set the table as best as you can." Our goal is to observe general behavior, i.e., whether participants use one or two hands, whether they focus on completely setting the table or correctly placing objects, the order of setting the table etc. We are interested in observing whether participants behave naturally, e.g., walk around the virtual table, catch falling objects etc., as well as their degree of presence.

4.2.3 Making real and virtual conditions equally difficult

To compare real and virtual tasks, we need to have approximately the same level of difficulty between the two. However, material properties (friction etc.) are only approximate and the physics simulator handles dynamics with impulses which can sometimes be unrealistic when running in real-time.

To perform a fair comparison, we performed a pilot study with four participants who performed the task of balancing the tray avoiding the posts and minimizing the number of balls dropped, in both the real and virtual environments. We measured difficulty by counting the number of balls dropped. The parameters we adjusted are the height of the borders of the virtual tray and the type of the real balls (see Fig. 4.17).

Because of the impulses, virtual balls tend to bounce more on the tray in the virtual environment than those in the real world. We thus increased the borders of the virtual tray. In addition, the real balls we used had different friction so we mixed two ball types: ping-pong balls and foam balls. The pilot study showed that equivalent difficulty level between real and virtual tasks is obtained with a tray border raised to 3/4 of the height of the balls, and mixing 3 foam balls and 6 ping-pong balls. We used this configuration in all experiments.

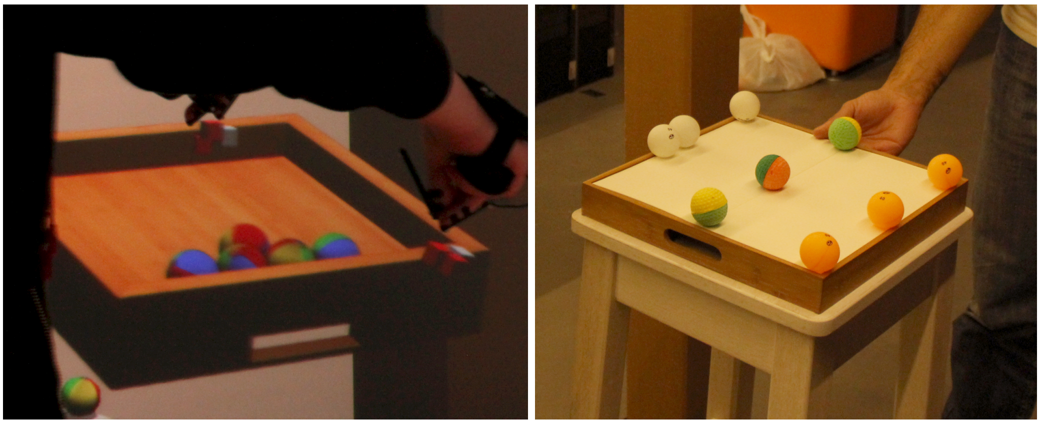


Figure 4.17: *Left: Virtual tray and balls. Right: Real tray and balls.*

4.3 Measurements

In each session, we record the head position and orientation of the user at each frame. We also record the position and orientation of the fingers and palm, as well as the time to complete each task and subtask. Every object collision and ball dropped are also recorded. We videotape the sessions, and manually extract the completion times for the real environment.

4.3.1 Objective metrics

During the usability task, we measure accuracy by recording the following:

(1) position of the tray with respect to the posts to make sure that the user does pass the tray between them;

(2) number of times the tray touches the posts;

(3) number of times the hands touch the posts;

(4) number of balls remaining on the tray when releasing the tray onto the table;

(5) number of balls inside the bowl after emptying the tray.

4.3.2 Subjective metrics

At the end of the experiment, participants complete a questionnaire. For each technique in virtual conditions (wand, one hand, and two hands), they are asked to rate various criteria on a Likert scale between 1 and 7. We evaluate: ease of use, fatigue caused by using the technique, sensation of "being there", plausibility of the interaction with the environment and of the reaction of the environment to actions, similarity to the real condition (for one hand and two hands), precision, naturalness and cybersickness. Participants are also asked to rate the similarity of each virtual task to the real task. Then, they have to answer open questions related to their strategies for using the interfaces, and their opinion on advantages, drawbacks and difficulties of each interface. Finally, they are free to make additional comments.

4.4 Results

We next present the statistical analysis of the results for both our objective measurements, i.e., speed and errors, and the responses to our subjective questionnaire. For the completion times, we performed a Shapiro test that rejected the normality hypothesis on the data distribution. Thus, we used a non-parametric Friedman test [Cunningham 2011] for differences among the conditions. Post-hoc comparisons were performed using Wilcoxon signed rank tests [Cunningham 2011] with a threshold of 0.05 for significance.

4.4.1 Objective measurements

We first performed a Friedman test on the time performances between the 5 conditions, i.e., 1- and 2-handed real, 1- and 2-handed direct manipulation and the wand. We use the following abbreviations: R1H and R2H are real, one- and two-handed conditions respectively, and DM1H and DM2H are virtual direct manipulation, one- and two-handed conditions respectively. The reported p-values were adjusted for multiple comparisons. We found a significant effect ($\chi^2 = 4.62$, $p < 0.001$) of condition. Post-hoc analysis revealed that the time to complete the task was significantly lower for R1H with a median time to completion of 61.33s compared to DM2H where time to completion was 113.8s, ($p < 0.001$). The time for R2H (median = 65.67s) was significantly lower than DM1H (median = 86.95s, $p = 0.04$), DM2H ($p < 0.001$) and Wand (median = 88.86s, $p = 0.04$). No significant effect was found between virtual conditions.

Errors were measured in two ways: first the number of balls lost throughout the task, and second the number of times the tray or the hands hit the posts in the virtual tasks. We also compare real and virtual conditions for both criteria.

We performed a Friedman test for the number of lost balls during the task. We found a significant effect of condition ($\chi^2 = 3.35$, $p = 0.007$). The number of lost balls was significantly lower for Wand compared to R1H and DM1H ($p = 0.007$ and $p = 0.01$ respectively), as revealed by post-hoc analysis.

Finally, we performed a Friedman test for the number of collisions during the task. The test revealed a significant effect of condition ($\chi^2 = 5.32$, $p < 0.001$). The number of collisions was significantly lower for R1H compared to all virtual conditions as shown by post-hoc analysis ($p < 0.001$ for DM1H , $p = 0.02$ for DM2H and $p < 0.001$ for Wand). The number of collisions was also significantly lower for R2H compared to all virtual conditions ($p < 0.001$ for DM1H, $p = 0.003$ for DM2H and $p < 0.001$ for Wand). We did not find any significant effect between the virtual conditions.

These results are summarized in Fig. 4.18 which shows "lower than" relationships between conditions for each parameter when it is significant.

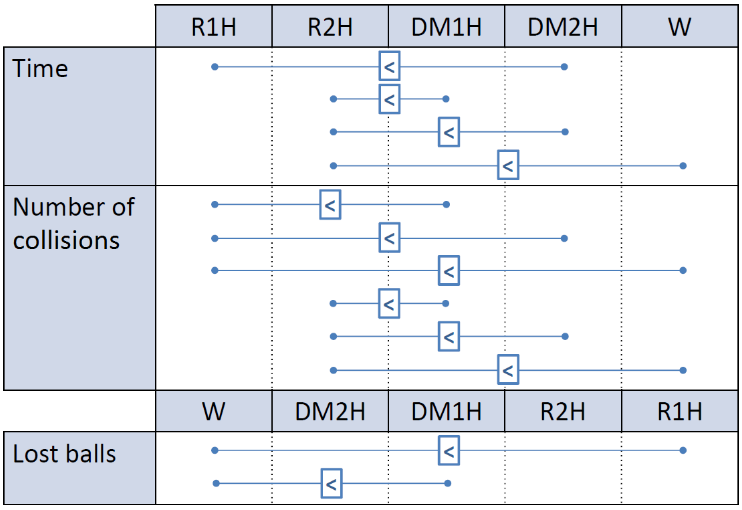


Figure 4.18: *Significant "lower than" relationships between conditions for objective measurements.*

4.4.2 Subjective measurements

For the responses to the subjective questionnaire, we performed a Friedman test for the different criteria between the three virtual conditions. We did not find any significant effect for *Plausibility* for the usability task, *Being there* for the user experience and *Cybersickness*. We found a significant effect for 7 criteria: *Ease of Use* for the usability task ($\chi^2 = 4.95$,

$p < 0.001$), *Being there* for the user experience ($\chi^2 = 2.69$, $p = 0.02$), *Plausibility* for the user experience ($\chi^2 = 2.54$, $p = 0.03$), *Fatigue* ($\chi^2 = 4.27$, $p < 0.001$), *Similarity with real experience* ($\chi^2 = 3.24$, p=0.003), *Precision* ($\chi^2 = 4.35$, $p < 0.001$) and *Naturalness* ($\chi^2 = 2.70$, $p = 0.02$) (see Fig. 4.20). Post-hoc analysis showed that Wand was preferred to DM1H and DM2H for *Ease of Use* during the usability task ($p < 0.001$ and $p = 0.003$ respectively), *Fatigue* ($p < 0.001$ and $p = 0.007$ respectively) and *Precision* ($p < 0.001$ for both). In contrast, DM1H and DM2H conditions were significantly better rated than Wand for *Being there* for the user experience ($p = 0.04$ and $p = 0.02$ respectively). The two hands condition was preferred to Wand for *Plausibility* for the user experience ($p = 0.03$), *Similarity with real experience* ($p = 0.003$) and *Naturalness* ($p = 0.02$).

These results are summarized in Fig. 4.19 which shows "lower than" relationships between conditions for each parameter when it is significant.

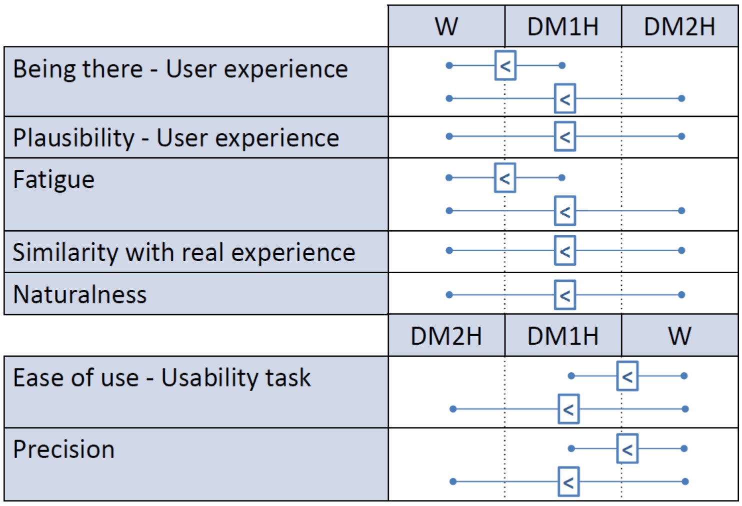


Figure 4.19: *Significant "lower than" relationships between conditions for subjective measurements.*

4.4.3 Discussion

Our goals were to evaluate the feasibility of direct manipulation in a fully immersive space, and its effect on presence as well as the similarity to real world manipulation. We hypothesized that the wand would be more precise and efficient than direct manipulation, but we believed that direct manipulation would positively affect the sense of presence. The experimental results support these hypotheses.

Concerning the first hypothesis, the results show that all virtual tasks took longer than the real world task and that the wand and direct manipulation are equivalent in terms of speed (even though the two-handed direct manipulation condition had a longer median completion time). We take this as an encouraging indication that direct manipulation does

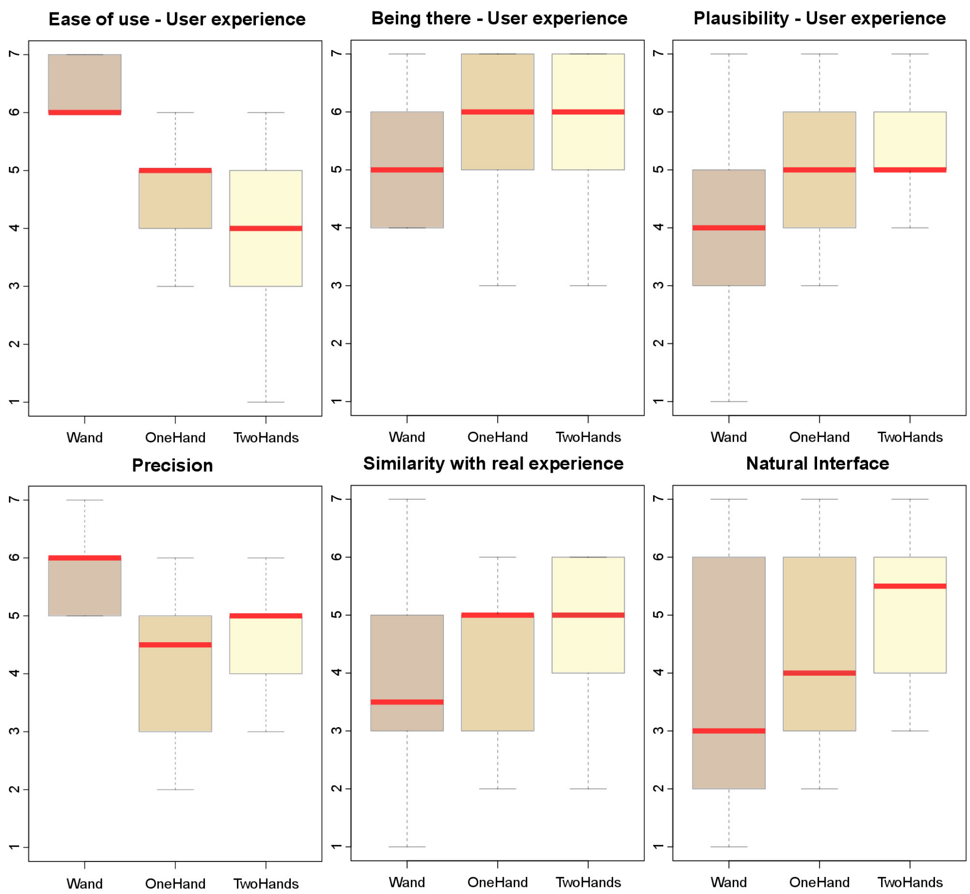


Figure 4.20: *Boxplots of the statistical results for some of the most interesting studied criteria. Each boxplot is delimited by the 25% quantile and 75% quantile of the distribution of the effect over the individuals. The median is also represented as a red line for each effect.*

not penalize speed. However, as a general remark, it also indicates that for such complex tasks involving balance and rotations, we are still not at the point where virtual tasks can be performed at the same speed as their real equivalent. In the real scene, the user is influenced by the kinesthetic perception (of touch, weight and muscle tension) which guides balance of the tray but increases fatigue. In contrast, this sense is missing in the virtual setting; note however that current solutions for haptics in immersive cubes are unsatisfactory and we thus chose not to opt for such a solution.

In terms of accuracy, users dropped fewer balls using the wand than with the one-handed virtual direct manipulation, as well as with the one-handed *real* life condition. There was no significant difference with the two-handed direct manipulation, real or virtual. Evidently, the wand is a "hyper-natural" interface in the terminology of Bowman et al. [Bowman 2012], so it is unsurprising that it allows better performance than the real world in some cases. The fact that this occurred for the one-handed case rather than two-hands is due to the inherent difficulty of the one-handed condition: the tray is a relatively long object, and a small movement of the fingers of the holding hand can result in a large movement of the tray, and thus a loss of the balls. When using two hands, the tray is more stable (see the introduction of Sec. 4.2). The above observations are true for the virtual setting, but also for the real tasks; several participants complained that the one-handed *real* task was hard and tiring.

The above two results for speed and accuracy indicate that the virtual direct manipulation is a feasible alternative for interaction. Similarly to the work by [Bowman 2012], there is an indication that direct manipulation can be considered better for applications such as training since performance is closer to the real world than the wand which *augments* interaction capabilities of the user.

The participants found the wand easier to use and less tiring. We also noticed that users subjectively considered the wand to be more precise than the direct manipulation even if the objective performance did not always confirm this, notably in terms of hitting the posts. Again, the "hyper-natural" aspect of the wand is a plausible explanation for this perception and this discrepancy concerning manual tasks.

For the user experience, participants rated the sense of *being there* to be higher for hands (both one- and two-handed) compared to the wand. In informal interviews after completing the study, participants explained that in the usability task they were so concentrated on completing the task that they did not pay much attention to the environment; this was not the case however for the user experience, where there was no constraint. Similarly, two hands were rated higher than the wand for the sense of *plausibility*, again in the user experience. We believe that this is not the case for the one-hand case because of the lack of precision when manipulating objects, as explained above.

The above results on the subjective ratings show that our direct manipulation interface does have an effect on the two components of presence, *plausibility* and *being there* [Slater 2009a]. In addition, for the two-handed case, participants perceived them as more natural and closer to the real experience than the wand. We believe that these are encouraging results, which support our hypothesis that direct manipulation can improve the sense of immersion in such virtual environments and provide an experience that is closer to reality than using more traditional device-based interfaces.

We also observed participants behavior informally, which revealed several interesting cases of their reactions to our virtual environments. The immersive cube offers a high level of immersion in and of itself; for example, participants attempted to place the wand on the virtual tables at the end of the different sessions. However, users tended to avoid walking through the virtual objects when they used their hands. In several cases, participants tried to catch the dropping objects as an automatic reaction when using direct manipulation. During the user experience, participants tended to place the objects in a specific order, as they do in real life: they begin with the plates, then the glasses, and they finish with the forks and knives. Participants also tend to develop strategies to use the different techniques in the user experience. When using the wand, participants would stay in a convenient place to pick and place objects from a distance. However, they found it harder to rotate the objects, although different techniques for rotation could be implemented to alleviate this problem. When using direct manipulation, they used both hands to adjust the orientation of the plates; otherwise, they used both hands to pick two objects separately, and thus can set the table faster. These behaviors witness the immersion of our virtual environments, as people behave as they do in real life.

Our results show that, even if the wand outperforms direct manipulation in terms of usability, our finger-based direct manipulation interface conveys a high sense of presence and naturalness. We also believe that improved tracking technology and possibly the use of five fingers will allow our approach to outperform the wand, while increasing the difference in presence between the two. Lastly, finger-based direct manipulation would also clearly benefit from haptic feedback.

4.5 Conclusions and future work

In this chapter, we have presented a complete system which integrates direct manipulation with real-time physics in a fully immersive four-sided cube, clearly demonstrating its utility and indicating some advantages over traditional 6 degree of freedom interfaces, in terms of immersion, presence, and similarity to the real world. Interaction is based on a physics engine, enhanced by a heuristic approach to manipulate objects in a virtual environment in a close-to-natural way. We examined feasibility of direct manipulation, leading us to test relatively complex tasks, involving translations and rotations of objects and maintaining balance while walking.

Our exploratory user study included a usability task, and a free-form user experience task. We tested our direct manipulation compared to the traditional wand, and for the controlled setting, we replicated the virtual task in the real world. Both the objective measures (speed, accuracy) and the responses to the subjective questionnaire indicated that direct manipulation is a feasible alternative to the more traditional wand interface. The results of our study also indicate that in several cases, especially when using two hands, the use of direct manipulation enhances the sense of presence in the virtual environment, and is perceived as being closer to reality.

Given that the feasibility of the interaction with direct manipulation is then quite clear, it is interesting to examine the different parameters of our system in separate, more specific studies (see Chapter 5). An interesting direction for future work also involves the use of a full hand model with complete tracking, similar to [Hilliges 2012] and [Borst 2005a].

CHAPTER 5

Evaluation of Direct Manipulation in an Immersive Cube: a controlled study

In this chapter, we will investigate rotations and translations in a controlled manner when manipulating 3D virtual objects. We will also study the cases of glove-based direct manipulation and flystick-based interaction, in the context of fully immersive spaces. To do so, we propose an experimental design in which we decompose the users' actions into single and multiple degrees of freedom through the use of carefully designed devices. We analyze how the users comprehend and anticipate the isolated movements to achieve their goal of reaching targets with increasing complexity. We also observe the difference in their behavior when facing virtual and real devices (i.e., how they apprehend the mechanisms).

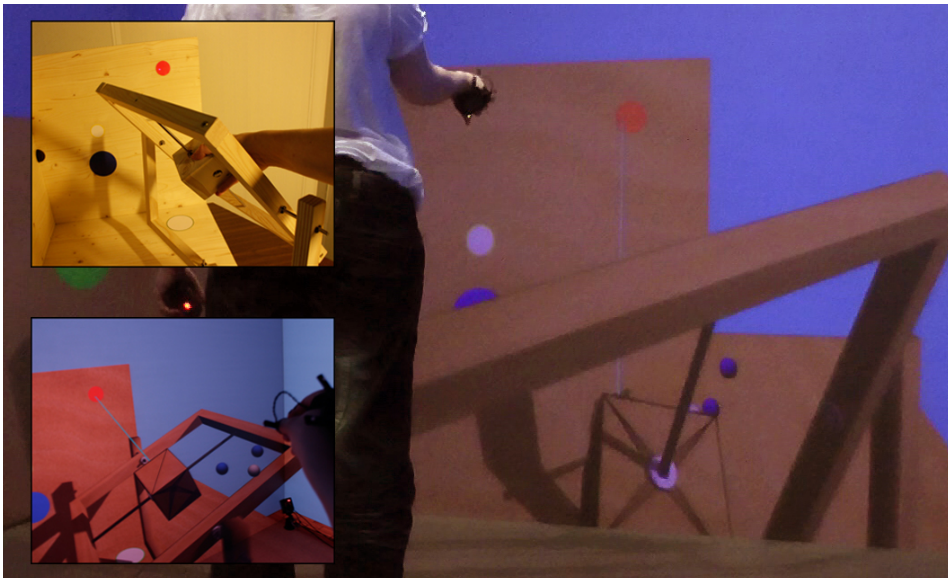


Figure 5.1: *A participant manipulating the device for 3D rotations with her hands. The insets show what the user sees, in both physical (top) and virtual (bottom) setups.*

Previous research has shown that it is best to reduce the use of devices in the design of fully immersive applications; such design tends to improve the realism and naturalness of the virtual experience, and also the user's sense of presence [Slater 2009a]. Such a search for natural interaction is justified by the fact that their intuitive aspect will ensure that training transfers to the real world [Bowman 2012] or improve industrial design [Moehring 2011]. A common solution to this is the use of glove-like devices which allow direct manipulation in virtual environments.

However, a legitimate question is whether such interfaces should always be preferred to more traditional wand-like devices, notably when the application focuses on performance, since it is reported that such interfaces can cause fatigue and frustration due to tracking problems [Buchmann 2004] (see Sec. 3.2 for details on tracking issues in the context of fully immersive spaces). Moreover, there has been little previous work on the effect of direct manipulation on the user's performance in a controlled context. In our previous experiment (see Chapt. 4), we were interested in studying direct manipulation compared to the use of a wand and to real world manipulations for complex tasks. We obtained interesting results: natural interaction increases presence, but wand-based interfaces outperform glove-based ones. However, this study involved complex tasks and many factors. It is thus also important to investigate these factors more in depth.

One of the fundamental and most influential parameters in object manipulation is the movement itself, and the interface may directly hinder or enhance it. Studying the efficiency of the movement is a classic way of evaluating an interface since the seminal work by [Fitts 1954]. In our context of virtual object manipulation, the involved tasks require movements with relatively few constraints. Thus we are not interested in the trajectory, contrary to steering law experiments [Accot 1997], but only in the final placement of the objects.

In our case, we chose to study the movements by decomposing them into main components (rotations and translations) and even further into single and multiple degrees of freedom. Our goals are to perform a carefully controlled study of 3D movements through their decomposition, and to design an experiment that can be reproduced in real and in virtual settings. The physical setup will serve as a baseline. From the results of our previous study, we strongly believe manipulation in this setup to be faster and more intuitive than in the virtual one (partly because of the lack of haptic feedback in the latter). Discrepancies are also thought to appear between the two virtual interfaces due to their fundamental and technical differences. Our preliminary hypotheses are that:

a) The reliability of the wand device should make it easier to use overall as we know that finger tracking technology presents significant noise issues (see Sec. 3.2.2),

b) The wand should perform better for translations, because then the wand comes down to a basic 3D cursor for pick and place (i.e. no grasping complexity),

c) Direct manipulation should be better when rotations are involved due to the cumbersomeness and resulting occlusions of the wand.

The hardware setup and software framework we use for virtual interaction are described in Chapt. 3, and the physical setup consists of the exact replica of the virtual evaluation devices in the vicinity of the immersive space.

The key contribution of this work is the careful design and creation of physical and virtual devices that permit this decomposition in a rigorous manner, and that allow users to perform tasks with measurable performance for each condition, in both real and virtual setups. It was thus of the utmost importance to implement a physical technique from which we could derive what can be expected in the virtual setup.

5.1 Experimental design

In this study, we are interested in decomposing 3D manipulation tasks into translation and rotation components, which are further decomposed into single and multiple degrees of freedom.

The experiment is a within subject design (every participant does all of the tasks) with two primary factors and two secondary factors. Our primary factors are the *input technique* and the *movement type*. Input techniques are *physical*, *direct manipulation* and *wand*. We decompose movement types according to the number of degrees of freedom (1, 3 or 6) and the nature of movement (translation, rotation or both), leading to 9 types. We specify the movement direction (x, y or z) for each 1 degree of freedom task as well as the movement nature (translation or rotation), which results in six types. We specify the movement nature (translation or rotation) for each 3 degree of freedom task, which results in two types. The 6 degree of freedom type permits free movement along all axes. Primary factors result in 27 primary conditions (see Table 5.1). We also introduce *target width* (large or small) and *target length* (short or long) as secondary factors to vary the difficulty of the tasks, leading to 4 secondary conditions. We thus have a total of 108 unique conditions.

We designed 5 devices: for 1 degree of freedom translations, 1 degree of freedom rotations, 3 degree of freedom translations, 3 degree of freedom rotations, and 6 degree of freedom free movement (see Fig. 5.2 and Sec. 5.2). We don't use trackballs since we want interaction spaces larger than desktop-sized ones.

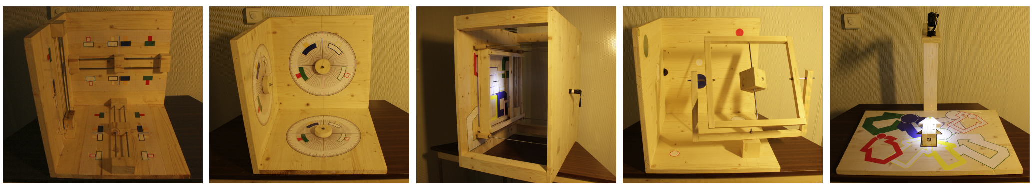


Figure 5.2: *Devices used to decompose motion. From left to right: 1D translations, 1D rotations, 3D translations, 3D rotations, free movement.*

The main concern and contribution of this work is thus the design of these evaluation

Movement type			Technique
Degrees of freedom	**Movement nature**	**Motion axis**	
1	Translation	X	Physical
			Direct manipulation
			Wand
		Y	Physical
			Direct manipulation
			Wand
		Z	Physical
			Direct manipulation
			Wand
	Rotation	X	Physical
			Direct manipulation
			Wand
		Y	Physical
			Direct manipulation
			Wand
		Z	Physical
			Direct manipulation
			Wand
3	Translation		Physical
			Direct manipulation
			Wand
	Rotation		Physical
			Direct manipulation
			Wand
6	Translation and rotation		Physical
			Direct manipulation
			Wand

Table 5.1: *Primary conditions of the experimental design*

devices and tasks. We needed to design physically and virtually feasible devices permitting these isolations of different motions, as well as tasks allowing us to measure performance. We defined a set of constraints to guide our decisions.

5.1.1 Design challenges

We defined the following guidelines:

- The devices should permit to isolate degrees of freedom, allowing us to achieve 1 and 3 degree of freedom tasks without the interference of unwanted degrees of freedom.
- The physical devices should be robust as they will be manipulated by many users

and we want the experiment to be reproducible in the same conditions for every participant.

- The devices should be large enough for significant measurements: the user has to be able to perform large movements as we want to significantly vary the amplitude of the movement to vary the difficulty of the tasks.
- The manipulated physical objects should be as light as possible to reduce fatigue as a) the experiment is long because we have many conditions and b) the user will not have actual objects in the virtual setup, and, nonetheless, we want the physical and virtual experiences to be as similar as possible.
- Due to the similarity requirement, we need to minimize friction in the physical setup as there is no friction in the virtual setup (see Sec. 5.1.2.2 for details).
- Due to the experiment duration and users' accumulated fatigue throughout the experiment, we also want the tasks to be relatively fast to perform, as we will need the users to perform several replications of all the tasks for stability and thus pertinence of the results.
- The tasks should be as simple and intuitive as possible so that the users can behave naturally: we want to minimize the complexity overload so that users can focus on the tasks and the movements themselves rather than on understanding how the mechanisms work. We thus restrict the actual manipulation to a single mobile object at a time.
- As a consequence of the simplicity requirement, and as the complexity is thought to increase with the number of degrees of freedom, we want the devices of higher complexity to build on what has been learnt with the devices of lower complexity.
- The devices should be easily realizable, both in real and in virtual conditions, for reproducibility.

Also, due to the constraints of the human limbs, object manipulation is dependent on the orientation of the arm [Ghez 1991, Lacquaniti 1992]. Thus, for single degree of freedom movements, we also account for the effect of the axis of the movement. The corresponding devices should thus provide a way to replicate the tasks in all 3 dimensions individually.

Concerning movements implying 3D translations, the user must be able to easily control her placement, looking at a single position. To do so, we decided to use 2D projections of the mobile object. We use a light source and shadow for this, as the shadow position and size are direct consequences of the 3D position of the object with respect to the source light (see Sec. 5.2.3 for details).

To obtain effective designs, we choose to derive our systems from existing well-known, well-established mechanisms which are as simple as possible.

5.1.2 Building and implementation details

5.1.2.1 Mechanical decisions and encountered problems of physical devices

The devices are made of pine, which is light and easy to work with, and which is robust enough for our purpose. Axes are made of stainless steel for robustness and rigidity; even if they are relatively heavy, they are not directly manipulated by the user, so that their weight does not interfere with the movement. To minimize friction, we use needle-bearings for rotational constraints, and we designed lubricated sliders for translational constraints. Concerning the latter, the wood will absorb the lubricant, resulting in a sort of auto-lubricating system for a while; when the oil is completely absorbed and the slider is dry, it can easily be re-oiled.

Concerning 1D devices, the corresponding tasks are performed on a panel, which is replicated twice, to build up a system of three panels (left, front and bottom), each one allowing the manipulation in a different direction. Due to this configuration, the visibility and handling ability of these devices is reduced for left-handed people. The devices could be adapted for them, but we preferred to keep the same configuration for all of our participants. We thus decided to run the experiment with right-handed people only.

As can be seen in Sec. 5.2, the tasks for 3D translations and 6 degree of freedom free movement involve projected shadows: the users have to place the mobile object so that its projected shadow fits into the specified targets. However, such tasks imply that the manipulation volume is reduced so that the projected shadows – and thus the targets – are of a reasonable size. We thus needed to make a compromise between the device size and the targets placement to avoid overlay. To solve this visibility issue, we decided to set the current target on top of the others in the virtual setup, and to complete the target contours with dot lines in the physical setup (see Fig. 5.3).

However, and despite our efforts to make physical and virtual as similar as possible, the physical constraints of visibility impose some limitations:

a) In both settings, the frame around the cube in the 3D translation device (see Fig. 5.10) also casts shadows.

b) In the physical setting, for both devices using shadows, the user's hand casts shadows, and may also hide the target (however, the user is allowed to move to improve her visibility).

c) In the virtual setting, lighting is based on shadow maps, whereas in the physical setting, lighting implies soft shadows which make some targets more difficult to validate.

For 3D translations, despite the robustness and rigidity of the stainless steel axes, the size of the device and thus the length of the axes implied some elasticity. Hence, we decided to strengthen the mechanism with extra wooden frames (see Fig. 5.10). Due to the

Figure 5.3: *Target visibility. Left: in the physical setup, the contours of the targets are completed with dot lines. Right: in the virtual setup, the current targer is displayed on top.*

increased weight to manipulate, we rotated the device on its side to reduce the effect of the frame weight: the user almost only needs to counterbalance for the weight of the mobile cube.

5.1.2.2 Use of the physics in the virtual setup

In this experiment, we use the same heuristic approach for manipulation as described in Sec. 4.1, which is built on the software framework detailed in Sec. 3.3.

However, contrary to our previous experiment (see Chapt. 4), where we had simple objects interacting through classic collisions, we have here complete mechanisms, implying constraints between objects which are crucial to respect the movement decomposition requirement. While the tasks are simpler to the user in this experiment, the internal handling of the objects behavior – which is transparent to the user – is more complex. As the intermediate library does not implement concave objects and considering our mechanisms, using the physics library to realistically simulate the constraints was not possible. Thus, the physics virtual world is only aware of the mobile objects the user will manipulate, and of the interaction widgets (fingertips and wand spheres), and has no gravity and no friction. The physics library is only used to detect collisions between the objects it is aware of, and the rest of the behavior is mathematically computed. This sometimes results in a slight lack of realism (such as the user's arm traversing the frame of the device for combined 3D rotations), but such a simple simulation proved sufficient for our experiment.

5.1.2.3 Wand visualization

During our previous experiment (see Chapt. 4), we noticed that, when manipulating objects with the wand, users tended to grab with the tool being far from the object, even when the user was relatively close to it. This lead to uncomfortable body configurations interfering with the manipulation, such as one's wrist being next to one's shoulder. We thought that this might be due to the significant length of the wand ray, which could be intuitively misleading despite the experimenter's instructions. We thus reduced the length of the ray and made it thicker for better visualization. However, many users still grabbed from the tip of the wand representation (such as when using a fork), but the discomfort was reduced due to its shortened size. When seeing users experiencing some discomfort, the experimenter had to clearly mention that the placement of the object on the ray had no importance on the grab.

Figure 5.4: *When grabbed, the object is highlighted.*

5.1.2.4 Visual feedback

In the virtual setup, users are provided with visual feedback on the manipulation widgets (fingertips and wand spheres), and on the manipulable objects themselves. Fingertips are represented as small spheres similar to the wand handles. The widgets color coding is exactly the same as described in Sec. 4.1.1. The feedback is reinforced by highlighting the mobile objects with visible black edges when they are grabbed (see Fig. 5.4).

5.2 Evaluation devices

As mentioned in Sec. 5.1.1, our designs are based on simple and well-established existing mechanisms.

For all of the devices, each target is presented as a set of 2 colored areas: one filled with the color, referred to as *solid color*, and a white area with a colored contour, referred to as *outline color*. The user will have to go from *solid color* to *outline color* back and forth (see Sec. 5.3.1.3 for details on task completion). As we have 2 *target widths* – large and small, respectively corresponding to easy and hard tasks – and 2 *target lengths* – short and long distance between colored areas – we have 4 levels of task difficulty, each one represented by a determined color: easy short is blue, easy long is green, hard short is yellow, and hard long is red (see Fig. 5.5).

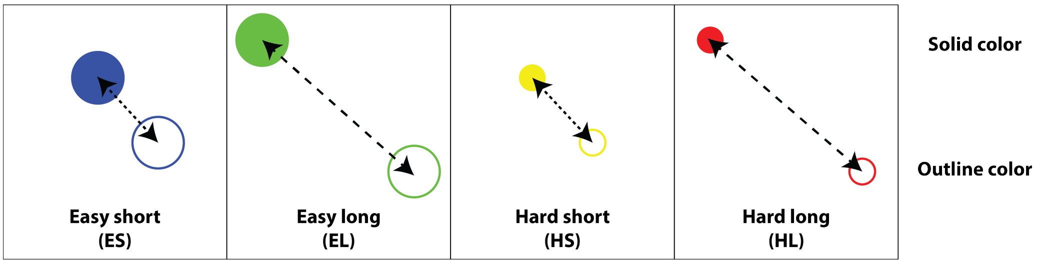


Figure 5.5: *Target classification with the example of dot shaped targets.*

5.2.1 1 degree of freedom translations

This device is composed of 3 perpendicular panels (front, bottom and left; 40 x 40 cm each), each of these presenting a linear slider (32 cm long) oriented in a different direction (x, y and z axis respectively). The first version of the slider was based on a modified drawer rail, but for ease of construction and simplicity, we preferred a mechanism based on double axis guidance from assembly machinery (see Fig. 5.6.) The user's movement is comparable to adjusting a linear slide potentiometer (e.g. for adjusting the thermostat of a toaster). The contact between wood and steel does not generate much friction, which is further reduced thanks to lubrication.

The user faces the front panel and, on each panel, manipulates a mobile wooden parallelepiped (4 x 6.8 x 3.2 cm) with a stainless steel needle (6 cm long) perpendicular to the translation axis. The targets are colored rectangular areas along the needle path. The user will have to place the mobile parts so that the needles are "inside" these areas, in terms of length (see Fig. 5.7). *Target widths* are 1.5 and 4 cm, and *target lengths* are 5 and 24 cm. On this device, the needles and targets are replicated on both sides of the translation axis for the targets to be always visible to the user, despite the occlusions from her hand.

5.2.2 1 degree of freedom rotations

This device is composed of panels similar to the 1D translation device, only each of the panels presents a rotating wooden disk (6.8 cm diameter), with a stainless steel needle

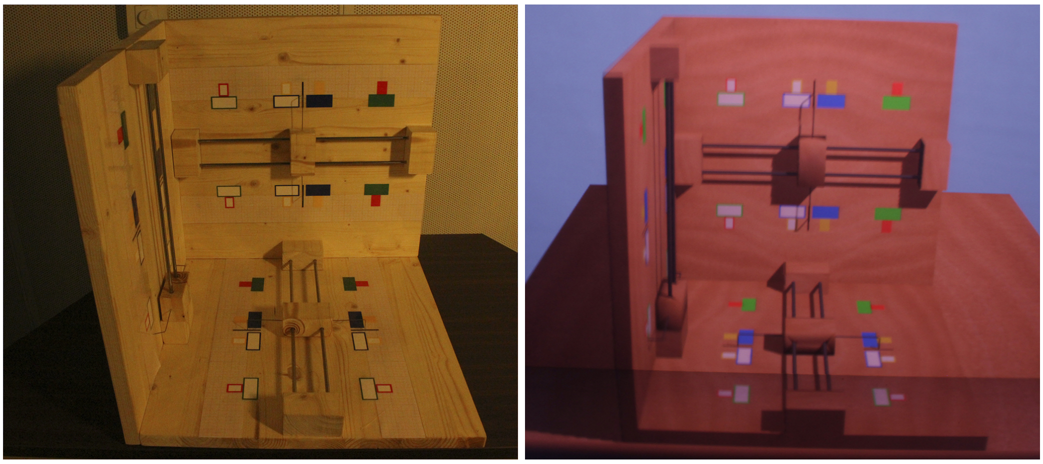


Figure 5.6: *1 degree of freedom translations. Left: physical device. Right: virtual device.*

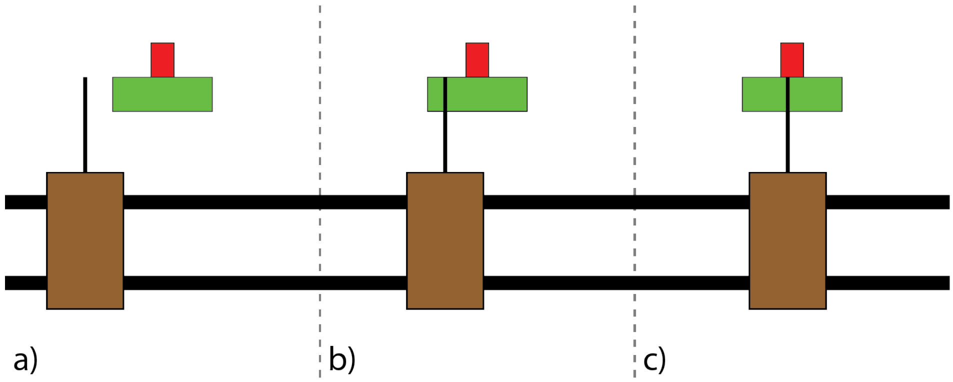


Figure 5.7: *Placement check for 1D translations. a) The placement is incorrect. b) The placement is correct for the green target only. c) The placement is correct for both the green and the red targets.*

(6.5 cm long) (see Fig. 5.8). These disks act as knobs, such as in most household appliances (e.g. tuners, ovens, dish-washers, etc.). Each knob rotates around a different axis. Friction is minimized by the use of needle-bearings between the disks and their axes. Needle-bearings are preferred here to ball-bearings as no translation is allowed.

Here again, the user faces the front panel to manipulate the knobs. This is critical when using this device: if the user faces the left panel to manipulate the left knob, she will end up in the exact same condition as when facing the front panel to manipulate the front knob. The targets are colored circular areas along the needle path. The user will have to rotate the knobs so that the needles are "inside" these areas, in terms of angle (see Fig. 5.9). *Target widths* are 10 and 40 degrees, and *target lengths* are 72 and 225 degrees. In order to respect a significant length difference between targets, and for visibility reasons, the user is asked to rotate the knob with the needle passing above its center.

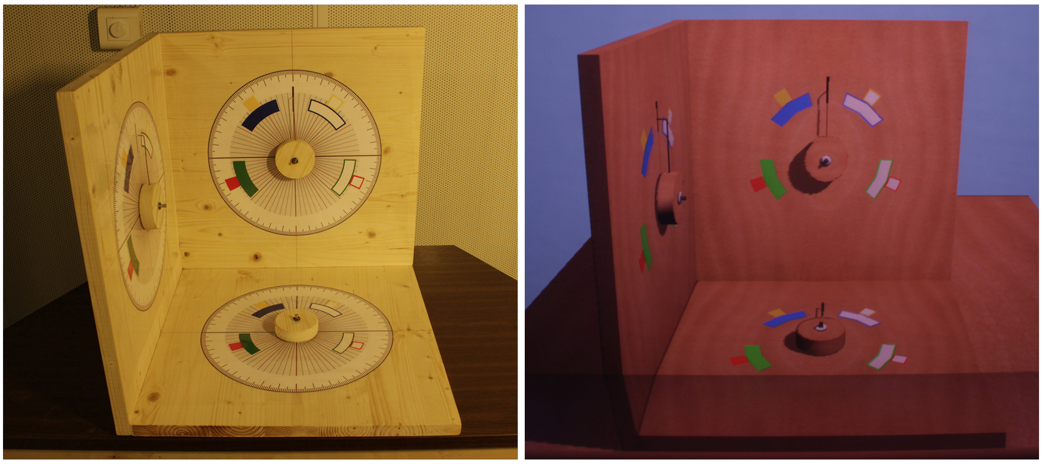


Figure 5.8: *1 degree of freedom rotations. Left: physical device. Right: virtual device.*

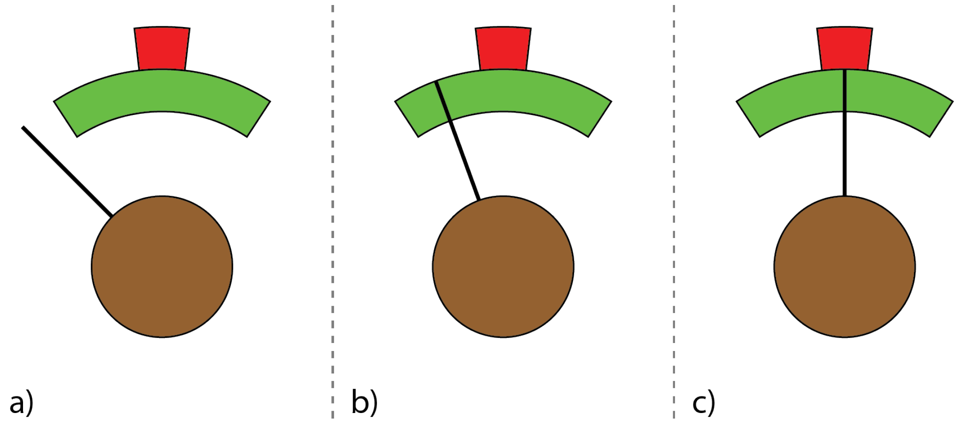


Figure 5.9: *Placement check for 1D rotations. a) The placement is incorrect. b) The placement is correct for the green target only. c) The placement is correct for both the green and the red targets.*

5.2.3 3 degree of freedom translations

With respect to our "building on" assumption (see Sec. 5.1.1), the 3D translation device is a simple extension of the 1D mechanism to 3D (see Fig. 5.10). A carter (80 x 80 x 51.6 cm) containing 4 steel axes (48 cm long) allows a frame to translate along the x axis. This frame itself holds 2 axes (44 cm long) allowing a smaller frame to move along the y axis. Finally, that smaller frame permits a cube to translate along the 2 z axis oriented steel axes it features (44 cm long). The user only manipulates the cube (6.8 cm per side), naturally involving the other objects movement.

The task, however, is more complex as the volumetric aspect of the movement implies a change in the target nature. We rejected the idea of using needles for each dimension separately as the checking would be visually scattered, which would be disturbing and

Figure 5.10: *3 degree of freedom translations. Left: physical device. Right: virtual device.*

inefficient for introducing unnecessary complexity. We wanted the user to focus on a single place to adjust her positioning. We thus chose to use the projected shadow of the mobile object onto a plane for 3D positioning. We inserted a lamp at the center of the right panel of the carter, and checked for its shadow projected onto the left panel. To account for placement tolerance in each dimension, we use a 2D tolerance which is much simpler. When at the target position, the object casts a shadow onto the left panel; the contour of this shadow is then extended to form an outer contour, and shrunk to form an inner contour. The combination of these outer and inner contours generates a valid patch. To validate a target, the user has to place the object so that its projected shadow is entirely inside this patch. This means that the shadow has to cover the whole inner contour, and cannot exceed the outer contour. In other words, the projected shadow contour must cross neither the inner contour nor the outer one (see Fig. 5.11). *Target widths* – thickness of the patch – are 2 and 5 cm, and *target lengths* – distance between the object center positions – are 13 and 31.4 cm.

5.2.4 3 degree of freedom rotations

The 3D rotation device is derived from the gyroscope (see Fig. 5.12). For ease of building, the rings are replaced with angular shapes. The central object is a cube (6.8 cm per side), rotating around a y axis oriented steel axis held by a square frame (34.4 cm per side, outer length). This frame itself rotates around the x axis, held by a U-shaped frame (45.4 cm long, 27 cm high, outer measures) so that the frame does not block against the user's arm. This last frame is held by the carter (48.2 x 48.2 x 50 cm), and rotates around the z axis. Here again, needle-bearings are used to minimize friction. We know that such a mechanism is prone to gimbal lock; however, this did not interfere with the manipulation as the user could easily rotate the cube to get out of this configuration.

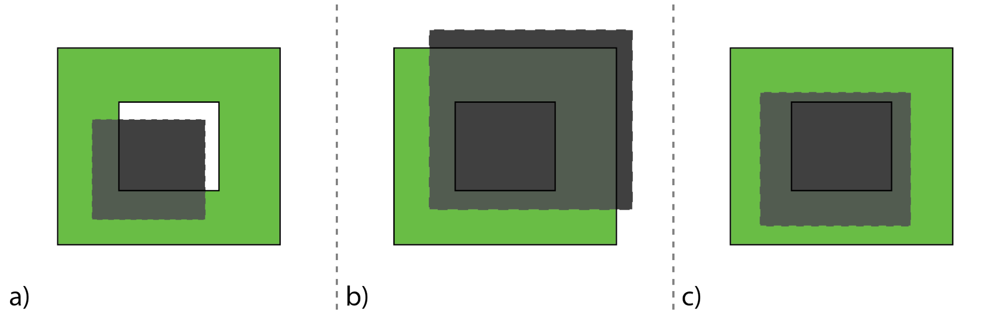


Figure 5.11: *Shadow task, placement check. a) The shadow does not completely cover the inner surface, the placement is not correct. b) The shadow exceeds the outer contour, the placement is not correct. c) The shadow contour is completely comprised between the inner and outer contours, the placement is correct.*

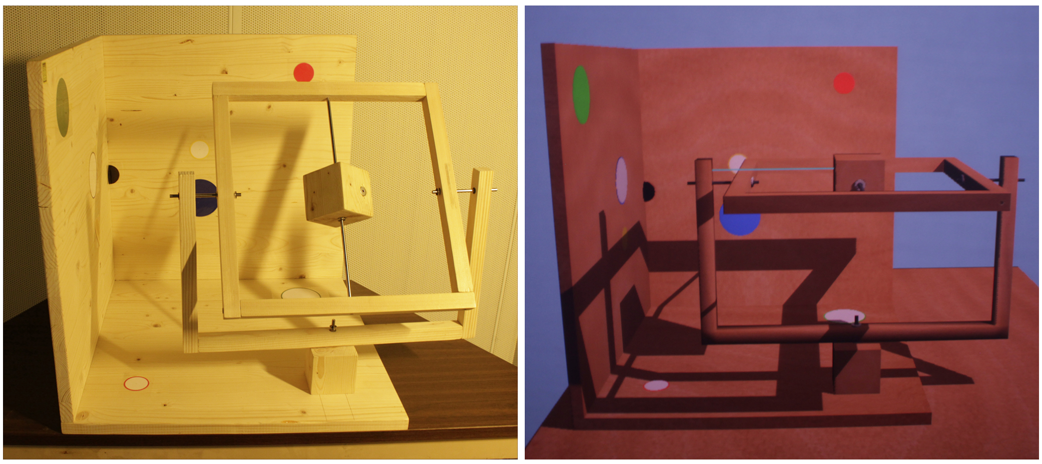


Figure 5.12: *3 degree of freedom rotations. Left: physical device. Right: virtual device.*

While the system here is more complex to comprehend than the 1D version, the task itself is very similar. The central cube is equipped with a red laser, which is automatically lit when inserted into the cube. The laser beam replaces the needle as the user has to point it inside circular targets placed onto the carter panels. *Target widths* – colored disk radius – are 2 and 4 cm, and *target lengths* – angle between the target centers, the center of rotation being the center of the cube – are 36.5 and 102.8 degrees.

5.2.5 6 degree of freedom free movement

The free movement device builds on the two 3D devices. The task is indeed a simple combination of the 3D rotation and translation tasks. No "system" is needed here as the movement is free. The shadow targets are simply displayed on a 60 x 40 cm plank featuring a 40 cm high bracket to hold the light (see Fig. 5.13). The mobile object is an anisotropic shape to disambiguate orientation possibilities; it is a sort of 3D arrow (9 x 12 x 5 cm bounding box). The box-shaped body of the arrow presents a hole to insert a laser emitter. The laser is again automatically lit when inserted.

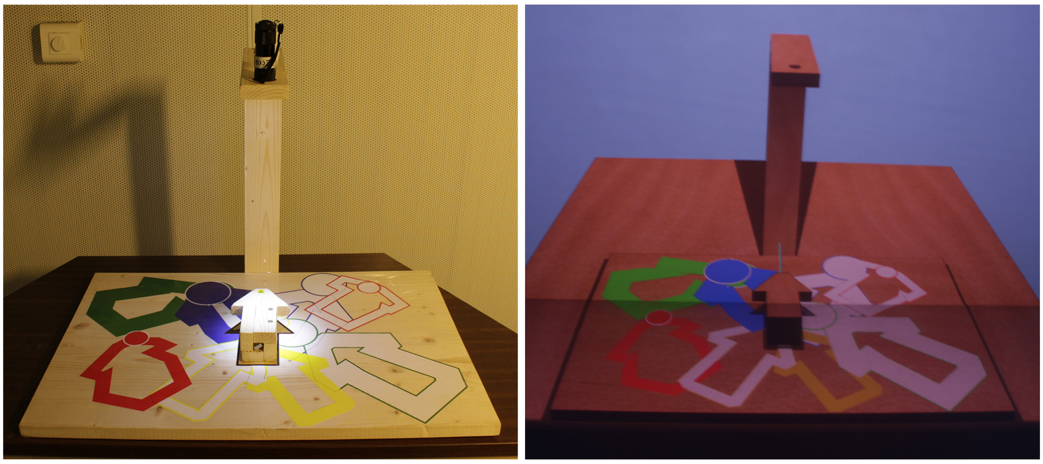


Figure 5.13: *6 degree of freedom free movement. Left: physical device. Right: virtual device.*

The targets combine a colored disk and a colored contour patch. To validate a target, the laser must be pointing inside the disk while the projected shadow of the arrow must be inside the patch, in the same way as described in Sec. 5.2.3. *Target widths* are 2 and 4 cm for the patch thickness, and 2 and 4 cm for the disk radius; *target lengths* – distance between the arrow bounding box center positions – are 9.7 and 27.8 cm.

5.3 Experimental procedure

5.3.1 Experiment proceedings

The experiment lasts 2.5 hours per participant.

The hardware used for the virtual devices is a four-sided immersive projection system with Infitec stereo and ART tracking both for the head and fingers, and featuring a surround sound system (see Sec. 3.1 for details).

5.3.1.1 General proceedings

Before starting the experiment, we perform a stereo-blindness test (see Sec. 5.3.3). The user is also informed that she can take a break anytime she needs.

The experiment is then organized into three groups of tasks, according to the input techniques. The physical set, or group of conditions for the physical devices, serves as a baseline and is always presented first, followed by the direct manipulation and wand sets. Each set includes five blocks of trials (one per device), each beginning with a practice session followed by four or twelve experimental trials. Each movement condition has four tasks, based on combinations of target size and length. For both training efficiency and logistics, we group the three 1D translation and the three 1D rotation trials together, for 12 trial types each. The 3D conditions are presented as individual blocks with 4 trial types each.

At the beginning of the physical set, participants are shown the 5 devices and, for each device, the experimenter shows how it works and explains the task.

The virtual session starts with the calibration process which provides settings for the subsequent direct manipulation and wand sets. It is composed of two steps: an eye position calibration to create a profile correcting the default projection in the immersive space (see Sec. 5.3.4), and a finger tracking device calibration to create a hand pattern specific to the user (see Sec. 3.2.1).

For each block within each set, the participant is invited to test the task in training sessions. The user performs 4 replications of the task for the easiest and hardest levels. She can repeat those until she feels comfortable with the manipulation. The experimenter observes the subject and guides her when necessary until the task is perfectly clear to the user and she is able to achieve the task fast enough alone. When the user feels confident enough, the experimenter stops the practice session and the participant performs the tasks according to the sequence defined in the protocol.

In the first training session of both virtual sets, the experimenter explains the principles of grasping and releasing and invites the participant to experiment with these until she feels comfortable.

5.3.1.2 Blocking and counter-balancing strategy

An experiment consists of 108 unique conditions organized into three sets. The physical set is always presented first, so we counter-balance the direct manipulation and wand sets for order, such that half the participants experience physical-direct manipulation-wand and half experience physical-wand-direct manipulation.

We also counter-balance for movement nature, such that half the users will start with rotations, and the other half will start with translations. We decided not to counter-balance for degrees of freedom as the complexity is thought to increase with the number of degrees of freedom. We thus preferred to let the user learn progressively as the experiment builds on previous steps to avoid complexity overload. Then, within a set, the blocks order will be: 1D rotations - 1D translations - 3D rotations - 3D translations - free movement for half of the participants, and 1D translations - 1D rotations - 3D translations - 3D rotations - free movement, for the other half.

For all tasks, we randomize the level of difficulty of the targets. For 1D tasks, we also randomize the axis of the movement. We use a Latin square to combine these randomizations for 1D tasks.

5.3.1.3 Task description

Each movement type has a neutral position. It guarantees that every participant starts with the same experimental configuration as this may influence the way they will grab the object, and then also their performance.

Each trial has a specified starting point and target position. Each trial involves moving a physical or virtual 3D object from the starting position to the specified target. Prior to each condition (i.e., a movement type associated to a target level), the object appears at a standard neutral position. The participant grasps the object, moves it to the indicated starting position, and presses the left button of a mouse held in her left hand. If the object is positioned correctly, a *correct* beep sounds and start time is recorded. The participant then moves the object to the target position and presses the button to record the end time. If the object is placed correctly, a *correct* beep sounds again, the former target acts as the starting position and the former starting position acts as the target for the subsequent trial. The participant moves the object to each successive target until reaching the final target of the condition, pressing the button each time the target is reached. Finally, the user repositions the object on the neutral position. Each successive target acquisition counts as a replication. If the object is not placed correctly on the starting position or any subsequent target, an *error* beep sounds, and the participant must reposition the object and click the button again. This records an error for the previous trial, and registers a new start time for the subsequent trial. Also, the user is not allowed to re-clutch until she reaches the final target of a given condition. If the object is released during the manipulation, the current trial has to be redone, meaning that the user has to click again on the starting position of the trial and then on its target. This records an error for the trial.

1D trials are very short and easy, so we include 6 replications for each condition. 3D trials are longer, making replications more costly in terms of time and fatigue, so we include 2 replications for each condition. For the training trials, we include 4 replications for each condition, independently of the number of degrees of freedom, and the trials for the easiest and hardest targets are replicated alternatively until the user feels comfortable with the manipulation.

For each condition, the first starting point is the solid color target. Participants are given the following instructions: *"Please go from solid color to outline color back and forth* N *times as quickly and as accurately as possible"*, replacing *N* by the actual number of replications.

5.3.2 Population

We ran the experiment with 16 participants, 12 men and 4 women. They were aged between 24 and 41 years old, with a mean age of 28.7 years, and a standard deviation of 4.9 years. Thirteen of them had no or little experience in virtual reality whereas three had a significant experience. All of them had normal vision in terms of color and stereo vision. All of them were right-handed or could easily manipulate objects with their right hand with reasonable speed and accuracy.

5.3.3 Stereo-blindness test

Before the experiment actually starts, each participant is invited to face the front screen of the immersive space where a random-dot stereogram is displayed (see Sec. 3.1.1). After the participant confirms that depth is correctly perceived, she is considered as not stereo-blind and the experiment continues. If not, the experiment is stopped.

5.3.4 Eye calibration

To calibrate the 3D position of the user's eyes in an accurate way, we decided to combine the protocol described by [Ponto 2013] with the simple test we performed in our previous study (see Sec. 4.2.2.1). They propose to use the perception of posts, in terms of position and distortion, to evaluate the parameters. Though accurate, this method implies subtle measurements which have to be perfectly explained to the users and carefully performed; the calibration will fail if this relatively complex step is unclear to the user. We thus allowed further adjustment with simple tests:

a) At the end of the protocol, when viewing a group of posts, the user is invited to point at a corner of her choice with her finger, and to take a step on the side, without physically moving her finger. She should see a gap appearing between the image of the corner and her actual fingertip, which is a direct consequence of the perceived movement of the static virtual objects. We thus correct the parameters to minimize this gap.

b) We then adjust the eye position so that the fingertips representation fits as much as possible with the user's actual fingertips. This is done within the experiment first environment to confront the user's viewing conditions to an actual meaningful scene.

5.4 Measurements

5.4.1 Objective metrics

We measure performance by recording a) the duration of a trial, and b) the error as success/fail outcome.

We enrich the data recorded in the virtual setup by storing, for each trial, the start and stop times, the quantified error or *delta* (difference between exact target position and position when validating, either in terms of angles or distances) and the fingertip position. Measuring start and stop times allows to double check the presence of errors in the collected data: for a given target, the user performs a bundle of replications in a row; thus, the end time of a replication should be the starting time of the following one.

In the physical setup, measures are extracted from video captures, using the mouse clicks on the audio track. Recording is performed using a hand-held camera so the experimenter can focus on the point of interest. To be as fair as possible with the virtual setup, the time needed by the experimenter to evaluate the validity of the placement is removed from the measures. For the analysis, we used the Atlas.ti[1] software which allows sequences of video to be annotated. Annotations can be further extracted into a structured text file, providing the name and duration of each task replication in our case.

To calculate the performance time, we take the median of the replications for each condition.

5.4.2 Subjective metrics

At the end of the experiment, users complete a 22-item cybersickness questionnaire [Viaud-Delmon 2000] evaluating the level of a group of physical sensations from 0 (no sensation) to 4 (strong sensation) for the whole virtual session.

Participants are also presented with a questionnaire evaluating the virtual techniques. For each device and for both techniques separately, they are asked to rate the following characteristics on a Likert scale from 1 to 7: similarity of the manipulation with the real experiment (1 = far from real, 7 = close to real), ease of use (1 = hard to use, 7 = easy to use), fun (1 = boring, 7 = fun), fatigue (1 = the user felt tired, 7 = the user didn't feel tired), perceived precision of the technique for selecting targets (1 = poor precision, 7 = good precision), perceived performance of the technique in terms of speed (1 = slow, 7 = fast), cybersickness (1 = the user felt sick/nauseous, 7 = the user felt no discomfort), general appreciation of the technique (1 = unsatisfied, 7 = satisfied).

Finally, they are asked if they had specific strategies for the techniques used, as well as the difficulties and advantages they identified for each technique. Participants could also

[1] http://www.atlasti.com/index.html

provide additional comments on each device and on the global experiment.

5.5 Results

5.5.1 Objective measurements

For time measurement analysis, we take the median of all replications for each movement type. Moreover, the reported confidence intervals (CI) present a confidence level of 95%[2].

5.5.1.1 1D translations

We first performed a Friedman non-parametric test to evaluate the effect of the technique on error rate. We found a significant effect of condition ($\chi^2 = 25.29$, $p < 0.001$). Post-hoc analysis revealed that the error rate was lower for the physical device (median = 0.0%, stddev = 1.0%) compared to wand (median = 6.3%, stddev = 6.4%) and direct manipulation (median = 5.6%, stddev = 4.3%).

For time completion analysis, we conducted a 4-way (i.e., Technique x Width x Distance x Axis) ANOVA with repeated measures. The main effects of Technique ($F_{2,30} = 61.32$, $p < 0.001$), Width ($F_{1,15} = 88.95$, $p < 0.001$), Distance ($F_{1,15} = 135.11$, $p < 0.001$) and Axis ($F_{2,30} = 6.78$, $p = 0.004$) on Time were statistically significant. Overall, the use of the physical device is faster than the wand (mean diff. = 0.507 sec, $CI = [0.340, 0.675]$) and than the direct manipulation (mean diff. = 0.634 sec, $CI = [0.441, 0.826]$); the wand is also significantly faster than the direct manipulation (mean diff. = 0.127 sec, $CI = [0.005, 0.248]$). With further analysis, we also found a significant interaction between Technique and Width ($F_{2,30} = 39.43$, $p < 0.001$), Technique and Distance ($F_{2,30} = 33.81$, $p < 0.001$) and Technique and Axis[3] ($F_{2.4,36.4} = 6.85$, $p = 0.002$). We further observe that using the physical device was significantly faster than wand and direct manipulation for each difficulty level. Using the wand is also systematically faster than the direct manipulation; however, the difference was only significant for the extreme levels of difficulty (i.e., *easy-short* and *hard-long*).

Performance results can be observed in Fig. 5.14.

In [Fitts 1954], the author performs some pointing task experiments and derives from his results a relation between speed, amplitude and tolerance in perceptual-motor activities, also known as Fitts's law. The movement mean time (MT) is a linear function of the task index of difficulty (ID):

$$MT = a + b.ID \tag{5.1}$$

[2]For post-hoc pairwise comparisons, CIs have been adjusted for multiple comparisons by using Bonferroni's adjustment

[3]When sphericity is violated, degrees of freedom have been corrected by using Greenhouse-Geisser correction

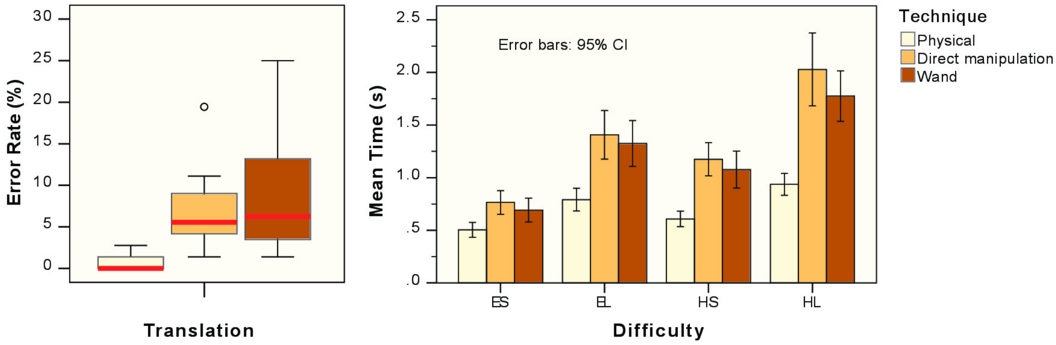


Figure 5.14: *Performance results for the 1D translation task. Left: error rate. Right: completion time.*

The index of difficulty is defined by a relation linking the target distance D and the target width W:

$$ID = log_2(1 + \frac{D}{W}) \tag{5.2}$$

Considering that our 1D translation task ends up being a constrained pointing task, and given the statistical significance of the effects of the target width and length on time, we plotted the mean time as a function of the target index of difficulty (see Fig. 5.15).

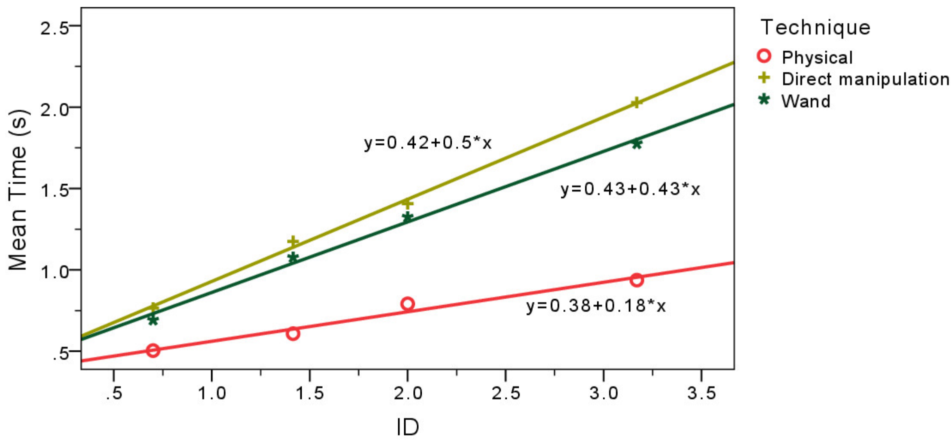


Figure 5.15: *As predicted by Fitts's law, completion time is proportional to the index of difficulty.*

When estimating the parameters, we obtain the table Tab. 5.2.

R^2 values show that our design closely follows Fitts's model. We can also observe that, when using virtual techniques, completion time increases faster with the level of difficulty than with the physical setup; whereas both virtual techniques are relatively close.

	a: 95% CI (sec)	b: 95% CI (sec)	R^2
Physical	[0.177, 0.583]	[0.082, 0.281]	0.968
Direct manipulation	[0.262, 0.587]	[0.425, 0.585]	0.997
Wand	[0.201, 0.655]	[0.322, 0.545]	0.993

Table 5.2: *Estimation of Fitts's model parameters for each input technique (linear regression)*

5.5.1.2 1D rotations

We first performed a Friedman non-parametric test to evaluate the effect of the technique on error rate. We found a significant effect of condition ($\chi^2 = 24.03$, $p < 0.001$). Post-hoc analysis revealed here also that the error rate was lower for the physical device (median = 0.7%, stddev = 1.0%) compared to wand (median = 6.9%, stddev = 5.3%) and direct manipulation (median = 9.0%, stddev = 7.0%).

For time completion analysis, we conducted a 4-way (i.e., Technique x Width x Distance x Axis) ANOVA with repeated measures. The main effects of Technique ($F_{1.2,18.5} = 83.31$, $p < 0.001$), Width ($F_{1,15} = 97.53$, $p < 0.001$), Distance ($F_{1,15} = 386.96$, $p < 0.001$) and Axis ($F_{2,30} = 33.04$, $p < 0.001$) on Time were statistically significant. Overall, the use of the physical device is faster than the wand (mean diff. = 0.756 sec, $CI = [0.602, 0.910]$) and than the direct manipulation (mean diff. = 1.105 sec, $CI = [0.793, 1.418]$); here again, the wand is significantly faster than the direct manipulation (mean diff. = 0.349 sec, $CI = [0.136, 0.562]$). With further analysis, we also found a significant interaction between Technique and Width ($F_{1.4,21.3} = 36.59$, $p < 0.001$), Technique and Distance ($F_{1.4,21.1} = 57.00$, $p < 0.001$) and Technique and Axis ($F_{2.4,36.9} = 4.16$, $p = 0.017$). We further observe that using the physical device was significantly faster than wand and direct manipulation, and the wand was significantly faster than direct manipulation, for each difficulty level.

Performance results can be observed in Fig. 5.16.

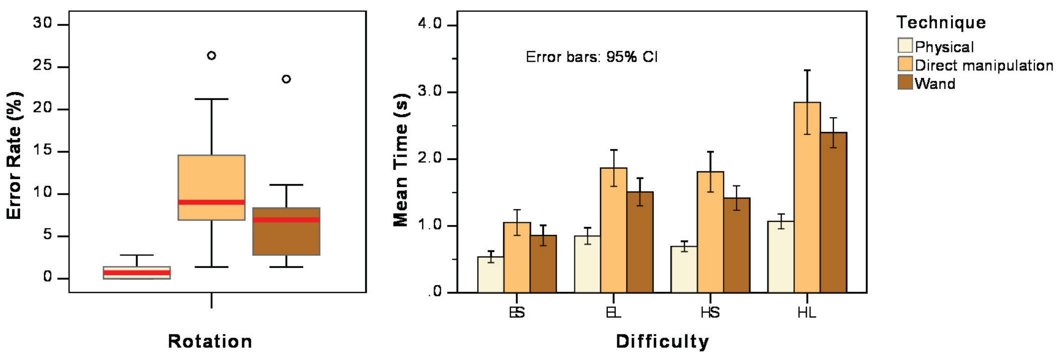


Figure 5.16: *Performance results for the 1D rotation task. Left: error rate. Right: completion time.*

Fitts's experiments only consider linear pointing tasks. However, similarly to [Stoelen 2010] and [Nguyen 2014], we were convinced that the principle should be similar for circular tasks, and the statistical significance of the effects of the target width and length on time confirmed this idea. We thus plotted the mean time as a function of the target index of difficulty (see Fig. 5.17).

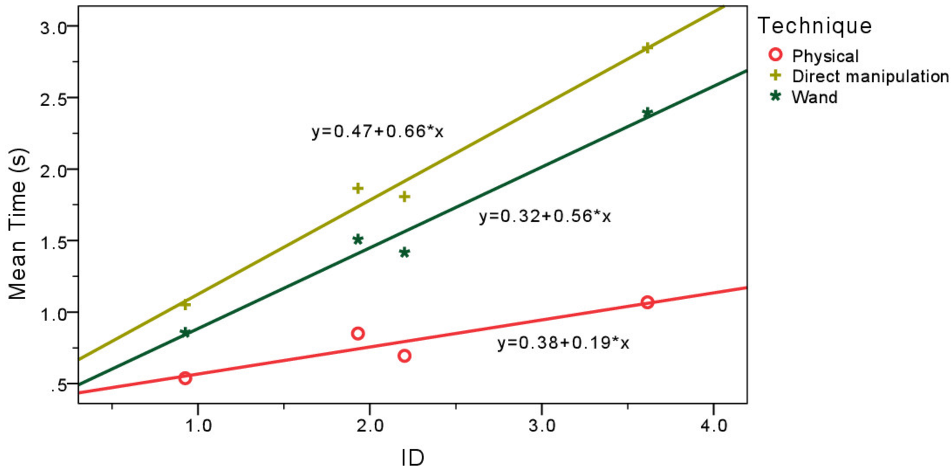


Figure 5.17: *1D rotations also follow the model described by Fitts: completion time is proportional to the index of difficulty.*

When estimating the parameters, we obtain the table Tab. 5.3.

	***a*: 95% CI (sec)**	***b*: 95% CI (sec)**	R^2
Physical	[-0.176, 0.932]	[-0.044, 0.423]	0.859
Direct manipulation	[-0.171, 1.104]	[0.389, 0.927]	0.982
Wand	[-0.354, 0.992]	[0.281, 0.849]	0.973

Table 5.3: *Estimation of Fitts' model parameters for each input technique (linear regression)*

R^2 values, even if reporting less correlation between time and difficulty than for the translation task, show that the design closely follows Fitts' model. Here again, we observe that completion time increases faster with the level of difficulty within the virtual setup than within the physical setup; whereas both virtual techniques are relatively close.

5.5.1.3 3D translations

We first performed a Friedman non-parametric test to evaluate the effect of the technique on error rate. We found a significant effect of condition ($\chi^2 = 7.73$, $p = 0.021$). Post-hoc analysis revealed here also that the error rate was lower for the physical device (median = 25.0%, stddev = 15.8%) compared to wand (median = 37.5%, stddev = 14.4%) and direct manipulation (median = 37.5%, stddev = 17.8%).

For time completion analysis, we conducted a 3-way (i.e., Technique x Width x Distance) ANOVA with repeated measures. The main effects of Technique ($F_{1.2,17.7} = 20.25$, $p < 0.001$), Width ($F_{1,15} = 42.46$, $p < 0.001$), and Distance ($F_{1,15} = 37.60$, $p < 0.001$) on Time were statistically significant. Overall, the use of the physical device is slower than the wand (mean diff. = -8.35 sec, $CI = [-13.54, -3.16]$) and than the direct manipulation (mean diff. = -10.96 sec, $CI = [-17.16, -4.76]$); the wand is significantly slower than the direct manipulation (mean diff. = -2.61 sec, $CI = [-4.86, -0.36]$). With further analysis, we also found a significant interaction between Technique and Distance ($F_{2,30} = 13.75$, $p < 0.001$), but not between Technique and Width ($F_{1.4,21.0} = 0.60$, $p = 0.50$). We further observe that using the physical device was significantly slower than both virtual techniques for large distance tasks.

Performance results can be observed in Fig. 5.18.

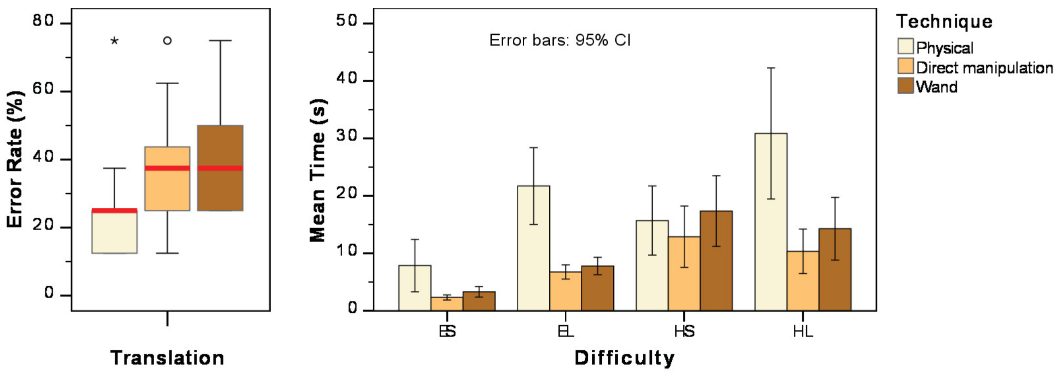


Figure 5.18: *Performance results for the 3D translation task. Left: error rate. Right: completion time.*

5.5.1.4 3D rotations

We first performed a Friedman non-parametric test to evaluate the effect of the technique on error rate. We found a significant effect of condition ($\chi^2 = 19.00$, $p < 0.001$). Post-hoc analysis revealed here also that the error rate was lower for the physical device (median = 0.0%, stddev = 3.1%) compared to wand (median = 12.5%, stddev = 13.3%) and direct manipulation (median = 25.0%, stddev = 12.5%).

For time completion analysis, we conducted a 3-way (i.e., Technique x Width x Distance) ANOVA with repeated measures. The main effect of Technique ($F_{2,30} = 33.75$, $p < 0.001$) on Time was statistically significant. However, the effects of Width ($F_{1,15} = 4.45$, $p = 0.052$), and Distance ($F_{1,15} = 0.19$, $p = 0.67$) on Time were not significant. Overall, the use of the physical device is faster than the wand (mean diff. = 2.41 sec, $CI = [1.28, 3.53]$) and than the direct manipulation (mean diff. = 4.27 sec, $CI = [2.78, 5.76]$); the wand is significantly faster than the direct manipulation (mean diff. = 1.86 sec, $CI = [0.30, 3.42]$).

Performance results can be observed in Fig. 5.19.

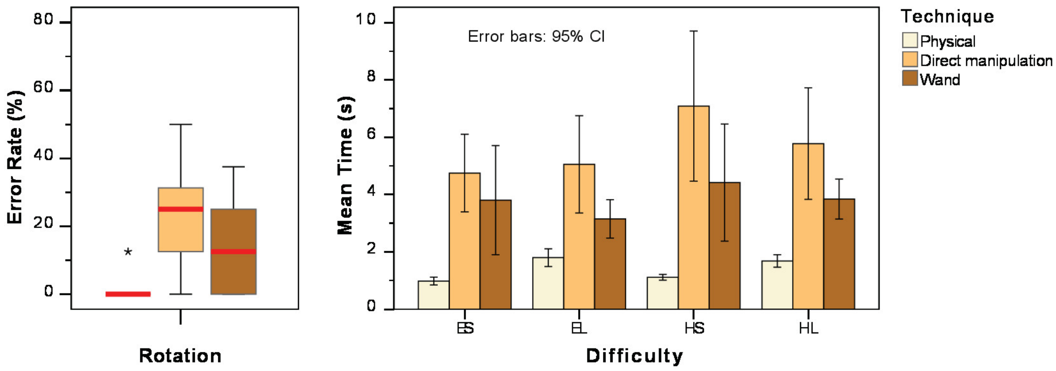


Figure 5.19: *Performance results for the 3D rotation task. Left: error rate. Right: completion time.*

5.5.1.5 Free movement

We first performed a Friedman non-parametric test to evaluate the effect of the technique on error rate . We found a significant effect of condition ($\chi^2 = 11.38$, $p = 0.003$). Post-hoc analysis revealed here also that the error rate was lower for the physical device (median = 12.5%, stddev = 12.8%) compared to wand (median = 37.5%, stddev = 17.1%) and direct manipulation (median = 37.5%, stddev = 12.8%).

For time completion analysis, we conducted a 3-way (i.e., Technique x Width x Distance) ANOVA with repeated measures. The main effects of Width ($F_{1,15} = 52.04$, $p < 0.001$), and Distance ($F_{1,15} = 31.51$, $p < 0.001$) on Time were statistically significant. However, the effect of Technique ($F_{2,30} = 1.98$, $p = 0.16$) on time was not significant. With further analysis, we also found a significant interaction between Technique and Distance ($F_{2,30} = 6.10$, $p = 0.006$), but not between Technique and Width ($F_{2,30} = 1.26$, $p = 0.30$).

Performance results can be observed in Fig. 5.20.

5.5.2 Subjective measurements

The results from the subjective questionnaire are summarized in Fig. 5.21. The questions *Q1* to *Q8* are those presented in Sec. 5.4.2, in the same order, such as Q1 = closeness to real, Q2 = ease of use, Q3 = fun, Q4 = fatigue, Q5 = precision, Q6 = speed, Q7 = cybersickness and Q8 = overall appreciation. We can observe that participants did not express any clear preference for any of the two virtual techniques. However, as expected, they found that direct manipulation was closer to real than the wand, and a Wilcoxon Signed Ranks Test

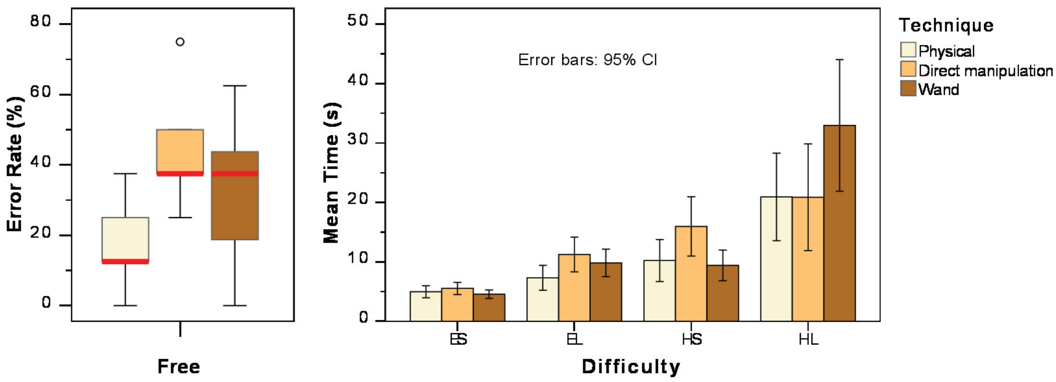


Figure 5.20: *Performance results for the free movement task. Left: error rate. Right: completion time.*

shows that this difference was statistically significant ($Z = -2.67$, $p = 0.008$). Also, with both virtual techniques, our design did not generate cybersickness.

5.6 Discussion

Our goal was to design devices reproducible in virtual and in real settings that allow us to carefully decompose and study 3D movements. The reason for this is that we believed that it is important to modify the complexity of the task motions in a controlled and progressive manner. We hypothesized that increasing complexity results in increased difficulty, due to the difficulty of anticipating the required motion which is constrained. We were interested in studying the impact of naturalness on this effect, and expected that there would be an advantage for direct manipulation.

Our hypotheses were thus that the wand would perform better for translation-based motions, but it would be outperformed by direct manipulation for rotations a) due to the cumbersomeness and occlusions of the wand and b) because it requires more anticipation (e.g., for 1D rotations, how to grab the object to be able to perform the complete required rotation without releasing the knob).

Our results show that the wand performed better than direct manipulation in terms of speed, but they were equivalent in terms of user appreciation. This does not confirm our hypothesis about rotations. We believe the main reason for this is tracking noise. As we can see in Sec. 3.2.2, the hand tracking signal is prone to trembling and jumps due to the limited number of markers and to occlusions, from the user herself and from the surrounding screens. This issue should disappear with technological progress. Another reason, as reported by the participants, is that the hand generated more occlusions than the wand as, in the latter case, users manipulate from a distance, the object being held by a thin virtual ray. We can thus suppose that visibility has more impact than cumbersomeness on

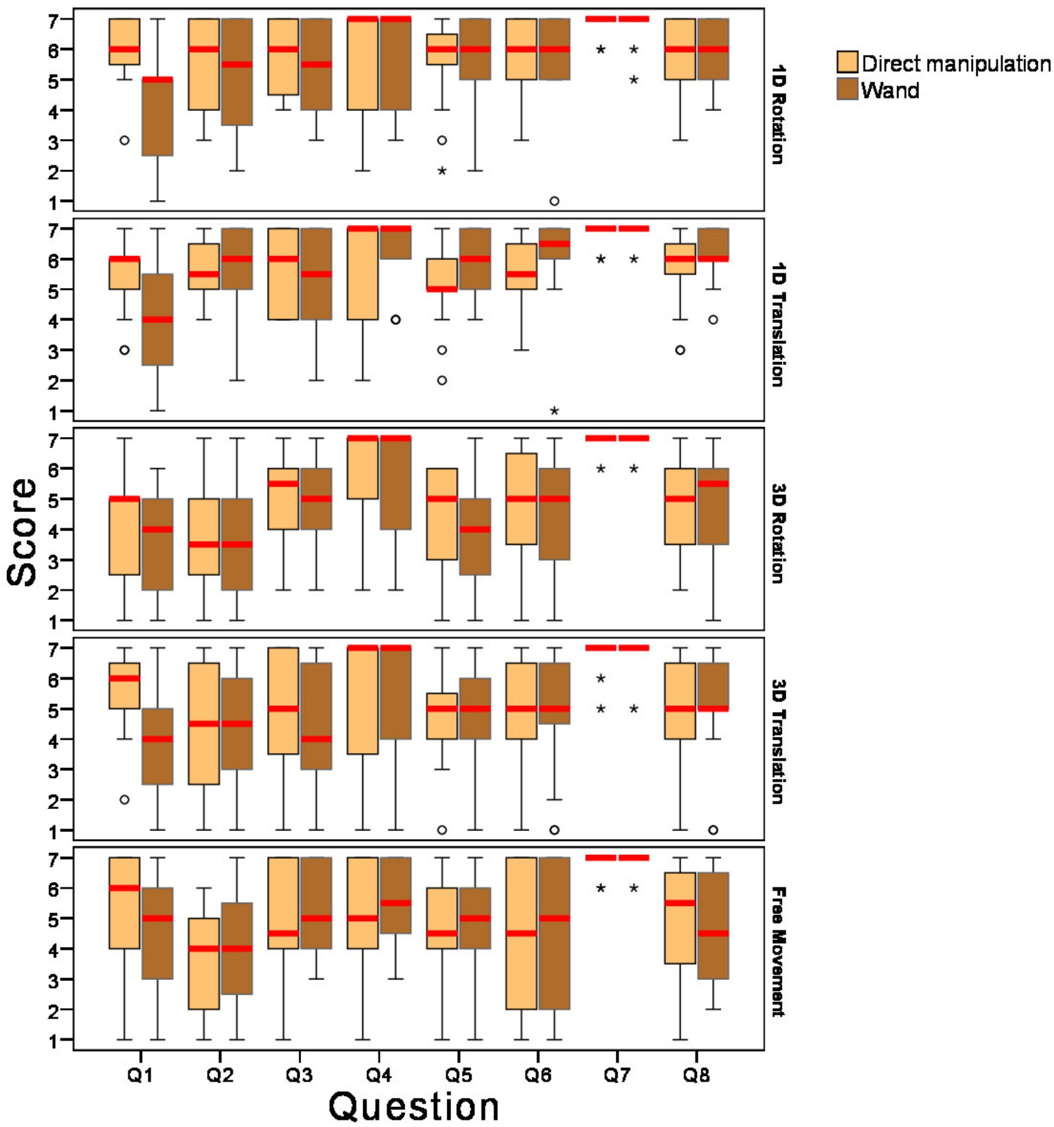


Figure 5.21: *Subjective responses to the questionnaire.*

performance.

Also, we observed that users systematically perform faster with the physical setup than with the virtual techniques. This confirms our previous observation that, technologically, we are still not at the point where virtual tasks can be performed at the same speed as their real equivalent, even for simple tasks.

The only exception to these observations is the case of 3D translations. For this task, direct manipulation performed slightly faster than the wand and the real setting. This is due to the loss of the hyper-natural advantage in this particular context. Indeed, for 1D translations, as the mobile object motion was constrained, users could translate the wooden piece by a simple rotation of the wrist, when they actually had to carry out the translation with the hands. For 1D and 3D rotations, the hyper-natural interface generated less occlusions while visibility was more crucial than for translations. On the contrary, for 3D translations, occlusions were not a problem due to the targets being away from the objects, but participants did not intuitively rotate the wrist and instead, followed the path of the cube. The movement was thus exactly the same as with hands, but with the cumbersomeness of an extra device. When switching from 1D to 3D, accurate placement becomes less intuitive. Also, for this task, both virtual techniques were significantly faster than the physical setup. This can be explained by the complexity of the physical device, as described in Sec. 5.1.2.1. This is a typical example of the compromises we had to make to accommodate the decomposition of a complex motion into simple components, and the mechanical consequences of these choices. The need to decompose translations into their three x, y and z components results in a set of three rails which involve weight and friction which does affect performance.

Another interesting result is the comparison with Fitts's model for 1D tasks. We can observe that the estimated parameters for translation and rotation are similar, leading to the conclusion that Fitts's model is valid for linear as well as angular distances, coherently with previous studies such as [Stoelen 2010] and [Nguyen 2014]. Moreover, the estimated linear correlation coefficient (R^2) proved the validity of our design for the three input techniques. The results confirmed that users perform better in the physical setup than in both virtual ones as the increase in completion time with task difficulty was lower. It also confirmed that the wand interface performed better than direct manipulation. Due to the complexity of the 3D tasks, we have not addressed the possibility of using Fitts's model in their design. Indeed, in Fitts's experiments, the tolerance (target width) is in the same dimension as the movement. However, our choice of using a 2D projection contour width as a tolerance when 3D translations are involved implies that the tolerance is not the same in the three dimensions. Also, such a task using shadow contours appeared to be quite complex as we observed a significant increase in the number of errors. A simpler evaluation method should thus be preferred.

The implemented devices allow us to evaluate the users' apprehension of movement decomposition. Designing such devices presents difficulties in terms of construction as it requires compromises between visibility, weight and robustness, and in terms of target placement due to the opposing constraints of varying difficulty levels while keeping the device dimensions in the range of human limb motions.

Our results show that the wand performs better than direct manipulation, but direct manipulation should be preferred when the sensation of closeness to reality is an important factor, such as in rehabilitation systems. We still believe that progress in finger tracking technologies will help direct manipulation interfaces get closer to reality in terms of capacities. As for now, an interesting direction is to study if a fair comparison between existing technologies would be valid. In our implementation, we added filtering to the wand to get the same latency for both interaction devices, even if the wand did not present a noisy signal. A fair comparison would consist in synthesizing the finger tracking noise pattern to apply it to the wand signal, and then filter both signals similarly. On one side, we would bring the performance of one interface to the level of the second, only considering its inherent usability rather than its technological quality; on the other hand, degrading an interface is not that relevant as no user will degrade a tool before using it.

5.7 Conclusion and future work

In this chapter, we have proposed a complete framework for the analysis of 3D movement through the design of a set of devices allowing the decomposition of 3D motion into movement type (rotation and translation) and single or multiple degrees of freedom (1D and 3D tasks). The devices are reproducible in real and in virtual setups. Our study clearly demonstrates that wand-based interfaces outperform direct manipulation in terms of performance, and confirms the naturalness of the experience while using finger tracking devices. Our design is based on simple and intuitive mechanisms, permitting the participants to only focus on completing the task.

Automating the measurements from physical devices would greatly improve the experiment as it would provide richer information about the natural performance, and thus more in-depth comparisons. In terms of concept, an interesting direction for future work is the design of a "super-device" combining the possibilities of our individual machines. Isolation of movement instinctively made us design separate devices, but this is not an absolute requirement. Such a device would provide an equivalent manipulation space for each movement type. We would also want to free ourselves from the shadow tasks as a) the tolerance is not equivalent in all 3 dimensions and b) it is prone to visibility issues as the hands and other parts of the mechanism also cast shadows. An early idea could be to combine the 3D task devices. A central sphere would incorporate 3 lasers of different color, each oriented in a different axis, thus creating a frame. The sphere would be held at the center of a gyroscope-based system, which would itself be the mobile piece of a combined slider mechanism for 3D translations. Locking/unlocking mechanisms would allow the enabled degrees of freedom to be switched, even allowing 2D tasks. They would need to be carefully designed. All the measurements would result in reading laser dot positions. Albeit interesting from a mechanics point of view, a haptic device could also simulate such a behavior with higher measurement accuracy. The scalable SPIDAR [Buoguila 2001], for example, would allow such movements while preserving an important magnitude, but the presence of cables would hinder complete hand rotations. However, the mechanical counterpart would be an economical and interesting alternative.

CHAPTER 6

Using interactive virtual reality for reminiscence therapy: a feasibility study

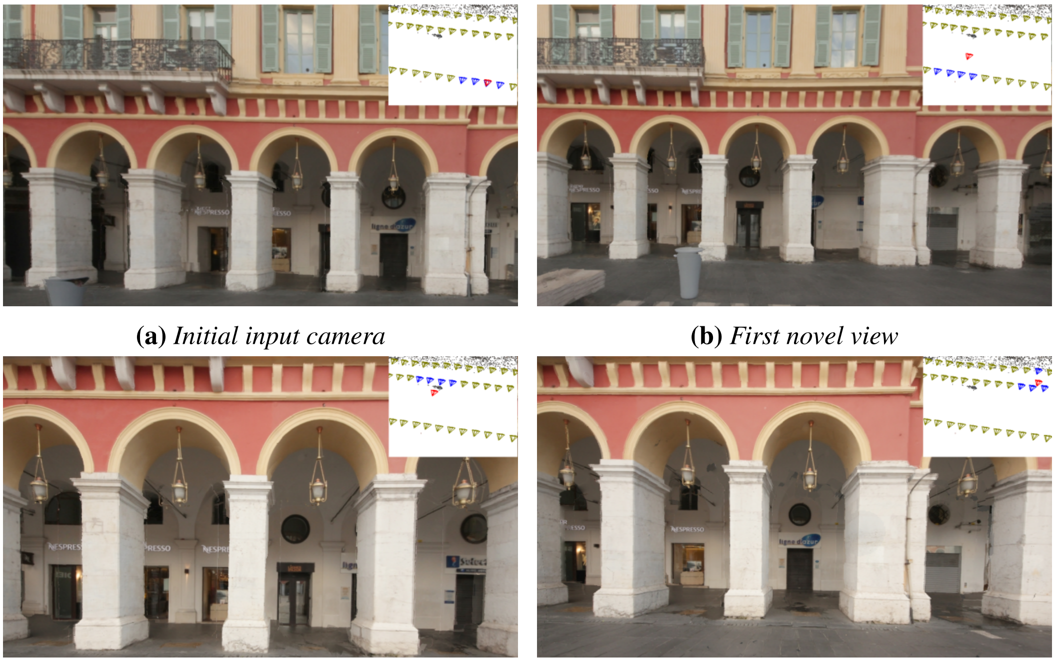
In this chapter, we will present an immersive virtual reality solution designed for reminiscence therapy, developed in collaboration with a clinical research group specializing in memory treatment and Alzheimer's disease[1]. Our system allows easy presentation of familiar environments thanks to highly-realistic image-based rendering, and supports close-to-natural gesture-based interaction with objects in the environment. To evaluate the effectiveness and utility of our system for reminiscence therapy, we perform a study with healthy elderly participants to test if our system can help with the generation of autobiographical memories. We adapt a verbal autobiographical fluency protocol to our virtual reality context, in which elderly participants are asked to generate memories based on images they are shown. We refer to the system as *IVIRAGE* – Image-based VIrtual ReAlity with GEstures – from now on.

Figure 6.1: *Left: A user in an image-based familiar environment, manipulating an apple with her right hand. Right: Our hardware setup, adapted for elderly people.*

Reminiscence therapy is an important intervention for dementia care. In traditional clinical settings, it involves the discussion of past activities, events or experiences with another person or group of people, usually with the aid of tangible props such as photographs,

[1] http://www.cmrr-nice.fr/?p=en-cobtek-presentation

household and other *familiar* items from the past, such as music or archive sound recordings. Autobiographical memory is one of the underlying processes of reminiscence therapy, and an essential component of the human memory. [Conway 2005] defined it as the mental representation of events from one's past (i.e., *episodic* autobiographical memory) together with semantic information about the self. There is evidence of some improvement of functional ability in people with dementia using reminiscence therapy [Woods 2005]. This traditional treatment is one of the most popular psychosocial interventions in dementia care, and is highly rated by staff and participants. It is thus interesting to develop and evaluate ways to improve reminiscence therapy using new technologies. In this work, we focused on *immersive virtual reality*. Our setup is a simplified version of that described in Sec. 3.1 (see Sec. 6.2 for details).



(a) *Initial input camera* **(b)** *First novel view*

(c) *Second novel view* **(d)** *Third novel view*

Figure 6.2: *(a) Image-based rendering with novel camera coincident with the input camera (red in inset); the result is identical to the input image. The user can move around freely in the area covered by the cameras. In (b)-(d) the method of [Chaurasia 2013] generates plausible novel views (red in inset) using the 4 views in blue.*

There has been promising previous work in virtual reality for memory treatments (e.g., [Brooks 2003]) and evaluation (e.g., [Gonneaud 2012]), including recent evidence that, for some cases, *interacting* with the environment can be beneficial for memory treatment and assessment [Plancher 2013]. In contrast to previous work however, virtual reality support for reminiscence therapy requires the presentation of environments which are particularly *familiar* to the patient. Creating a realistic 3D model (e.g., of a patient's neighborhood or landmarks of her city) is far too expensive using traditional manual modeling. In contrast, image-based modeling and rendering [Shum 2007] provide a much easier way to achieve this goal, since only a small set of photographs is required as input. It is however only very recently that image-based rendering has achieved a sufficient level of quality and efficiency to be suitable for free-viewpoint navigation [Chaurasia 2013], and thus for virtual reality. The high level of realism of [Chaurasia 2013] can be seen in Fig. 6.2. In parallel, the advancement of finger tracking technologies and physics simulations now make it possible to provide intuitive, close-to-natural interaction with the virtual environment as we have seen in the previous chapters.

In this study we investigate whether our immersive virtual reality system is useful in stimulating autobiographical memories in healthy elderly adults. To ensure that the virtual reality experience we propose conveys the sense of familiarity, we introduce highly realistic image-based virtual reality. Using this approach, virtual environments are constructed by simply taking a few photographs of the scenes; this allows easy access to highly realistic representations of *familiar* and personalized environments. Our system also supports close-to-natural interaction with objects in the environment by combining finger tracking, gestures and physics simulation, providing a high level of interactivity and thus action. Interactive image-based rendering can then be performed, permitting free-viewpoint navigation in the environment. Navigating in a highly realistic environment and interacting with objects opens up new possibilities for reminiscence therapy. Clinicians adapted the verbal autobiographical fluency protocol commonly used in reminiscence therapy [Dritschel 1992] to our interactive virtual reality setting. In this test, healthy elderly participants are asked to generate memories based on the images they see.

We designed and ran an experiment in which we compare the number of autobiographical fluency responses obtained in an unknown image-based virtual environment to those obtained in a familiar image-based virtual environment, to determine whether the highly realistic environments generated by our approach are powerful enough to convey a sense of familiarity to participants, and thus increase the number of generated memories. In particular, realism needs to be sufficient so that the sense of Presence and notably Place Illusion [Slater 2009b] are strong. We also investigate whether our system is as effective in generating memories as current standard clinical practice by comparing to a gray-screen control condition and a static photograph of a familiar environment which serves as a baseline, since photographs of familiar locations are used extensively in traditional reminiscence therapy. We test whether *IVIRAGE* can generate a similar number of memories as the standard approach. However, there is a significant distinction between the two: to allow in-place navigation and interactive manipulation of objects, the photographs used for the virtual environments have a narrow field-of-view to provide a scale 1 immersive experience. In contrast, photographs used in traditional autobiographical fluency usually have large field-of-views (e.g., a panoramic view), with the potential to elicit many memories. This gives an advantage to the traditional reminiscence therapy condition compared to the *IVIRAGE* condition with the familiar environment; we thus expect the number of memories generated by the familiar virtual environment to be less than or at best equal to the familiar photo. However, we hypothesize that providing the ability to interact with objects in the scene and navigate in the highly realistic environment will be effective for reminiscence therapy. We also hypothesize that the immersion in the familiar scene will result in a higher number of responses, which will indicate that *IVIRAGE* can convey familiarity.

To be successful, we need to satisfy two fundamental requirements. First, we need to provide a working system that allows navigation using image-based rendering and gestures in an immersive setting and second, we need to demonstrate that our system and protocols are well tolerated by elderly participants and can be as effective as traditional reminiscence therapy. In the following, we present our system that satisfies the first requirement. In the results, we present data from our experiment demonstrating that our system and protocol are well tolerated by elderly participants.

6.1 Specific previous work

We first provide some additional background on autobiographical memory, and then briefly discuss related work on closely-related image-based rendering. See Sec. 2.3 for a review of previous work on rehabilitation and virtual reality, and in particular work concerning memory and the elderly.

6.1.1 Autobiographical memory

The retrieval of autobiographical memory is a complex process, involving mental traces of past events and related semantic knowledge. These differ e.g., in content, imagery or emotional intensity. A fundamental dimension of an autobiographical memory is its level of detail, which is thought to be a function of a person's "database" of memory traces, reconstructive processes, mental schemata, self goals and different retrieval strategies [Conway 2000].

Episodic autobiographical memory, especially the richness of details, is impaired early in the course of Alzheimer's disease or even in the preclinical phase, while semantic memories are spared until moderate stages, indicating a dissociation between both memory systems [Berna 2012, Seidl 2011]. Autobiographical memory is used to develop a coherent sense of self, emotions and future plans [Bluck 2003]. It thus has a rich role in motivating behavior and in daily life. Many different methods have been used to study autobiographical memory [Griffith 2012]; one approach is to stimulate memories using word, visual or other sensorial cues.

We concentrate on verbal autobiographical fluency tests [Dritschel 1992]. Such tests are widely used to assess language abilities and executive functioning [Miller 1984]. These tasks involve rapid associative exploration and retrieval of words based on phonemic (e.g., starting with a specified letter of the alphabet) or semantic/categorical (e.g., animals) criteria over a brief timed interval.

6.1.2 Image-Based Rendering

View-dependent texture mapping [Debevec 1996] and the unstructured lumigraph [Buehler 2001] are early examples of image-based rendering methods which allow interactive viewing. However, both suffer from disturbing visual artifacts if the reconstructed 3D geometry is not perfect. There have been many more recent methods which improve various aspects of image-based rendering (see [Shum 2007] for an overview). However, true interactivity with quality at a level acceptable for virtual reality has remained elusive, despite advances in the last few years [Eisemann 2008, Goesele 2010].

Most image-based rendering systems can be classified into two types: proxy based [Buehler 2001, Eisemann 2008] and image warping approaches [Stich 2011, Chaurasia 2011, Chaurasia 2013]. The former apply a view-dependent texture on the 3D model of the scene while the latter warp input images and blend the warped images to synthesize the final result. Proxy-based approaches require an accurate 3D model which is essentially impossible to obtain using state of the art 3D reconstruction [Furukawa 2009].

The recent work on depth synthesis and warp-based rendering [Chaurasia 2013], does provide rendering quality which is acceptable for some level of virtual reality display. The quality of the rendered images – given specific restrictions which we develop below – is definitely acceptable for the type of immersive virtual reality we will use in this study, despite some residual visual artifacts. The resulting realism is high, as can be seen in Fig. 6.2, even in views away from the input photographs used to reconstruct the scene. We summarize this algorithm in Sec. 6.3 since it is central to our study.

6.2 Experimental design

The goals for our interactive virtual reality system for reminiscence therapy are thus:

- to provide easy capture of familiar and personalized environments, while allowing realistic interactive display and navigation;
- to allow intuitive interaction with virtual objects within the image-based virtual environments, using gestures, based on finger tracking.

6.2.1 Limitation to a single screen

The algorithm of [Chaurasia 2013] which we use in our system (see Sec. 6.3) is limited to single screen displays for two main reasons.

First, at every frame, the four input views “closest" to the current view are chosen, and then their “superpixels” warped. The algorithm can be used independently to synthesize a different view for each of the four screens of the immersive cube; however, it is important to note that the subset of input images used to synthesize each screen will be different because the side and bottom screens have orientations perpendicular to that of the front screen. This fundamentally means that there can be discontinuities across screens. Although not proven, this should be less of a problem with proxy-based image-based rendering [Buehler 2001, Eisemann 2008] because at least the same proxy is used for all the screens. We however used [Chaurasia 2013] because the rendering quality of proxy-based approaches is expected to be below the bare minimum required to conduct these experiments. This problem is further complicated for the bottom screen because of its downwards orientation. Studies [Vangorp 2013] reveal that large angles between the orientations of input camera and novel camera can lead to severe perspective distortions. Thus, the bottom screen can be effectively rendered only using input images that are pointing downwards. This would lead to very cumbersome capture using specialized camera rigs.

Second, the shape-preserving warp of “superpixels”, which compensates for missing depth, assumes that depth over the “superpixels” has a low gradient, and that “superpixels” are close to front-facing. The front-facing property and the depth gradients are problematic at screen corners, e.g., when “superpixels” from the same image are used across two screens.

For our experiments we are thus limited to a single screen; consequently, we only use the front wall of the immersive space described in Chapt. 3, including head and finger tracking. The lack of 4-screen image-based rendering did not prove detrimental for the purpose of our experiments, because the target of these experiments is to draw the attention of the subjects to specific tasks, which can be well accommodated within a single screen. Nonetheless, it is true that multi-screen image-based rendering could potentially afford an even greater level of immersion which would be of much utility for future studies. To this end, a direction for future work is to try to combine the benefits of proxy-based approaches for consistent view synthesis, with high quality image-based rendering using image warp approaches.

6.2.2 Specific hardware setup

The fact that our participants are elderly (adults over 60 years old) implies specific precautions in the hardware setup. To avoid any risk of unsteadiness or falling, participants sat on a chair installed in front of the display screen, at a distance of 1m, during the experiments. A small stool is placed next to the chair for the experimenter. Both are placed on a carpet reinforced with a wood plate, in order to protect the bottom screen of the immersive space (see Fig. 6.3).

To enrich the sensation of immersion in the virtual reality environments, we added

Figure 6.3: *Hardware setup of the experiments.*

ambient spatialized 3D sound. We typically play sounds such as street, car and ambient noise, crowds walking and talking etc.

6.3 Image-based rendering integration

To achieve the best possible visual quality, we used the approach of [Chaurasia 2013]. This approach is designed for free-viewpoint navigation in the scene, making it suitable for interactive virtual reality. Its integration into our system required some adaptations for tracked display and the combined display of synthetic objects and image-based rendering, which we describe below.

One important difference with traditional virtual reality is that for rendering we use the calibrated positions and orientations of the cameras of the input photographs. Care must be taken to apply appropriate transformations, especially when combining image-based and synthetic imagery.

This section is part of [Chaurasia 2014], but we repeat it here for completeness.

6.3.1 Image-based rendering

Capture and preprocessing. The method first pre-processes the input images by running standard 3D reconstruction [Snavely 2006, Furukawa 2009] and oversegments the input images using [Achanta 2012]. The oversegmentation divides the input images into macro regions called *superpixels* (see Fig. 6.4), each containing some 3D depth samples. The approach then synthesizes plausible depth in poorly reconstructed regions, compensating for lack of 3D geometry.

(a) *Input image*

(b) *Superpixel oversegmentation*

Figure 6.4: *Superpixel oversegmentation.*

The positions and orientations of the cameras used for the familiar scene are shown in Fig. 6.5.

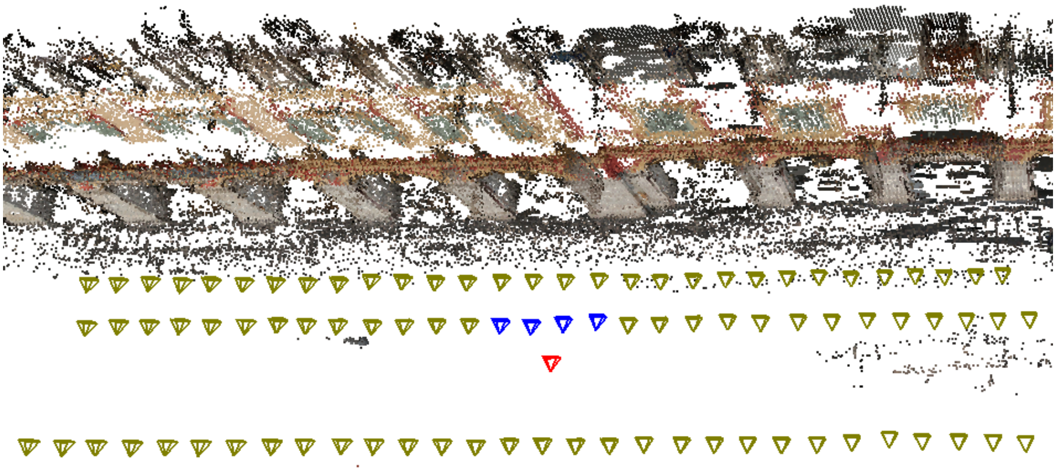


Figure 6.5: *The placement of the cameras for the capture of the familiar image-based virtual environment (see Sec. 6.5.5). This set of cameras allows navigation in a significantly large portion of the scene (*75×30 *meters shown here as top view).*

Rendering. At run time, each novel view is synthesized by pre-selecting 4 input images for the viewpoint. Each superpixel of each selected image is warped to the novel view using a shape-preserving warp. The warped images are then blended to create the final novel view, shown in Fig. 6.6. More details can be found in [Chaurasia 2013].

Figure 6.6: *The four side images show the warped superpixels of 4 input images pre-selected to render the current novel view. The central image shows the blended result of the 4 warped images.*

6.3.2 Adapted image-based rendering for virtual reality

The approach in [Chaurasia 2013] is developed for desktop applications where the user navigates using a mouse, whereas in our immersive space setup, the novel camera position V_{novel} is provided by the head tracker in real-time. The captured 3D scene is transformed so that the front screen of the immersive space is aligned with the x-axis of the scene. The virtual camera used to display the front screen has orientation $(0, 1, 0)$ and up vector $(0, 0, 1)$. These vectors remain the same irrespective of the position and orientation of the head. Thus, in OpenGL terminology, the modelview matrix is given by:

$$M_{\text{f,novel}} = \texttt{gluLookAt}\left(V_{\text{novel}}, V_{\text{novel}} + (0, 1, 0), (0, 0, 1)\right) \tag{6.1}$$

where, the subscript f denotes the front screen. The perspective matrix $P_{\text{f,novel}}$ is constructed using the standard approach of joining the head position with the four corners of the screen [Cruz-Neira 1993].

We select input images whose modelview matrices are "most similar" to that of the novel camera. Thus, to render the image, we select the input images whose center of projection is closest to V_{novel} and orientation is similar to $(0, 1, 0)$ in this coordinate system. Once we pre-select the input images, we warp them using the overall projection $C_{\text{f,novel}}$ given by $P_{\text{f,novel}}.M_{\text{f,novel}}$. The warped images are blended as described in [Chaurasia 2013].

6.3.2.1 Navigation

The above approach seamlessly handles head movement within the immersive space. To allow long range navigation, we use a pointing interface described in Sec. 6.4, where the user indicates the direction to follow. This gives an overall transformation T for the head position. However, the head position is updated asynchronously by the head tracker. Therefore, instead of transforming the head position by T, we transform the scene and input cameras by T^{-1} with equivalent effect. To transform the scene, we apply T^{-1} on each 3D depth sample used for image-based rendering. We transform the input cameras by transforming the center of projection as well as the modelview or extrinsic matrix.

While our system provides realistic images, like any image-based method it is restricted to representing content which actually exists in the input images. When the user leaves this region, visual artifacts appear. To avoid this, and to ensure that all participants navigate in the same regions, we limit navigation to the zone where artifacts are very small.

6.3.2.2 Rendering synthetic objects with image-based rendering

An important part of our system is the ability to manipulate and render synthetic objects (see Fig. 6.7). To allow this, we modify the rendering pipeline of the image-based rendering as follows. We first assign a single depth value to each superpixel, i.e., the median of depth samples of the superpixel. When the superpixel is warped to the novel view, we re-project this depth value into the novel view using the following equation:

$$d_{\mathrm{f,novel}} = C_{\mathrm{f,novel}}.C_{\mathrm{input}}^{-1}.d \tag{6.2}$$

where, d is the depth of the superpixel, C_{input} and C_{novel} are the overall projection matrices of the input and novel camera respectively. While rendering warped superpixels, we write the novel depth into the OpenGL depth buffer. Finally, we render the synthetic objects with the depth test enabled. This places the synthetic objects at the correct depth in the scene giving the correct (dis)occlusion effects.

6.4 Gestures

Another key requirement of our system is support for direct manipulation using finger tracking. We build on the approach presented in Chapt. 4, which combines ART-based finger tracking with a physics engine (see Chapt. 3), and provides realistic and close-to-natural interaction with objects in the immersive setting. Visual feedback of their thumb, index and middle fingers is presented as red, green and blue cylinders respectively. An example of gesture-based interaction is shown in Fig. 6.7. The majority of participants had no previous virtual reality or gaming expertise, and were all elderly people; we thus preferred to avoid complicating the task with a wand-like device.

To increase the sense of immersion in the image-based virtual environments and to allow participants to become familiar with the finger tracking, we start the immersive experimental conditions with participants interacting with the environment in two ways: direct manipulation of virtual objects and navigation inside the environment. We thus inserted 3D objects inside the image-based rendering scenes as explained in Sec. 6.3.2.

(a) *Familiar image-based virtual environment*

(b) *Unknown image-based virtual environment*

Figure 6.7: *Photographs of the screen of the gesture-based interface. The finger positions are given by the colored cylinders.*

6.4.1 Object manipulation

The participants can directly grab, release, translate and rotate virtual objects in 3D space using their tracked fingers. Our implementation is the same as in Chapt. 4 (see Sec. 4.1 for details): the objects behave as naturally as possible, and the physics library does not control objects being manipulated. The participant simply has to place her thumb and index onto an object to grab it, and open her fingers to release it.

6.4.2 Navigation gesture

To navigate inside the image-based virtual environments, the participants have to point in the direction they want to follow with the index finger and make a pinch gesture with the thumb and middle finger. During pilot tests, we realized that elderly people are prone to hand arthritis and that the pointing gesture was thus too constraining. Hence, we chose that the direction is specified by the palm orientation so that the pointing can be approximate. However, we kept instructing participants to point, for clarity and simplicity. Given this specification of direction, the gesture is independent of hand location and thus participants can place their hand on their knees to avoid arm muscular fatigue, which is particularly important for our elderly participants.

6.5 Experimental procedure

Our assessment methodology procedure is an adaptation of traditional autobiographical fluency [Dritschel 1992].

The experimental procedure has three main steps: inclusion, preparation and exposure to the different conditions. Before starting the experiment, two steps are performed: a stereo-blindness test (see Sec. 6.5.4) and hand-tracking device calibration (see Sec. 3.2.1).

6.5.1 Population

The participants were 13 healthy elderly participants with varied education level (from elementary school to university), 5 women and 8 men, recruited by the clinical Memory Center research group. The mean age was 66.84 years, with a standard deviation of 4.33 years. Each participant underwent a cognitive behavioral and motor assessment. Subjects not living in the city of experimentation, with a clinical history of neuropsychiatric disease or dementia, with a Mini Mental Score [Folstein 1975] lower than 28 (see Sec. 6.5.3) were excluded as well as subjects with dizziness or motor instability. All participants gave their informed consent before beginning the study. Ethical approval was received from the competent Ethics Committee.

6.5.2 Autobiographical memory assessment

In the autobiographical fluency test adapted for the experiment, participants had 2 minutes to generate as many memories as possible related to the environment they see, using short sentences. At the end of this time period, the experimenter assessed the quality of the recollection using the *Remember/Know* procedure [Gardiner 1998]. The subject was asked to indicate for each of her recollections if it was a *Remember* – i.e., conscious recollection of many vivid contextual details, such as “when" and “how" the information was learned – or a *Know* – i.e., the sensation that the memory/event/environment has been seen before, but not being able to pin down the reason why.

The subjects verbal output was recorded during the experiment. The audio recordings were analyzed by a speech linguist independent from the design of our study. Derived scores include the *Total* number of memories, as well as the number of *Remember* and *Know* recollections. In addition to the autobiographical fluency, subjects had also to complete questionnaires assessing the acceptability and the qualities of the experiences during the different conditions and a standard Presence questionnaire (see Sec. 6.6.2).

6.5.3 Clinical inclusion

Clinical inclusion aims at verifying if the subject is a healthy adult and not a patient suffering from a memory disease. Before starting the experiments in the immersive space, each participant spends about 45 minutes with a clinician in a separate room.

After a motor clinical examination, the following tests are performed: mini mental score [Folstein 1975], short cognitive battery [Robert 2003], frontal assessment battery [Dubois 2000], and the apathy inventory [Robert 2002]. This ensures that the participant is physically and mentally fit to participate in the experiment. If successful, the experimenter accompanies the participant to the immersive space. If the participant is considered as a patient, the experiment is canceled. In practice, all participants succeeded the inclusion and took part in the study.

In addition, the participants are asked to complete a pre-immersion PANAS questionnaire (see Sec. 6.6.2.1).

6.5.4 Stereo-blindness test

After clinical inclusion, each participant is invited to sit on the chair in the immersive space and a stereo-blindness test is performed (see Sec. 3.1.1 for details). The experiment continues only if the participant confirms that depth is correctly perceived. All participants had correct stereo vision.

6.5.5 Environments

We have four different experimental conditions. First, we have two baseline conditions: a control condition which is a grey-colored blank screen and a photograph of a familiar location, which corresponds to the use of photos for traditional autobiographical fluency. We then have two conditions with *IVIRAGE*, namely an unknown and a familiar environment. The four conditions are shown in Fig. 6.8, (a)-(d). We captured the two scenes in (c) and (d) by taking 90 and 35 photographs for each of *FamIBVE* (familiar image-based virtual environment) and *UnknoIBVE* (unknown image-based virtual environment) respectively. The placement of the cameras used for the first environment is shown in Fig. 6.5. The total time for photography, reconstruction and preprocessing is between 2-4 hours, while rendering is real-time.

The 2D environments consist in displaying images on the screen. Baseline conditions are displayed as simple images, without stereo vision nor tracking. Both *IVIRAGE* conditions are constructed and displayed with the method of [Chaurasia 2013].

Grey baseline The first condition (a), which we refer to as *Grey* consists of a blank, medium grey screen.

(a) *Grey*

(b) *FamPhoto*

(c) *FamIBVE*

(d) *UnknoIBVE*

Figure 6.8: *The four conditions used in our experiment.*

Known landmark photograph The second condition (b), referred to as *Familiar Photo (FamPhoto)*, represents a well-known location in the city of the experiment.

Known IBVE We refer to the third condition (c) as *Familiar image-based virtual environment (FamIBVE)*. It consists of an image-based representation of a known landmark square in the city of the experimentation.

Unknown IBVE The fourth condition (d) is the *Unknown image-based virtual environment (UnknoIBVE)* which is captured in a public housing neighborhood, and contains buildings without recognizable features.

6.5.6 Exposure to the environments

After the hand-tracking device calibration, the actual experiment starts. The four environments are presented one after the other to each participant. Since the Grey environment is a control stimulus, its order has little influence in our analysis. We thus chose to perform this in the second position for all the groups, reducing their number. The other three environments are randomized in 6 different groups. Because we are studying the contributions of *IVIRAGE* in reminiscence therapy, and comparing it to the traditional photo-based reminiscence therapy, the order of the *FamPhoto* is important.

In all the environments, the participant sits on the chair and the clinician sits next to her on the stool.

For the baseline conditions, the participant is presented with the photograph or blank screen for 15 seconds, and then the verbal fluency session begins.

For the *IVIRAGE* conditions, the session starts with the interactive gesture manipulation of synthetic objects. This involves manipulating synthetic 3D objects in the virtual environments. Specifically, a plate with two dishes (one next to the other) is presented to the participant. The right dish contains 3 apples: two red and one green. The apples are dynamic rigid objects following gravity rules, controlled by the physics engine (see Sec. 6.4.1). We show these objects in Fig. 6.7. Spatialized sound feedback related to the dynamic virtual objects is provided when the apples are removed or placed on the dishes; a spatialized ambient sound is also played (street, car and crowds noise). The participant is then invited to navigate in the environment using the pointing gesture for about 45 seconds. At the end of the navigation, the verbal fluency test is performed.

6.6 Measurements

6.6.1 Objective metrics

The speech linguist recorded the number of responses and determined those that are conscious responses (*Remember*) or more vague recollections (*Know*) (Sec. 6.5.2). Only *Remember* values are relevant for our study.

6.6.2 Subjective metrics: post-exposure questionnaires

After the autobiographical fluency test each participant completed a set of questionnaires.

6.6.2.1 General questionnaire

A general questionnaire was presented on a tactile tablet to each participant at the end of each condition. The participant was asked whether the environment (except *Grey*) was recognized (YES/NO), and a continuous scale (Visual Analogue Scale (VAS) [Aitken 1969]) from 0 (red colour) to 10 (green colour) was used to rate the Familiarity with the environment (except *Grey*) and Emotion, Anxiety, Motivation, Security and Fatigue for all conditions.

We used a PANAS pencil-and-paper questionnaire (affect measures questionnaire, [Thompson 2007]) for the users to self-report responses to questions to assess their current feeling or basic predisposition. We used the 20 items version [Watson 1988]. For each feeling or emotion, the user had to indicate her feelings and emotions right at the moment on a 5 scores scale: from 1 (very slightly or not at all) to 5 (extremely). In addition to being filled before the beginning of the experiments (see Sec. 6.5.3), this questionnaire is filled after the two conditions *UnknoIBVE* and *FamIBVE*.

6.6.2.2 Cybersickness and presence questionnaire

For the *IVIRAGE* conditions (*UnknoIBVE* and *FamIBVE*), we tested cybersickness with the 22-item questionnaire of [Viaud-Delmon 2000], and the sense of presence using the questionnaire[2] presented in [Schubert 2001].

The cybersickness questionnaire is also a pencil-and-paper questionnaire; it contains 22 physical sensations. For each one of them, the participant is asked to indicate the level: from 0 (no sensation) to 4 (strong sensation).

The presence questionnaire makes it possible to evaluate the *spatial presence*, the *involvement* and the *experience realism* [Regenbrecht 2002, Schubert 2001].

At the end of the experiments, the participants are asked to fill a *General Acceptability* questionnaire. It aims at evaluating the *satisfaction*, the *interest in* and the *comfort* of the entire experience. Each participant has to give a score (from 1 to 10) for each one of these 3 items. They also completed a PANAS [Thompson 2007] questionnaire for clinical evaluation.

The participants were also asked to provide any other feedback they felt appropriate.

6.7 Results

We present the results of our quantitative measurements as well as qualitative evaluation elements.

6.7.1 Objective measurements

Since the number of *Remember* responses is the most significant measure in the present study, Fig. 6.9 illustrates the number of such answers for each participant for the *IVIRAGE* conditions.

[2] http://www.igroup.org/pq/ipq/index.php

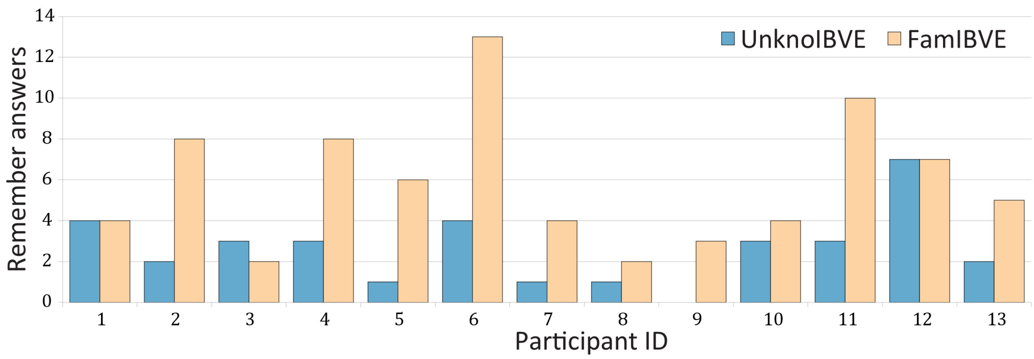


Figure 6.9: UnknoIBVE *vs.* FamIBVE*:* Remember *answers.*

The median number of *Remember* responses for *Grey*, *UnknoIBVE*, *FamIBVE* and *FamPhoto* are respectively: 1, 3, 5 and 5. We notice that the exposure to image-based virtual environments results in a larger number of responses for the familiar case than for the neutral, in all cases but one.

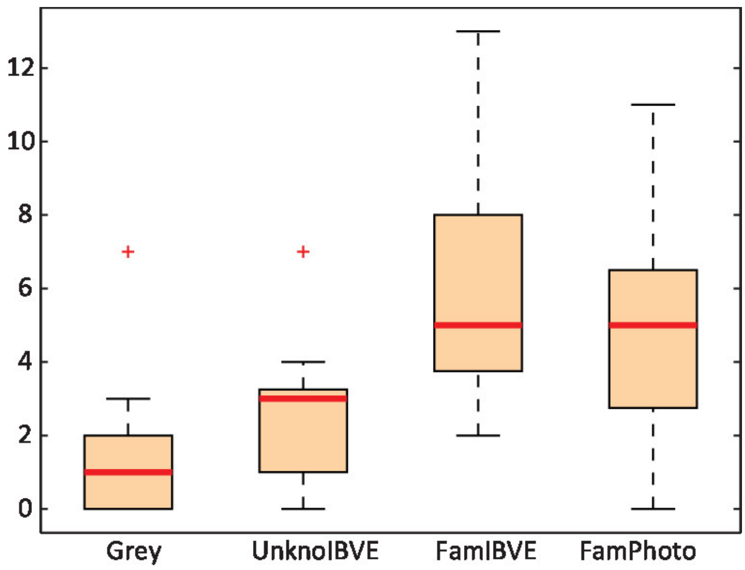


Figure 6.10: Remember *data for the 4 conditions. The red bar is the median; we also show quartiles and distribution.*

The *Remember* responses for the 4 conditions are shown in Fig. 6.10. The median number of *Remember* responses are very close for *Familiar IBVE* and *Familiar Photo*. We performed the non-parametric Friedman test [Cunningham 2011] on this data. A significant main effect of condition was found for *Remember* ($\chi^2 = 18.22$, $p = 0.0004$). No effect of the group order was found ($\chi^2 = 6.23$, $p = 0.82$).

We then ran a Wilcoxon signed ranks test [Cunningham 2011] to determine whether the differences between conditions are significant. For *Remember*, *Grey* is significantly different from *FamIBVE* ($p = 0.0002$) and from *FamPhoto* ($p = 0.00056$); *UnknoIBVE* is significantly different from *FamIBVE* ($p = 0.004$) and *FamPhoto* ($p = 0.01$). No other significant differences were found.

6.7.2 Subjective measurements

6.7.2.1 General questionnaire

All the participants indicated that they recognized the *FamIBVE* but not the *UnknoIBVE* (except one participant who knows both environments (see Sec. 6.7.3)). The results from the familiarity question show a high score for the *FamIBVE* (mean = 7.92, standard deviation = 1.86) and for the *FamPhoto* (mean = 8.74, standard deviation = 1.44) and a low score for the *UnknoIBVE* (mean = 3.28, standard deviation = 3.02). The results for the Emotion, Motivation and Security are shown in Tab. 6.1.

	Grey			*UnknoIBVE*			*FamIBVE*			*FamPhoto*		
	E	*M*	*S*	*E*	*M*	*S*	*E*	*M*	*S*	*E*	*M*	*S*
Mn	4.29	4.84	7.86	3.82	6.44	7.46	7.23	8.59	9.4	7.30	8.64	9.43
SD	3.54	3.14	3.00	3.27	3.41	3.28	1.56	1.3	0.47	1.98	0.99	0.38

Table 6.1: *Acceptability and qualities of the experiences (Emotion* (E)*, Motivation* (M)*, Security* (S)*) during the different conditions.* Mn *and* SD *respectively refer to the mean and the standard deviation.*

Emotion is highest in *FamPhoto* (mean = 7.30, standard deviation = 1.98) and *FamIBVE* (mean = 7.23, standard deviation = 1.56), while it was low for *Grey* (mean = 4.29, standard deviation = 3.54) and *UnknoIBVE* (mean = 3.82, standard deviation = 3.27). Mean Motivation is highest (mean = 8.59, standard deviation = 1.3) in *FamIBVE*, while mean Security for *FamIBVE* (mean = 9.7, standard deviation = 0.47) is on a par with *FamPhoto* (mean = 9.43, standard deviation = 0.38).

We first ran a Levene normality test [Cunningham 2011] for the condition data for Emotion, Motivation and Security. We found that only Motivation follows a normal distribution. We thus ran a two-way ANOVA test [Cunningham 2011] on the Motivation data and a Friedman test on the Emotion and Security data. The results are as follows: there is a significant main effect of condition for Emotion ($\chi^2 = 13.64$, $p = 0.0034$) and Motivation ($F = 5.56$, $df = (3, 12)$, $p = 0.0030$); there is no effect for Security ($\chi^2 = 4.00$, $p = 0.26$). There is no significant effect of group for all these variables.

The Wilcoxon signed ranks test showed significant differences for Emotion between *Grey* and *FamIBVE* ($p = 0.01$); *Grey* and *FamPhoto* ($p = 0.013$); *UnknoIBVE* and *FamIBVE* ($p = 0.003$); *UnknoIBVE* and *FamPhoto* ($p = 0.0012$). Using the same statistical test, Motivation is significantly higher for *FamIBVE* compared to *Grey* ($p = 0.018$), *FamPhoto* compared to *Grey* ($p = 0.0073$) and *UnknoIBVE* compared to *FamPhoto* ($p = 0.034$).

Finally, participants did not experience fatigue (mean = 0.87, standard deviation = 0.27 on a scale of 0 to 10) or anxiety (mean = 2.3, standard deviation = 0.15 on a scale of 0 to 10) during the entire experiment across all conditions.

The results for other parameters (on a scale of 0 to 10) assessing the general acceptability are shown in Tab. 6.2.

	Interest	Satisfaction	Comfort
Mn	8	8.18	8.42
SD	2.62	2.62	2.69
Md	8.9	9.1	9.5
Va	6.90	6.87	7.1

Table 6.2: *General acceptability of the conditions. Mn, SD, Md and Va respectively refer to the mean, standard deviation, median, and variance.*

The PANAS test reported in Tab. 6.3, provides measures of positive and negative affect (PA and NA) pre- and post-immersion. We can see that the *Unknown IBVE* reduces the PA, while the *Familiar IBVE* has a lesser effect. We can also see that the image-based virtual environments do not introduce NA. However, the differences were not statistically different.

	Pre-Immersion		UnknoIBVE		FamIBVE	
	PA	NA	PA	NA	PA	NA
Mn	32.18	11.9	26.18	11.72	29.72	10.9
Sd	5.32	2.94	8.53	3.03	9	1.57
Md	30	11	25	10	30	10

Table 6.3: *PANAS: positive affect (PA) and negative affect (NA) schedule. Mn, SD and Md respectively refer to the mean, standard deviation and median.*

6.7.2.2 Cybersickness and presence questionnaire

The results of the Presence questionnaires have high overall scores (median = 5 on a scale of 0 to 6) for *General Presence* and *Spatial Presence* [Schubert 2001]. The *Experienced Realism* and *Involvement* have medium scores (median = 3 on a scale of 0 to 6) for both *UnknoIBVE* and *FamIBVE*. The Friedman test showed no significant main effect of condition ($\chi^2 = 1.32, p = 0.87$)

Also, participants did not experience sickness during the entire experiment across all conditions. This is shown by the data from the cybersickness questionnaire (mean = 0.46, standard deviation = 0.1 for *UnknoIBVE* and mean = 1.42, standard deviation = 0.3 for *FamIBVE* on a scale of 0 to 88).

6.7.3 Informal qualitative evaluation

We did not perform a formal qualitative evaluation, but we report some informal feedback and anecdotes from the experimental session.

All participants found the experience provided by *IVIRAGE* to be very interesting and stimulating; they expressed their satisfaction with many positive comments at the end of the sessions, both for interaction and navigation. One participant (number 3 in Fig. 6.9) had worked in the construction industry, and had actually worked on the construction of our unknown environment. Evidently, the unknown environment was actually familiar for this participant, which is reflected in the results of her fluency test.

In our familiar environment, there used to be a newsagent store which was recently closed, replaced by a fashion store with a much younger clientele. This newsagent had significant emotional value for the elderly population, and the vast majority of our participants commented on this issue, stating that they found this to be “a disgrace”, “really a pity” etc.

Several participants commented that, despite the high level of realism, they found that the scene was “deserted”. We provided the (true) explanation that the scenes are situated in the very early morning, and thus no people are present. However, the addition of virtual humans and crowds would definitely be a plus, but this is left for future work.

6.7.4 Discussion

From a clinical perspective the results of our study underline two main points which hold promise for the future: 1) it is possible to use *IVIRAGE* in a population of elderly subjects and 2) the system can stimulate conscious recollections of autobiographical memory.

Concerning the first point, the physical setup used (a chair or a bench) shows that this approach adapts well to populations e.g., with limited mobility. The responses to the questionnaires on Emotion, Motivation and Security show that the technological setup was well tolerated by our elderly participants since the data of *FamIBVE* are comparable to the traditional approaches as tested by *FamPhoto*. The fact that *IVIRAGE* can adapt the virtual environment to the personal history and surroundings of participants holds particular promise from a clinical perspective, since it will be able to cognitively stimulate low mobility subjects.

Concerning the second point, previous work has already demonstrated the utility of virtual reality for cognitive/memory processes (e.g., [Brooks 2003, Plancher 2012]). The increased number of memories with *IVIRAGE* for familiar vs. unknown scenes, shows that our system can help generate conscious autobiographical memories. This has several clinical implications. The goal of cognitive impairment treatment is to provide information on the subjects long term functional status. In this field, the ecological validity or generalizability of neuropsychological test results, beyond the clinical environment, is a crucial component of the evaluative process [Kibby 1998]. The highly realistic environments of *IVIRAGE* are both ecological and generalizable, and give the additional opportunity to create an environment familiar to the patient and therefore closer to her real life.

Our study also shows benefits from the use of virtual reality in a therapeutic context. Our results indicate a positive effect of *IVIRAGE* on autobiographical memory which is a core component of reminiscence therapy. The number of generated memories with *FamIBVE* is on a par with those from the traditional reminiscence therapy protocol which involves photographs: this shows that our approach is at least as effective as traditional therapy, additionally confirming the applicability of our approach. Our immediate future objective is to use our *IVIRAGE* system with patients in a clinical setting.

One of the great challenges in reminiscence therapy is whether different technologies maintain users motivation and engagement when confronting them with a repetitive series of training challenges, either cognitive or physical. This is of particular interest given that apathy i.e., the disorder of motivation, is the most frequent behavioral disturbance in elderly subjects with Alzheimer's disease and related disorders. The interactivity and realism of *IVIRAGE* give ample opportunity to improve motivation for training processes, with many positive consequences for patients. Engagement, defined as the act of being occupied or involved with an external stimulus [Cohen-Mansfield 2009] is the most challenging task for professional carers. Understanding individual interests may guide intervention parameters to improve engagement and potential benefits. Image-based capture and display allow tailored intervention for individuals suffering from cognitive problems. It is reasonable to expect that motivation can be further improved if the training process using virtual reality is based on image based to the person life.

From a technical standpoint, the higher number of generated memories for *FamIBVE* compared to *UnknoIBVE*, is a very strong indication that image-based rendering does convey familiarity in our immersive setting. Similar to recent studies on the effect of realism for virtual reality [Yu 2012], such indirect measures of realism are often more powerful indicators than, for example, direct questionnaires. The strong Presence scores also underline the ability of image-based rendering to convey a sense of "being there".

Despite the advantage explained in the introduction of the wide panoramic view captured for *FamPhoto* compared to the intrinsic narrow field-of-view of *FamIBVE*, we see that the number of responses is the same for both conditions, confirming the effectiveness of *IVIRAGE* for reminiscence therapy. We expect that the performance of *FamIBVE* will increase when the solution for multiple screens is available, since the field-of-view for the image-based virtual environments will be much larger.

The most pressing algorithmic direction of future work is the improvement of the underlying image-based rendering algorithms. This includes the development of a solution for multiple screens, thus providing full immersion. The addition of virtual humans to populate the scenes is another interesting challenge that should render the approach even more engaging and useful.

6.8 Conclusion and future work

The potential clinical implications of the use of *IVIRAGE* – if successful – are numerous. For example, the virtual environments provided by *IVIRAGE* are ecologically valid and generalizable, which is crucial for memory evaluation [Kibby 1998]. Another aspect concerns the realism and interactivity of *IVIRAGE*, which hold great promise for motivation and engagement [Cohen-Mansfield 2009], which are also key components of successful reminiscence therapy.

Our contributions can thus be summarized as follows:

- We introduce the *IVIRAGE* system, which demonstrates the first usage of free-viewpoint image-based rendering, combined with finger tracked gestures and physics, in an immersive setting for reminiscence therapy.
- The results of our study show that the number of autobiographical memories generated during the exposition to a familiar environment presented in *IVIRAGE* is higher than that for an unknown environment. This strongly indicates that image-based rendering can convey familiarity of a given scene, which is an essential requirement for the use of virtual reality in reminiscence therapy.
- Our study also shows that the *IVIRAGE* can be useful for reminiscence therapy, since acceptability scores indicate that the system is well tolerated by elderly patients, and by results that show that our system generates memories as effectively as traditional reminiscence therapy.

We thus strongly believe that the extension of our system to full immersion and panoramic views will make *IVIRAGE* even more efficient, and that reminiscence therapy will thus greatly benefit from virtual reality technologies.

The results of our study shows that image-based techniques offer great promise for reminiscence therapy, and for virtual reality in general. Our *IVIRAGE* system has numerous advantages over traditional 3D assets used in virtual reality. Even though it was not the focus of this study, the fact that only a few casual photographs are required to create a scene that can be used for virtual reality is an advantage with significant consequences. The level of realism obtained by the imagery, despite some residual artifacts, is at least as good as that produced at great cost with traditional means e.g., manual modeling.

In conclusion, our study has given strong first evidence of the utility of image-based rendering and close-to-natural gesture interaction for immersive virtual reality systems, in the context of memory treatments and in particular reminiscence therapy. We are confident that our work will open the way to a much wider adoption of such technologies. The ease of creation of the image-based virtual environments has the potential to not only greatly impact the use of virtual reality for therapy, but also in other applications, since personalized assets can now be created and realistically displayed at low cost.

From the clinical standpoint, this work opens up several interesting directions. Following the experiment, it is possible to use virtual reality for autobiographical memory stimulation. This needs now to be done in subjects suffering from mild to moderate memory disturbances. In this perspective, it is of the utmost importance to adapt the virtual reality solution to a clinical setting. This is the case of the memory center in Nice, France, which has recently been equipped with a virtual reality space featuring 3D stereoscopic screens, in order to perform such experiments in real conditions. This wall is smaller than the front screen of our immersive space, but large enough to occupy most of the user's field of view and thus generate a certain sense of immersion. However, it would be interesting to study the effect of the size of the screens on the number of recollections as a measure of presence.

CHAPTER 7

Conclusion and future work

The goal of this thesis was to evaluate gesture-based interaction in the context of fully immersive virtual reality spaces. Our contributions consist in the proposition and evaluation of a combined heuristics-based and physics-based solution for direct manipulation, the description of a complete framework for the analysis and decomposition of movement in 3D manipulations, and a feasibility study of the use of virtual reality and gestures for reminiscence therapy.

7.1 Evaluation of a new solution for direct manipulation

Our direct manipulation solution provides a simple and close-to-natural way of handling 3D virtual objects for general purpose and relatively complex tasks within fully immersive CAVE-like systems. It combines a heuristic approach based on a finite state machine and a simple real-time physics simulation to handle collisions. Our hand manipulation interface is robust to tracking noise and simulation instabilities. It allows users to grasp, translate, rotate and release objects in space intuitively, and to combine these manipulations to perform complex tasks such as maintaining balance while walking. We evaluated feasibility of direct manipulation, and clearly demonstrated its utility and advantages over wand-like interfaces, in terms of presence and naturalness, even if the wand outperforms direct manipulation in terms of speed and accuracy. Our results for fully immersive setups are coherent with [Bowman 2012]: direct manipulation could be considered better for applications such as training since performance is closer to the real world than the wand which, however, augments interaction capabilities of the user.

Evaluation consisted in a user study involving a usability task for analysis and a free-form user experience for observation. Our direct manipulation approach was compared to a wand interface and to a real world replica for the usability task. The observation task clearly revealed the sense of presence experienced by users when using direct manipulation since their automatic behaviors were those of everyday life. Concerning the analysis, our results showed that virtual tasks took longer than real world task, strongly indicating that virtual techniques are not yet at the point where they can perform at the same speed as their real equivalent. We explained this through the lack of haptic feedback in the virtual setups which, however, provides an important guide in real conditions, and we recalled that current solutions for tactile feedback are not appropriate for immersive setups.

7.1.1 Future work

This study revealed two interesting directions for future work.

First, using a full hand model with complete tracking would allow for more complex manipulations, and thus provide even more natural and immersive experiences. It would require us to adjust the heuristic approach accordingly, but, more importantly, it would need improvements in tracking technology before becoming realizable. The ultimate goal would be to incorporate haptic feedback once the technology for immersive setups has been perfected.

Second, we noticed that our study involved many parameters, such as the movement itself, task difficulty, involved skills, etc. Thus, performing specific studies to analyze the impact of those factors individually would be interesting.

7.2 Framework for movement decomposition

Our framework allows us to study 3D motion for direct manipulation interfaces in fully immersive setups. It proposes a design of devices that allow the careful decomposition of movements according to their nature (i.e., rotation or translation) and degrees of freedom (i.e., 1D or 3D tasks). The devices are simple and intuitive, using well-established existing mechanisms that make them easily reproducible in real and virtual settings and that allow users to focus on completing the task rather than on understanding how they function. They involve 1D translations, 1D rotations, 3D translations, 3D rotations, and 3D free movement, each within a separate device. We evaluate the impact of motion type on the global performance through simple pointing and orienting tasks, and compare to the performance obtained with a wand-based interface.

As in our previous experiment, our results clearly show that using finger tracking increases the sensation of immersion while the wand outperforms direct manipulation in terms of performance. We also confirm that none of the virtual techniques can reach the capacities of physical setups, even for simple tasks. Our analysis showed that increasing the task complexity increases the difficulty of anticipating the required motion. We also reported that designing such devices requires compromises between visibility, weight and robustness that affect performance. Finally, we demonstrated that our devices for 1D translations and rotations followed Fitts's model, and that the estimated parameters of the model were coherent with our previous observations on performance of the techniques.

7.2.1 Future work

A first and simple way to improve the design would be to automate the physical devices for measurements since this would provide richer information for further analyses.

Another interesting direction would be to study the effect and relevance of a fair comparison between input techniques since, for now, finger tracking devices suffer significantly from noise and instabilities. The noise pattern of the finger tracking devices could be synthesized and applied to the wand, and both input signals could then be filtered similarly. This would make the performance of both devices equivalent, but would alter one tool. However, this could give a hint about the capacities of direct manipulation over wand interfaces once the technology provides better tracking.

Lastly, it would be interesting to design a single device that would allow for any movement decomposition. Such a "super-device" could couple a gyroscope-based mechanism with a combined slider system to manipulate a central object which would hold three lasers oriented to build a light frame. Carefully designed locking/unlocking mechanisms would enable only the desired degrees of freedom. This would allow every degree of freedom combination, and thus even 2D tasks which are not considered in our study.

7.3 Feasibility study for reminiscence therapy

Our system evaluates the feasibility of using virtual reality technologies for reminiscence therapy. It presents a solution which combines easy presentation of highly realistic environments on a single stereo screen, natural interaction with 3D virtual objects and free-viewpoint navigation within the environments. The system provides ecologically valid and generalizable virtual environments, and the realism and interactivity allowed are crucial for patients' motivation and engagement, and thus for the success of the therapy. Our system involves a verbal autobiographical fluency test specifically adapted to virtual reality. We also adapted our setup to older adults who may suffer from limited mobility. Our study revealed a high acceptance and tolerance rate from elderly participants.

Users were presented with familiar and unknown virtual environments as well as with a classic photograph of a known place. The study revealed that the system was capable of eliciting memory as effectively as traditional therapy, and that it can convey familiarity of a given scene as more memories were generated within familiar environments. The system can thus stimulate conscious recollections of autobiographical memory. This proves that immersive virtual reality systems combining close-to-natural gesture interaction and highly realistic rendering are useful in the context of reminiscence therapy.

7.3.1 Future work

Two main directions for future work arise from this work.

The immediate future objective is to use the system in a clinical setting with patients suffering from mild to moderate memory disturbances. Thus, the memory center in Nice, France, is in the process of adapting this solution to a clinical setting since they are being equipped with 3D stereoscopic screens in order to perform such experiments in real conditions.

Algorithmically, an important improvement would be to adapt the system to fully immersive multi-screen spaces to provide panoramic views, and to add virtual crowds to populate the scene and thus increase the sense of presence.

7.4 Concluding remarks

In all of our experiments, quantitative results clearly prove that using gesture-based interfaces feels more natural than wand-based interfaces. However, the wand clearly outperforms direct manipulation in terms of speed or accuracy. The two main reasons for this are a) the technological aspect of the devices, and b) the hyper-natural characteristic of wand-like tools. This shows that the chosen interface must always be adapted to the application requirements. When everyday life is replicated and commonplace objects are involved, gesture-based interfaces should be preferred, but they can be coupled with tool-based interfaces. From a subjective point of view, we observed very positive reactions from all participants when interacting and manipulating virtual objects with their hands. This was also true for elderly participants, which clearly shows the potential of such direct manipulation interfaces which provide interaction close to real life.

Unfortunately, such studies can be very difficult to perform because of the technical aspects linked to the constraints of fully immersive spaces, such as the integration of physics simulation or the finger tracking suffering from noise. Indeed, CAVE-like spaces are wide, closed and black. This raises visibility problems for vision-based systems, and non trivial range or complexity issues in the placement of sensors, resulting in noise and instabilities in the signal. We think that this is the main reason why our results show such a performance advantage for wand-based interface, contrary to what we expected. However, we truly believe that technological progress should alleviate this issue.

Another important contribution of this work is the comparison with real conditions, which we performed for our first two experiments. The physical setup provides a baseline to the experiment and a solid foundation for our research. Moreover, it has been really interesting to observe and compare people's reactions towards virtual and physical setups, and to see, for example, that people automatically try to catch falling virtual objects as they do in real life. We think that in-depth studies of interface paradigms that are reproducible in virtual and real settings are a very promising direction for future research.

APPENDIX A

FRANÇAIS
Introduction

Pendant plus de cinquante ans, la réalité virtuelle a gagné de plus en plus en popularité, et elle continue de voir son matériel progresser et ses possibilités d'applications se diversifier. Elle est désormais au croisement de nombreux domaines et elle commence à connaître un usage étendu. Elle fournit aux utilisateurs des expériences réalistes, à travers la combinaison configurable d'un affichage stéréo haute qualité, de son spatialisé, d'un système de tracking et d'appareils de retour de force. Cette quête du réalisme amène de plus en plus les designers d'applications à opter pour une interaction manuelle en raison de l'aspect naturel qu'on leur connaît, et pour des installations complètement immersives car elles sont reconnues pour fournir une sensation de présence plus importante dans les environnements virtuels. Dans cette thèse, nous nous intéressons donc au contexte difficile de l'utilisation de gestes en espace complètement immersif.

Depuis ses débuts, la réalité virtuelle a été utilisée intensivement pour l'entraînement militaire ou les simulations de vol, et il a été prouvé que les interfaces naturelles, parce qu'elles sont intrinsèquement intuitives, peuvent aider au transfert de l'entraînement virtuel vers le monde réel [Bowman 2012]. Les avancées matérielles ont aussi rendu possibles de nouvelles applications telles que le design industriel. Les tests en réalité virtuelle permettent de réduire le nombre d'itérations de design et donc les coûts de prototypage, et d'accélérer la mise sur le marché.

Mais de nos jours, nous assistons à l'explosion de la réalité virtuelle. Une première raison à cela est sa large accessibilité au public à travers le très populaire domaine des jeux vidéo. L'important progrès technologique a rendu de nouvelles interfaces accessibles aux masses. Premièrement, en 2010, la Kinect[1] a permis aux utilisateurs de contrôler des jeux vidéo avec leur propre corps. L'appareil, initialement conçu pour des espaces d'interaction larges, a une deuxième version conçue pour les utilisateurs de PC depuis 2012, étendant d'avantage son usage. Ensuite, en 2012, le contrôleur Leap Motion[2] a permis aux utilisateurs d'interagir avec des applications de bureau directement avec leurs mains nues. Une autre innovation concerne le matériel d'affichage. Le très récent casque HMD (head mounted display) Oculus Rift[3] devrait être disponible en 2015 et abordable

[1]http://www.xbox.com/fr-FR/Kinect
[2]https://www.leapmotion.com/
[3]http://www.oculusvr.com/rift/

pour une majorité de joueurs; il est déjà supporté par plusieurs jeux vidéo.

Dans ce qui suit, nous présentons plusieurs domaines qui ont motivé ces travaux de thèse.

A.1 Réalité virtuelle et réhabilitation

Une raison essentielle de l'essor de la réalité virtuelle est le fait qu'elle est considérée par beaucoup comme l'avenir de la réhabilitation clinique en général, et du traitement des lésions cérébrales en particulier [McGee 2000, Rose 2005, Weiss 2006], devenant un élément à part entière de l'évaluation et de la réhabilitation cognitives, motrices et fonctionnelles. Des études ont prouvé que la réalité virtuelle est très efficace pour l'apprentissage actif, qu'elle fournit des environnements puissants, sûrs et réalistes, qu'elle est parfaitement adaptable aux besoins des patients, qu'elle motive considérablement les patients, qu'elle permet d'enregistrer des mesures objectives de performance, et qu'elle fournit des modes de retour alternatifs [Weiss 2006]. De tels flexibilité, capacité d'immersion et contrôle de l'interaction la rendent tout à fait appropriée pour la réhabilitation. En effet, la réalité virtuelle s'est montrée efficace pour les thérapies par exposition et est de plus en plus utilisée en complément de techniques traditionnelles pour le traitement d'une large gamme de troubles psychologiques tels que les phobies et d'autres troubles neurologiques tels que le trouble de stress post-traumatique, la négligence spatiale ou les troubles du spectre autistique. Plus récemment, ces technologies ont été utilisées intensivement pour la réhabilitation mémorielle, notamment dans le contexte du traitement de la maladie d'Alzheimer où l'entraînement par réalité virtuelle a un impact significatif sur les troubles liés à l'âge. La réalité virtuelle a été utilisée pour évaluer les mémoires prospective, épisodique et visuo-spatiale, et présente l'avantage d'être reproductible entre les laboratoires et les patients. En général, la réalité virtuelle obtient des taux d'acceptabilité et de tolérance élevés à la fois de la part des sujets sains et déficients. Cependant, l'utilisation de la réalité virtuelle avec des personnes âgées requiert des considérations spécifiques dans le design du système car ils sont plus sujets à la fatigue, et l'arthrite limite leurs capacités de mouvement; la tolérance doit être évaluée.

A.2 Présence et immersion

Une réaction aussi positive face à la réalité virtuelle s'explique par la sensation de présence que les patients ressentent face aux environnements virtuels. [Slater 2009a] présente cette notion comme l'effet combiné de l'illusion d'emplacement, qui est la sensation d'être dans un endroit réel, et l'illusion de plausibilité, qui est l'illusion que la situation et les évènements se passent réellement. Cependant, une telle sensation dépend grandement des limitations techniques du système de réalité virtuelle. Plusieurs études rapportent les limitations de l'utilisation de la réalité virtuelle. Dans le contexte de l'évaluation clinique, le coût de tels systèmes les rend difficilement abordables pour des cliniciens indépendants. De plus, malgré le réalisme et le contrôle permis, l'implication de cliniciens dans le processus reste cruciale, car le design et les résultats des expériences sont sujets aux biais et interprétations.

L'élément-clé derrière toutes les solutions proposées est le progrès technologique, puisque le compromis entre les limitations et les besoins détermine l'approche utilisée. Depuis le travail fondamental de [Sutherland 1968] qui a conçu le premier HMD, les appareils d'affichage ont beaucoup évolué et se sont beaucoup diversifiés, tout comme les appareils d'interaction, permettant de créer des systèmes plus riches et plus complexes.

A.3 Affichages

En ce qui concerne les affichages, l'immersion a été un centre d'intérêt fondamental. Alors qu'un moniteur peut être suffisant pour des opérations simples, la plupart des applications utilisent des affichages stéréo et combinent les espaces de manipulation et de visualisation à travers cette perception de la profondeur. Le tracking de la tête est aussi une caractéristique critique, car il permet aux utilisateurs de voir depuis leur propre point de vue. Les premières solutions comprennent les systèmes sous forme de table, tels que le Responsive Workbench [Krüger 1994], et des affichages de la taille d'un mur. Ce sont des affichages par projection, combinés à des systèmes de tracking et de son intégrés. Cependant, de telles installations ne fournissent qu'une immersion restreinte puisque l'utilisateur voit évidemment le monde réel alentour. Pour pallier ce problème, les systèmes de type CAVE [Cruz-Neira 1993] combinent plusieurs de ces affichages muraux pour former un grand espace de travail avec un niveau d'interaction élevé. Une autre solution très précoce pour l'immersion totale est l'utilisation de HMDs [Sutherland 1968]. Ils offrent de larges espaces de travail et empêchent l'utilisateur d'être distrait par le monde réel. Ils ont été utilisés de manière intensive dans les applications d'entraînement et de réhabilitation.

A.4 Appareils d'interaction

En ce qui concerne les appareils d'interaction, les récents progrès combinés des simulateurs physiques et du matériel d'interface gestuelle ont mené à un grand intérêt pour ce type d'interaction, à la fois pour la réalité virtuelle et pour la réalité augmentée. Les gants sont l'appareil le plus courant pour obtenir des informations quant à la posture et au mouvement des mains de l'utilisateur car ils préservent la dextérité de la main. La manipulation d'objets virtuels (telle que pointer, attraper, translater, tourner et relâcher) est devenue plus précise et intuitive. En effet, dans les années 80, le matériel disponible n'était pas assez rapide pour supporter des applications interactives, comme le rapportent [Sturman 1989]. Des approches plus simples ont donc été adoptées. Les interfaces basées sur un vocabulaire définissent un ensemble de gestes pour chaque opération, décrivant un langage. Elles sont efficaces, mais a) le nombre de gestes doit être limité en raison de la surcharge d'apprentissage, et b) les gestes doivent être correctement conçus pour prendre en compte le fonctionnement de la main et la fatigue. Une autre approche, relativement naturelle et puissante, consiste à combiner des gestes descriptifs et la parole pour créer des interactions multimodales intuitives [Bolt 1980, Koons 1994, Latoschik 1998]. Cependant, le modèle d'interaction le plus intuitif et naturel est la manipulation directe, avec laquelle les utilisateurs se comportent comme dans la vie réelle. De telles approches suscitent beaucoup d'intérêt car les progrès logiciels et matériels les rendent plus réalisables, et elles se sont avérées plus appropriées pour l'apprentissage actif [Bowman 2012]. Avant le développement de moteurs physiques largement disponibles, les designers optaient pour des approches heuristiques, fournissant des opérations logiques régissant l'interaction. Les simulateurs physiques ont amené la réalité virtuelle à un tout autre niveau, permettant un retour naturel et donc une interaction avec l'environnement vraiment intuitive. Ils ont aussi mené à des systèmes plus complexes puisque les représentations virtuelles de la main doivent être gérées minutieusement pour appliquer des forces externes réalistes aux objets virtuels de la scène. Les simulations physiques permettent des interactions plus réalistes, renforçant donc l'illusion de plausibilité. Par conséquent, elles sont fortement recommandées dans les applications nécessitant une sensation de présence forte, comme en réhabilitation. Enfin, un plus grand niveau d'immersion peut être atteint grâce au retour tactile. Il a été reporté par [Sturman 1989] que le retour visuel seul ne peut pas fournir suffisamment d'indices pour que les utilisateurs se sentent directement présents, et la qualité visuelle ne peut pas compenser une telle limitation. Cependant, les appareils haptiques disponibles à l'heure actuelle sont limités à des espaces de travail relativement petits et ne sont pas compatibles avec des espaces larges et complètement immersifs. Ils peuvent être utilisés avec des HMDs, mais la navigation physiques reste limitée. Une exception à cela est le SPIDAR extensible [], qui simule le retour de force dans des espaces de type CAVE. Cependant, la présence de câbles empêche une navigation complètement libre dans l'espace.

A.5 Evaluation de l'usabilité

Le fait de fournir de nouvelles solutions est bien pour faire naître de nouvelles idées et donc des interfaces innovantes. En revanche, évaluer les avantages et inconvénients de ces solutions est aussi très important pour décider des directions à prendre pour les futures recherches. De telles évaluations jaugent l'expérience des utilisateurs en termes d'immersion et d'usabilité, et l'efficacité de l'interface en termes de design et de performance atteignable. Elles combinent des mesures objectives et subjectives pour fournir des informations riches à propos de la légitimité de l'interface. Alors que les interactions naturelles permettent de meilleures compréhension spatiale et précision, sont plus intuitives et garantissent le transfert de l'entraînement dans le monde réel [Bowman 2012], le finger-tracking en espace immersif reste difficile, en raison de la nécessité de calibrer les appareils et de l'apparition de bruit dans le signal dû aux occlusions des repères par le corps de l'utilisateur ou par les murs. Au contraire, les appareils de type manette sont bien établis (conçus pour éviter les occlusions, signal fiable), et de telles interfaces hyper-naturelles augmentent les capacités humaines et donc l'interaction, facilitant les actions des utilisateurs [Bowman 2012].

A.6 Vue d'ensemble

Dans cette thèse, notre but est d'évaluer l'interaction gestuelle dans le contexte des espaces de réalité virtuelle complètement immersifs. Notre intuition est que la performance et l'expérience des utilisateurs sont directement liées aux mouvements requis par les tâches, du fait du mécanisme des membres humains et des outils d'interaction. Plus spécifiquement, nous comparons les interactions reposant sur des gestes naturels et sur une manette pour la manipulation directe. Nous proposons donc des designs expérimentaux impliquant les principaux mouvements quotidiens (Chapitre 4), et développons des appareils permettant d'évaluer l'influence des mouvements sur les tâches en les décomposant de manière contrôlée (Chapitre 5). Dans les deux cas, nous comparons les systèmes virtuels à un équivalent réel qui donne un aperçu des problèmes concernant le design expérimental et de l'interface. Nous testons aussi l'usage d'une interface gestuelle simple avec des personnes âgées dans le contexte de la thérapie par réminiscence (Chapitre 6).

Nous considérons donc ce qui suit comme des contributions de cette thèse:

- **Évaluation de la manipulation directe avec finger-tracking pour des tâches complexes en cube immersif:** Nous proposons une solution pour manipuler des objets virtuels 3D de manière quasi-naturelle pour des tâches générales dans un espace immersif cubique. Notre solution couple du finger-tracking avec un moteur physique temps-réel, combinés à une approche heuristique d'interaction manuelle, qui est robuste au bruit de tracking et aux instabilités de simulation. Nous évaluons notre interface à travers des manipulations relativement complexes, comme maintenir des objets en équilibre pendant que l'on marche dans le cube. La performance en utilisant le finger-tracking est comparée à la performance avec une manette à six degrés de liberté, et les tâches analysées sont aussi effectuées en conditions réelles. Nous avons aussi demandé à nos participants d'effectuer une tâche libre pour observer leur niveau de présence perçu dans la scène. À notre connaissance, c'est la première fois qu'un tel système est évalué dans le contexte de tâches d'ordre général en espace virtuel complètement immersif.

- **Évaluation contrôlée de la manipulation directe à travers la décomposition du mouvement:** Nous proposons un framework pour l'évaluation minutieuse de la manipulation directe d'objets virtuels 3D en espace complètement immersif. Nous nous intéressons spécifiquement à l'impact du type de mouvement sur la performance globale. Pour cela, nous avons conçu des appareils qui permettent de décomposer le mouvement selon sa nature (rotation et translation), mais aussi en degrés de liberté uniques et multiples. Nous évaluons la manipulation directe à travers des tâches de pointage et d'orientation, et comparons à la performance obtenue avec une manette à six degrés de liberté. Les appareils conçus pour la décomposition du mouvement ont aussi été répliqués en réel. À notre connaissance, c'est la première fois que la compréhension du mouvement décomposé pour la manipulation directe est évaluée en espace complètement immersif.

- **Étude de faisabilité sur l'utilisation de la réalité virtuelle pour la thérapie par réminiscence:** Nous présentons une nouvelle solution de réalité virtuelle pour la thérapie par réminiscence, qui est une intervention populaire dans le soin de la démence. Notre système immersif permet de facilement présenter des environnements familiers hautement réalistes en utilisant la technique innovante de rendu à base d'images de [Chaurasia 2014], et fournit une interaction naturelle avec les objets virtuels compris dans ces environnements. Nous évaluons l'efficacité de notre système à travers une étude utilisateur. Des participants âgés et sains sont soumis à un protocole de fluence verbale autobiographique adapté pour la réalité virtuelle, au cours duquel ils doivent générer des souvenirs en se basant sur les images qui leur sont montrées. Le système a été testé pour des environnements inconnu et familier afin d'évaluer sa capacité à évoquer l'aspect familier d'une scène donnée. À notre connaissance, c'est la première fois que l'utilisation de la réalité virtuelle avec rendu à base d'images est évaluée pour la thérapie par réminiscence.

Le reste de cette thèse est structuré de la façon suivante:

- Dans le Chapitre 2, nous discutons de l'état de l'art en réalité virtuelle en termes de technologie, d'évaluation et d'application des interfaces.
- Le Chapitre 3 présente le framework matériel et logiciel que nous utilisons dans nos expériences, ainsi que les défis soulevés par l'interaction gestuelle dans le contexte spécifique des espaces complètement immersifs.
- Le Chapitre 4 évalue une interface de manipulation directe pour des tâches générales relativement complexes. L'interface combine une approche heuristique et une simulation physique.
- Le Chapitre 5 évalue l'effet du mouvement sur la manipulation 3D à travers la décomposition du mouvement.
- Le Chapitre 6 combine un rendu hautement réaliste et une interaction naturelle pour évaluer l'usabilité de la réalité virtuelle dans la thérapie par réminiscence pour les personnes âgées.

- Dans le Chapitre 7, nous résumons les résultats de cette thèse et proposons des directions pour de prochains travaux.

APPENDIX B

FRANÇAIS
Résumé

Dans cette thèse, nous proposons trois études permettant de mettre en œuvre et d'évaluer une solution de manipulation directe, et nous la soumettons au test d'une application concrète dans le domaine clinique.

Ce chapitre présente un résumé de ces études, ainsi qu'une description rapide des dispositifs matériels et logiciels utilisés.

B.1 Dispositif

B.1.1 Matériel

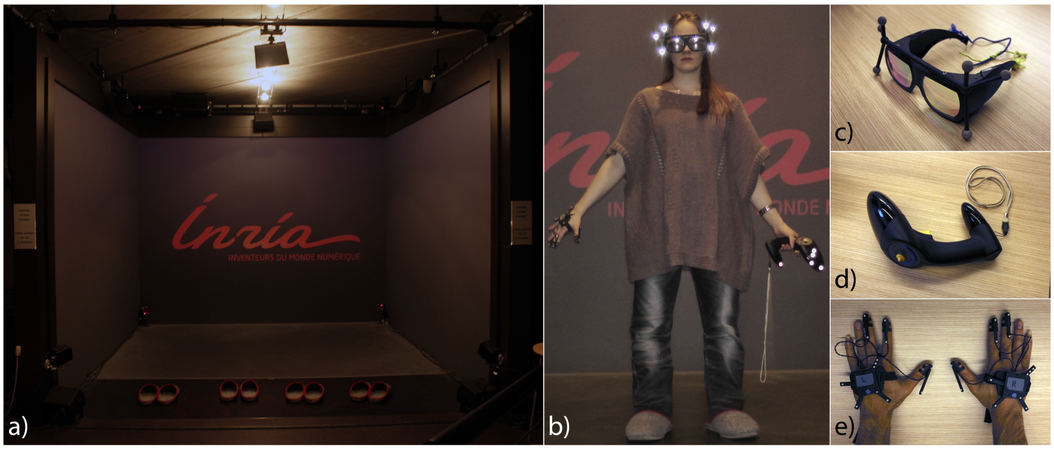


Figure B.1: *Installation expérimentale. Le cube immersif est visible en a) avec huit caméras infrarouges et un système de son spatialisé. Au milieu, b) un utilisateur porte les appareils détectés: c) les lunettes Infitec avec balises, d) la manette ART et e) les gants ART de suivi des doigts.*

Notre système de réalité virtuelle consiste en un cube immersif (Barco iSpace[1]) composé de quatre écrans rétro-projetés (trois murs et le sol) et équipé d'un système de suivi

[1] http://www.barco.com/en/products-solutions/visual-display-systems/3d-video-walls/multi-walled-stereoscopic-environment.aspx

par caméras infrarouges (technologie ART[2]) ainsi que d'un système de son spatialisé 6.1. La vision stéréoscopique est fournie grâce à la technologie Infitec[3].

Le système optique de suivi utilise huit caméras infrarouges, et récupère la position et l'orientation dans le cube d'une manette ART, des paumes et doigts de l'utilisateur grâce à des gants ART, et de la tête de l'utilisateur grâce à des balises montées sur les lunettes. Ce dernier suivi permet notamment d'adapter les images affichées au point de vue de l'utilisateur.

En théorie, six caméras devraient être suffisantes pour le suivi de l'utilisateur. En pratique, ce n'est pas suffisant pour le suivi des mains qui est sujet aux zones "d'ombre" où le signal est instable (bruit et coupures) en raison des occlusions des balises par l'utilisateur lui-même ou par la proximité des murs. Deux caméras ont donc été ajoutées. Les appareils de suivi des doigts sont calibrés individuellement pour chaque utilisateur, et le signal est filtré de manière spécifique. De plus, les utilisateurs bénéficient d'un retour visuel grâce à l'affichage d'une représentation virtuelle des paumes et doigts.

B.1.2 Logiciel

Comme une partie de la population est insensible à la vision stéréoscopique, et donc ne perçoit pas la profondeur par comparaison des images fournies par les deux yeux, nous proposons un test de perception. En effet, la perception de la profondeur est capitale pour effectuer les tâches virtuelles car il n'y a pas d'ajustement possible grâce au retour tactile. Nous affichons sur l'écran du fond un stéréogramme à points aléatoires. Il s'agit de nuages de points qui, lorsque regardés à travers des lunettes stéréoscopiques, produisent une sensation de profondeur, avec des objets apparaissant comme devant ou derrière l'écran. L'expérience ne continue que si ces objets sont perçus.

Notre système logiciel est basé sur la librairie de graphe de scène OpenSceneGraph[4], sur laquelle se greffe une couche logicielle développée en interne, et qui gère le système de suivi, la communication entre machines, et l'affichage stéréoscopique. Pour nos expériences, nous avons intégré le moteur physique libre Bullet[5] et l'avons adapté à nos besoins.

B.2 Évaluation d'une nouvelle solution pour la manipulation directe

Dans cette étude, nous analysons les avantages et inconvénients de l'interaction directe manuelle et avec manette dans le contexte des espaces immersifs larges. Nous présentons

[2] http://www.ar-tracking.com
[3] http://www.infitec.net/
[4] http://www.openscenegraph.org/
[5] http://www.bulletphysics.org

donc une solution de manipulation à partir de suivi des doigts dans un cube immersif et la comparons à la manipulation avec une manette traditionnelle à 6 degrés de liberté. Les utilisateurs peuvent donc interagir de manière quasiment naturelle pour effectuer des tâches d'ordre général modérément complexes.

Figure B.2: *Un utilisateur dans le cube immersif manipulant un plateau avec deux mains.*

Les buts de cette étude sont d'évaluer:

a) si l'interaction manuelle est une alternative réalisable aux interfaces traditionnelles en espace immersif pour des tâches d'ordre général modérément complexes,

a) les effets de l'interaction manuelle sur la présence,

a) la ressemblance avec la manipulation réelle.

B.2.1 Approche heuristique

La manipulation des objets virtuels est régie par un ensemble de règles logiques. Lorsqu'ils ne sont pas sélectionnés, les objets sont gérés par le moteur physique. Dès que les représentations virtuelles des doigts entrent en contact avec un objet, celui-ci est sélectionné et son comportement est déterminé de manière heuristique. L'utilisateur peut attraper, relâcher, translater et tourner les objets dans l'espace, et combiner ces actions pour effectuer des tâches plus complexes comme maintenir un objet en équilibre. Les objets peuvent être manipulés à une ou deux mains, et les transitions entre l'un et l'autre sont gérées par une machine à états.

B.2.2 Protocole expérimental

Nous avons effectué l'expérience avec 18 participants âgés entre 24 et 59 ans. La session se compose d'une étape de calibration de la distance interoculaire, d'une phase d'entraînement pour l'apprentissage des techniques d'interaction, d'un test d'utilisabilité, et d'une tâche libre. Le test d'utilisabilité est également effectué en conditions réelles, tandis que les autres phases ne sont réalisées qu'en conditions virtuelles, avec manette et avec mains (1 et/ou 2).

Le test d'utilisabilité consiste en la manipulation d'un plateau contenant des balles. L'utilisateur doit le soulever, le tourner et le faire passer entre des poteaux avant de le vider dans un saladier et de le ranger sur une étagère. Nous relevons le nombre de collisions du plateau et des mains avec les poteaux, ainsi que le nombre de balles perdues au cours de la manipulation. Le temps est également mesuré.

La tâche libre est une tâche d'observation en temps limité. Les utilisateurs disposent de 4 minutes pour mettre la table pour 6 personnes, et peuvent développer des stratégies d'utilisation des techniques pour optimiser leur performance. Leurs réactions ainsi que leur usage des techniques sont observés.

À la fin de l'expérience, les participants remplissent un questionnaire évaluant leur appréciation des techniques en termes de facilité d'utilisation, fatigue causée, sensation de présence, plausibilité de l'interaction, aspect naturel de l'interaction, ressemblance au monde réel, précision, rapidité, et sensation de nausée.

B.2.3 Résultats

Nos mesures ont montré qu'aucune des techniques virtuelles n'atteignait encore les performances du monde réel. Nous avons observé que les performances étaient meilleures avec une manette en termes de rapidité et de précision, mais que les utilisateurs se sentaient plus immergés dans le monde virtuel lorsqu'ils manipulaient avec les mains.

Nos résultats objectifs et subjectifs ont montré que l'interaction manuelle était une alternative réalisable aux techniques traditionnelles telles que les manettes. Nous avons aussi montré en particulier que la manipulation bi-manuelle favorisait la sensation de présence dans l'environnement virtuel et était perçue comme plus proche de la réalité.

B.3 Framework pour la décomposition du mouvement

Dans cette étude, nous analysons de manière contrôlée les rotations et translations mises en œuvre lors de la manipulation d'objets 3D. Nous nous intéressons en particulier aux interactions manuelle et avec manette dans le contexte des espaces complètement immersifs. Nous proposons donc un design expérimental dans lequel nous décomposons les actions des utilisateurs en degrés de liberté individuels et multiples grâce à des appareils conçus à cet effet. Nous étudions comment les utilisateurs comprennent et anticipent les mouvements isolés pour atteindre leurs objectifs de complexité croissante, et nous observons leur comportement face aux appareils réels et virtuels.

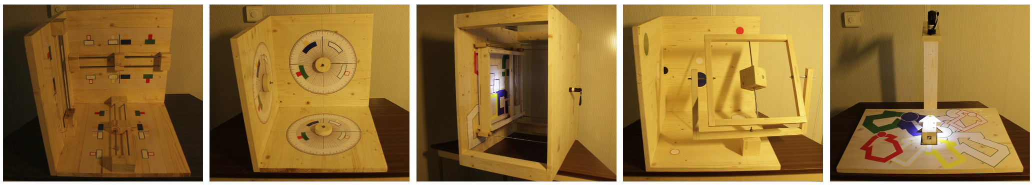


Figure B.3: *Les appareils utilisés pour décomposer le mouvement. De gauche à droite: translations 1D, rotations 1D, translations 3D, rotations 3D, mouvement libre.*

Nos buts sont d'effectuer une étude contrôlée des mouvements en 3D à travers leur décomposition, et de concevoir une expérience reproductible en conditions réelles et virtuelles.

B.3.1 Design expérimental

L'étude comporte 2 facteurs primaires: la technique utilisée (manuelle, avec manette, ou en condition réelle) et le type de mouvement (translations et rotations 1D individuelles suivant chaque axe, translations et rotations 3D individuelles, et mouvement libre ou translations et rotations 3D combinées); ainsi que 2 facteurs secondaires: la largeur et l'espacement des cibles.

Certaines lignes directrices ont été identifiées pour le design des mécanismes. Les appareils conçus doivent être robustes et simples à utiliser, avec un poids et une friction limitée pour préserver au maximum la similitude entre les conditions réelle et virtuelles, permettre d'effectuer des tâches rapides avec des amplitudes suffisantes pour varier la difficulté, et permettre un apprentissage progressif.

Les appareils conçus pour décomposer le mouvement sont basés sur des mécanismes existants simples et bien établis. Les translations 1D sont des glissières, les rotations 1D des boutons rotatifs, les translations 3D un système combinant des glissières, et les rotations 3D une adaptation d'un gyroscope. Le mouvement libre ne nécessite pas de mécanisme et consiste seulement en un objet non isomorphe à manipuler dans l'espace. Pour les mouvements 1D, l'utilisateur doit placer dans la cible une aiguille fixée à la partie mobile. Les translations 3D et le mouvement libre nécessitent de placer un objet de manière à ce que le contour de son ombre se trouve dans la cible. Enfin, les rotations 3D et le mouvement libre nécessitent de pointer un laser dans la cible.

B.3.2 Protocole expérimental

Nous avons effectué l'expérience avec 16 participants âgés entre 24 et 41 ans. La session expérimentale se compose d'une étape de calibration de la distance interoculaire, puis de la réalisation des tâches sur chacun des appareils, en conditions réelle et virtuelles. Les participants effectuent les mouvements 1D, 3D, puis libre, en alternant rotations et translations d'un utilisateur à l'autre. Les cibles sont des lots de zones de couleur, avec une couleur par niveau de difficulté, et deux zones par niveau: une pleine et une "creuse" (contour coloré uniquement). Les participants reçoivent l'instruction de déplacer l'objet de la cible pleine à la cible creuse, et vice versa, aussi rapidement et précisément que possible. À chaque positionnement, ils doivent appuyer sur un bouton pour valider leur action et passer à la suivante. Un retour sonore leur indique si le positionnement est correct. Pour chaque tâche, une phase d'apprentissage et d'entraînement est proposée.

Nous mesurons le temps pour chaque positionnement ainsi que le nombre d'erreurs. À la fin de l'expérience, les utilisateurs complètent un questionnaire évaluant leur tolérance à la réalité virtuelle, ainsi que leur appréciation des techniques proposées en termes de similitude avec la réalité, facilité d'utilisation, fatigue, motivation, précision et rapidité perçues.

B.3.3 Résultats

Nos résultats ont montré là encore que, globalement, la manette était plus performante que l'interaction manuelle en termes de rapidité et précision, mais que l'utilisation des mains semblait plus proche de la réalité. Ici aussi, les performances avec les techniques virtuelles ne sont pas aussi efficaces qu'en conditions réelles. Ceci n'a en revanche pas été le cas pour les translations 3D, dont le système a posé des problèmes mécaniques (difficulté de manipulation et problèmes de visibilité).

Nous avons également observé que les tâches effectuées en 1D suivaient le modèle de Fitts qui relie le temps de réalisation et le niveau de difficulté de la tâche.

B.4 Étude de faisabilité pour la thérapie par réminiscence

Dans cette étude, nous présentons une solution pour la thérapie par réminiscence en réalité virtuelle immersive, développée en collaboration avec une équipe de recherche clinique spécialisée dans le traitement de la mémoire et la maladie d'Alzheimer. Notre système permet de présenter facilement des environnements familiers grâce à une technique de rendu à base d'images hautement réaliste, et supporte une interaction gestuelle quasi-naturelle avec les objets présents dans l'environnement. Pour évaluer l'efficacité et l'utilité de notre système pour la thérapie par réminiscence, nous effectuons une étude avec des personnes âgées saines pour tester si notre système peut aider à générer des souvenirs autobiographiques.

Nous adaptons un protocole de fluence verbale autobiographique à notre contexte de réalité virtuelle, dans lequel les participants doivent générer des souvenirs en fonction des images qui leur sont montrées.

Figure B.4: *Gauche: un utilisateur dans un environnement familier rendu à partir d'images, manipulant une pomme virtuelle avec sa main droite. Droite: notre installation matérielle, adaptée pour les personnes âgées.*

Les objectifs de cette étude sont d'évaluer si notre système peut:

a) évoquer la familiarité des environnements représentés et

a) être utile à la thérapie par réminiscence de par le nombre de souvenirs générés.

B.4.1 Design expérimental

Nous présentons un système intégrant une technique de rendu à base d'images permettant une navigation libre dans l'environnement et la solution d'interaction manuelle présentée dans la première étude (Chap. 4). L'algorithme de rendu étant limité à un seul écran, nous n'utilisons que l'écran du fond de notre cube immersif. De plus, nos participants étant des personnes âgées (plus de 60 ans), des précautions particulières ont été prises: un tapis et des chaises ont été installés dans le cube, de façon à ce que le participant et le clinicien puissent s'asseoir, et n'aient pas besoin de porter de chaussons. L'immersion, en revanche, est renforcée grâce aux sons d'ambiance spatialisés.

Outre la manipulation d'objets virtuels, les utilisateurs peuvent interagir avec l'environnement en navigant dans l'espace grâce à une gestuelle simple et adaptée à la population ciblée.

B.4.2 Protocole expérimental

Nous avons effectué l'étude avec 13 participants âgés de plus de 60 ans. L'expérience se compose d'une exposition à quatre conditions suivie, chaque fois, d'une session de fluence verbale autobiographique. Les quatre conditions consistent en une condition contrôle constituée simplement d'un écran gris, d'une condition simulant un test de fluence autobiographique traditionnel présentant une photographie panoramique d'un environnement connu, et de deux conditions d'évaluation du système de réalité virtuelle représentant un environnement inconnu et un familier. Dans ces deux derniers, les participants commencent par manipuler des objets virtuels se trouvant à leur portée dans la scène, puis naviguent dans l'environnement en pointant la direction dans laquelle ils souhaitent se déplacer.

Nous mesurons le nombre de souvenirs évoqués pour chaque condition, et un questionnaire est rempli après chaque exposition. Ce questionnaire évalue la familiarité de l'environnement ainsi que le degré d'émotion, anxiété, motivation, sécurité et fatigue ressenties. Les participants remplissent également des questionnaires évaluant leur niveau de présence et sensation de nausée pour les conditions utilisant la réalité virtuelle.

B.4.3 Résultats

Nos mesures ont montré que le système de réalité virtuelle était bien toléré par les personnes âgées. Le nombre de souvenirs autobiographiques générés montre aussi que notre solution est réaliste et capable d'évoquer la familiarité d'un environnement donné car ces souvenirs sont plus nombreux pour l'environnement familier que pour l'environnement inconnu. Enfin, le nombre de souvenirs générés est similaire à celui obtenu lors de thérapies par réminiscence traditionnelles, ce qui laisse espérer des résultats et une efficacité encore meilleurs lorsque la solution pour 4 écrans sera développée et que les scènes seront peuplées. En effet, la photographie panoramique utilisée représente nécessairement plus d'éléments qu'un point de vue classique du fait de son large champ de vision.

APPENDIX C

FRANÇAIS
Conclusion et travaux futurs

Le but de cette thèse était d'évaluer les interactions basées sur les gestes dans le cadre des espaces de réalité virtuelle complètement immersifs. Nos contributions consistent en la proposition et l'évaluation d'une solution à la fois physique et heuristique pour la manipulation directe, la description d'un framework complet pour l'analyse et la décomposition du mouvement lors de manipulations 3D, et une étude de faisabilité sur l'utilisation de la réalité virtuelle et des gestes pour la thérapie par réminiscence.

C.1 Évaluation d'une nouvelle solution pour la manipulation directe

Notre solution de manipulation directe offre un moyen simple et quasiment naturel de manier des objets virtuels 3D pour des tâches d'ordre général et relativement complexes, dans des systèmes complètement immersifs de type CAVE. Elle combine une approche heuristique basée sur une machine à états finis et une simulation physique simple et en temps réel pour gérer les collisions. Notre interface de manipulation manuelle est robuste au bruit de tracking et aux instabilités de simulation. Elle permet aux utilisateurs d'attraper, de translater, de tourner et de relâcher des objets intuitivement dans l'espace, et de combiner ces manipulations pour effectuer des tâches complexes comme maintenir des objets en équilibre pendant que l'on marche. Nous avons évalué la faisabilité de la manipulation directe, et avons clairement démontré son utilité et ses avantages par rapport aux interfaces de type manette, en termes de présence et d'aspect naturel, même si la manette est plus efficace que la manipulation directe en termes de rapidité et de précision. Nos résultats pour les espaces complètement immersifs sont cohérents avec ceux rapportés par [Bowman 2012] : la manipulation directe pourrait être considérée comme plus appropriée pour des applications telles que les entraînements car l'expérience est plus proche du monde réel qu'avec la manette qui, en revanche, améliore les capacités d'interaction de l'utilisateur.

L'évaluation consistait en une étude utilisateur comprenant une tâche d'évaluation de l'usabilité pour l'analyse, et une expérience utilisateur libre pour l'observation. Notre approche de manipulation directe a été comparée à une interface avec manette et à une réplique en conditions réelles pour la tâche d'évaluation de l'usabilité. La tâche d'observation a clairement révélé la sensation de présence ressentie par les utilisateurs lors de l'usage de la manipulation directe car leurs réflexes étaient ceux du quotidien. En ce qui concerne l'analyse, nos résultats ont montré que les tâches virtuelles ont pris plus de temps que les tâches réelles, indiquant fortement que les techniques virtuelles ne sont pas encore au stade où elles peuvent concurrencer leur équivalent réel en termes de rapidité. Nous avons expliqué ceci par le manque de retour tactile dans les systèmes virtuels qui, en revanche, offre un guidage important en conditions réelles, et nous avons rappelé que les solutions actuelles de retour tactile ne sont pas satisfaisantes pour les systèmes immersifs.

C.1.1 Travaux futurs

Cette étude a révélé deux directions intéressantes pour de prochains travaux.

Premièrement, l'utilisation d'un modèle complet de main avec tracking de tous les doigts permettrait des manipulations plus complexes, et donc offrirait des expériences encore plus naturelles et immersives. Cela nous demanderait de modifier l'approche heuristique en conséquence, et, surtout, cela nécessiterait l'amélioration des technologies de tracking avant de devenir réalisable. Le but ultime serait d'intégrer un retour tactile quand la technologie pour les espaces immersifs aura été perfectionnée.

Deuxièmement, nous avons remarqué que notre étude impliquait de nombreux paramètres, tels que le mouvement en soi, la difficulté des tâches, les compétences requises, etc. Par conséquent, il serait intéressant d'effectuer des études spécifiques pour analyser l'impact de ces facteurs individuellement.

C.2 Framework pour la décomposition du mouvement

Notre framework nous permet d'étudier les mouvements 3D pour les interfaces de manipulation directe dans les espaces complètement immersifs. Nous proposons un design d'appareils permettant une décomposition minutieuse des mouvements selon leurs nature (i.e., rotation ou translation) et degrés de liberté (i.e., tâches 1D ou 3D). Les appareils sont simples et intuitifs, utilisant des mécanismes existants bien établis qui les rendent facilement reproductibles en réel et en virtuel, et qui permettent aux utilisateurs de se concentrer sur la tâche à accomplir plutôt que sur la compréhension de leur fonctionnement. Ils font intervenir des translations 1D, des rotations 1D, des translations 3D, des rotations 3D, et des mouvements 3D libres, chacun grâce à un appareil spécifique. Nous évaluons l'impact du type de mouvement sur la performance globale grâce à de simples tâches de pointage et d'orientation, et comparons à la performance obtenue avec une interface de type manette.

Comme dans notre expérience précédente, nos résultats montrent clairement que l'usage du finger-tracking augmente la sensation d'immersion alors que la manette surpasse la manipulation directe en termes de performance. Nous confirmons aussi qu'aucune des techniques virtuelles ne peut atteindre les capacités du système physique, même pour des tâches simples. Notre analyse a montré qu'augmenter la complexité de la tâche augmente la difficulté d'anticipation du mouvement requis. Nous avons aussi rapporté que le design de tels appareils nécessite de faire des compromis entre la visibilité, le poids et la robustesse qui affectent la performance. Enfin, nous avons démontré que nos appareils pour les translations et rotations 1D suivent le modèle de Fitts, et que les paramètres estimés pour le modèle étaient cohérents avec nos précédentes observations sur la performance des techniques.

C.2.1 Travaux futurs

Un premier et simple moyen d'améliorer le design serait d'automatiser les appareils physiques pour les mesures car cela fournirait des informations plus riches pour de prochaines analyses.

Une autre direction intéressante serait d'étudier l'effet et la pertinence d'une comparaison équitable entre les techniques puisque, pour l'instant, les appareils de finger-tracking présentent un bruit et des instabilités significatifs. Le modèle de bruit des appareils de finger-tracking pourrait être synthétisé et appliqué à la manette, et les deux signaux d'entrée pourraient ensuite être filtrés de manière similaire. Ceci rendrait les performances des deux appareils similaires, mais détériorerait l'un des outils. Cependant, cela donnerait une idée des avantages que pourrait avoir la manipulation directe sur la manette lorsque la technologie fournira un meilleur tracking.

Enfin, il serait intéressant de concevoir un appareil unique permettant n'importe quelle décomposition du mouvement. Un tel "super-appareil" pourrait coupler un mécanisme basé sur le gyroscope avec un système de glissières combinées pour manipuler un objet central qui comporterait trois lasers orientés de manière à former un repère lumineux. Des mécanismes de verrouillage/déverrouillage minutieusement conçus permettraient de n'autoriser que les degrés de liberté souhaités. Ceci permettrait n'importe quelle combinaison de degrés de liberté, et donc même des tâches 2D qui ne sont pas prises en compte dans notre étude.

C.3 Étude de faisabilité pour la thérapie par réminiscence

Notre système évalue la faisabilité de l'utilisation des technologies de réalité virtuelle pour la thérapie par réminiscence. Nous présentons une solution qui combine une présentation facile d'environnements virtuels hautement réalistes sur un écran stéréo, une interaction naturelle avec des objets 3D virtuels, et une navigation libre dans les environnements. Le système fournit des environnements virtuels proches du réel et généralisables, et le réalisme et l'interactivité permis sont cruciaux pour la motivation et l'implication des patients, et donc pour la réussite de la thérapie. Notre système implique un test verbal de fluence autobiographique adapté spécifiquement à la réalité virtuelle. Nous avons aussi adapté notre installation aux personnes âgées qui peuvent souffrir d'une mobilité réduite. Notre étude a révélé des taux d'acceptation et de tolérance élevés de la part des participants âgés.

Nous avons présenté aux utilisateurs des environnements virtuels familier et inconnu ainsi qu'une photographie classique d'un endroit connu. L'étude a révélé que le système était capable de susciter des souvenirs aussi efficacement que la thérapie traditionnelle, et qu'il peut évoquer l'aspect familier d'une scène donnée puisque plus de souvenirs ont été générés dans l'environnement familier. Le système peut donc stimuler des souvenirs conscients de la mémoire autobiographique. Ceci prouve que les systèmes de réalité virtuelle immersive combinant une interaction gestuelle proche du réel et un rendu hautement réaliste sont utiles dans le contexte de la thérapie par réminiscence.

C.3.1 Travaux futurs

Deux directions principales apparaissent pour de prochains travaux à partir de cette étude.

Le prochain objectif immédiat est d'utiliser le système dans un contexte clinique avec des patients souffrant de troubles de la mémoire légers à modérés. Dans ce but, le centre mémoire de Nice, en France, est en train d'adapter cette solution à une installation clinique car il est en train d'être équipé d'écrans 3D stéréoscopiques afin d'effectuer de telles expériences en conditions réelles.

En termes d'algorithme, une amélioration importante serait d'adapter le système à des espaces complètement immersifs à plusieurs écrans afin de fournir une vision panoramique, et d'ajouter des foules virtuelles pour peupler la scène et donc augmenter la sensation de présence.

C.4 Remarques finales

Dans toutes nos expérimentations, les résultats quantitatifs prouvent clairement que l'utilisation d'interfaces gestuelles est ressentie comme plus naturelle que les interfaces avec manette. Cependant, la manette surpasse clairement la manipulation directe en termes de rapidité et de précision. Les deux principales raisons à cela sont a) l'aspect technologique des appareils, et b) le caractère hyper-naturel des outils de type manette. Ceci montre que l'interface choisie doit toujours être adaptée aux besoins de l'application. Quand le quotidien est répliqué et que des objets communs sont impliqués, les interfaces gestuelles devraient être préférées, mais elles peuvent être couplées à des interfaces reposant sur des outils. D'un point de vue subjectif, nous avons observé des réactions très positives de la part de tous nos participants lors de l'interaction et de la manipulation d'objets virtuels avec leurs mains. Ceci s'est aussi avéré pour les participants âgés, ce qui montre clairement le potentiel de telles interfaces de manipulation directe qui offrent une interaction proche du réel.

Malheureusement, de telles études peuvent être très difficiles à effectuer à cause des aspects techniques liés aux contraintes des espaces complètement immersifs, tels que l'intégration d'une simulation physique ou le fait que le finger-tracking soit sujet au bruit. En effet, les espaces de type CAVE sont larges, fermés et sombres. Ceci soulève des problèmes de visibilité pour les systèmes basés sur la vision, et des problèmes non triviaux de portée et de complexité dans le placement de capteurs, dont résultent du bruit et des instabilités dans le signal. Nous pensons que c'est la principale raison pour laquelle nos résultats montrent un tel avantage de performance pour les interfaces de type manette, contrairement à ce que nous attendions. Cependant, nous pensons sincèrement que les progrès technologiques devraient résoudre ce problème.

Une autre contribution importante de ces travaux est la comparaison avec les conditions réelles, qui a été effectuée pour nos deux premières expériences. L'installation physique fournit une condition de contrôle pour l'expérience et une base solide pour notre recherche. De plus, cela a été vraiment intéressant d'observer et de comparer les réactions des gens face aux systèmes virtuel et physique, et de voir, par exemple, que les gens essaient automatiquement d'attraper les objets virtuels qui tombent comme dans la vie réelle. Nous pensons que des études approfondies de paradigmes d'interfaces qui soient reproductibles en virtuel et en réel sont une direction très prometteuse pour de futures recherches.

Bibliography

Personal publications

[Bousseau 2011] Adrien Bousseau, **Emmanuelle Chapoulie**, Ravi Ramamoorthi and Maneesh Agrawala. *Optimizing Environment Maps for Material Depiction*. Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering), vol. 30, no. 4, July 2011. (Not cited.)

[Cabral 2011] Marcio Cabral, Peter Vangorp, Gaurav Chaurasia, **Emmanuelle Chapoulie**, Martin Hachet and George Drettakis. *A Multimode Immersive Conceptual Design System for Architectural Modeling and Lighting*. Proceedings of IEEE 3DUI (technote), 2011. (Not cited.)

[Cirio 2012] Gabriel Cirio, Peter Vangorp, **Emmanuelle Chapoulie**, Maud Marchal, Anatole Lécuyer and George Drettakis. *Walking in a Cube: Novel Metaphors for Safely Navigating Large Virtual Environments in Restricted Real Workspaces*. IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Virtual Reality), vol. 18, no. 4, 2012 . (Cited on page 18.)

[Taffou 2012] Marine Taffou, **Emmanuelle Chapoulie**, Adrien David, Rachid Guerchouche, George Drettakis and Isabelle Viaud-Delmon. *Auditory-visual integration of emotional signals in a virtual environment for cynophobia*. Proceedings of Cybertherapy, September 2012. (Cited on page 22.)

[Chapoulie 2014a] **Emmanuelle Chapoulie**, Rachid Guerchouche, Pierre-David Petit, Gaurav Chaurasia, Philippe Robert and George Drettakis. *Reminiscence Therapy using Image-Based Rendering in VR*. Proceeding of the IEEE Virtual Reality Conference, 2014. (Not cited.)

[Chapoulie 2014b] **Emmanuelle Chapoulie**, Maud Marchal, Evanthia Dimara, Maria Roussou, Jean-Christophe Lombardo and George Drettakis. *Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube*. Virtual Reality Journal, 2014. (Not cited.)

References

[Accot 1997] Johnny Accot and Shumin Zhai. *Beyond Fitts' Law: Models for Trajectory-based HCI Tasks*. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, CHI '97, pages 295–302, New York, NY, USA, 1997. ACM. (Cited on pages 19 and 58.)

[Accot 1999] Johnny Accot and Shumin Zhai. *Performance Evaluation of Input Devices in Trajectory-based Tasks: An Application of the Steering Law*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '99, pages 466–472, New York, NY, USA, 1999. ACM. (Cited on page 19.)

[Achanta 2012] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk. *SLIC Superpixels Compared to State-of-the-Art Superpixel Methods*. IEE Trans. PAMI, vol. 34, no. 11, pages 2274–2282, 2012. (Cited on page 91.)

[Agarawala 2006] Anand Agarawala and Ravin Balakrishnan. *Keepin' it real: pushing the desktop metaphor with physics, piles and the pen*. CHI '06, pages 1283–1292, 2006. (Cited on page 13.)

[Aitken 1969] Robert C Aitken. *Measurement of feelings using visual analogue scales*. Proceedings of the royal society of medicine, vol. 62, no. 10, page 989, 1969. (Cited on page 99.)

[Banville 2012] Frédéric Banville and Pierre Nolin. *Using virtual reality to assess prospective memory and executive functions after traumatic brain injury*. Journal of Cybertherapy & Rehabilitation, vol. 5, no. 1, page 45, 2012. (Cited on pages 14, 23 and 24.)

[Barry 2009] Susan R. Barry. Fixing my gaze: A scientist's journey into seeing in three dimensions. Basic Books, 2009. (Cited on page 28.)

[Bekele 2013] Esubalew Bekele, Zhi Zheng, Amy Swanson, Julie Crittendon, Zachary Warren and Nilanjan Sarkar. *Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments*. IEEE Trans. VCG (VR'13), vol. 19, no. 4, pages 711–720, 2013. (Cited on pages 14 and 23.)

[Berna 2012] Fabrice Berna, Peter Schönknecht, Ulrich Seidl, Pablo Toro and Johannes Schröder. *Episodic autobiographical memory in normal aging and mild cognitive impairment: A population-based study*. Psychiatry research, vol. 200, no. 2, pages 807–812, 2012. (Cited on page 88.)

[Bluck 2003] Susan Bluck. *Autobiographical memory: Exploring its functions in everyday life*. Memory, vol. 11, no. 2, pages 113–123, 2003. (Cited on page 88.)

[Bolt 1980] Richard A. Bolt. *"Put-that-there": Voice and Gesture at the Graphics Interface*. In Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '80, pages 262–270, New York, NY, USA, 1980. ACM. (Cited on pages 3, 10 and 118.)

[Borst 2005a] Christoph W Borst and Arun P Indugula. *Realistic virtual grasping*. In Virtual Reality, 2005. Proceedings. VR 2005. IEEE, pages 91–98. IEEE, 2005. (Cited on page 56.)

[Borst 2005b] C.W. Borst and A.P. Indugula. *Realistic virtual grasping*. In IEEE VR 2005, pages 91–98, 2005. (Cited on page 13.)

[Bowman 2012] D.A. Bowman, R.P. McMahan and E.D. Ragan. *Questioning naturalism in 3D user interfaces*. Communications of the ACM, vol. 55, no. 9, pages 78–88, 2012. (Cited on pages 1, 3, 4, 8, 9, 10, 20, 54, 55, 58, 109, 115, 118, 119 and 131.)

[Brooks 2003] BM Brooks and FD Rose. *The use of virtual reality in memory rehabilitation: current findings and future directions*. NeuroRehabilitation, vol. 18, no. 2, pages 147–157, 2003. (Cited on pages 23, 24, 86 and 104.)

[Bruce 2009] Morgan Bruce and Holger Regenbrecht. *A virtual reality claustrophobia therapy system-implementation and test*. In Proc. IEEE VR, pages 179–182, 2009. (Cited on pages 14, 21 and 22.)

[Buchmann 2004] Volkert Buchmann, Stephen Violich, Mark Billinghurst and Andy Cockburn. *FingARtips: gesture based direct manipulation in Augmented Reality*. In Proc. GRAPHITE '04, pages 212–221, 2004. (Cited on pages 7, 9, 17 and 58.)

[Buehler 2001] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler and Michael Cohen. *Unstructured lumigraph rendering*. In SIGGRAPH, ACM Proc., pages 425–432, 2001. (Cited on pages 89 and 90.)

[Buoguila 2001] Laroussi Buoguila, Masahiro Ishii and Makoto Sato. *Scaleable spidar: a haptic interface for human-scale virtual environments*. In Haptic Human-Computer Interaction, pages 182–193. Springer, 2001. (Cited on pages 4 and 84.)

[Cabral 2005] Marcio C. Cabral, Carlos H. Morimoto and Marcelo K. Zuffo. *On the usability of gesture interfaces in virtual reality environments*. In Proc. CLIHC '05, pages 100–108, 2005. (Cited on page 7.)

[Chaurasia 2011] Gaurav Chaurasia, Olga Sorkine and George Drettakis. *Silhouette-Aware Warping for Image-Based Rendering*. Computer Graphics Forum, EGSR Proc., vol. 30, no. 4, pages 1223–1232, 2011. (Cited on page 89.)

[Chaurasia 2013] Gaurav Chaurasia, Sylvain Duchene, Olga Sorkine-Hornung and George Drettakis. *Depth synthesis and local warps for plausible image-based navigation*.

ACM Trans. on Graphics (TOG), vol. 32, no. 3, pages 30:1–30:12, 2013. (Cited on pages 86, 87, 89, 90, 91, 92, 93 and 97.)

[Chaurasia 2014] Gaurav Chaurasia. *Algorithmes et analyses perceptuelles pour la navigation interactive basée image*. These, Université Nice Sophia Antipolis, Feb 2014. (Cited on pages 5, 91 and 120.)

[Cohen-Mansfield 2009] Jiska Cohen-Mansfield, Maha Dakheel-Ali and Marcia S Marx. *Engagement in persons with dementia: the concept and its measurement*. The American journal of geriatric psychiatry, vol. 17, no. 4, page 299, 2009. (Cited on pages 105 and 106.)

[Conway 2000] Martin A Conway and Christopher W Pleydell-Pearce. *The construction of autobiographical memories in the self-memory system*. Psychological review, vol. 107, no. 2, page 261, 2000. (Cited on page 88.)

[Conway 2005] Martin A Conway. *Memory and the self*. Journal of memory and language, vol. 53, no. 4, pages 594–628, 2005. (Cited on page 86.)

[Corbett-Davies 2013] Sam Corbett-Davies, Andreas Dünser and Adrian Clark. *An advanced interaction framework for augmented reality based exposure treatment*. In Proc. IEEE VR, 2013. (Cited on pages 14 and 22.)

[Cruz-Neira 1993] Carolina Cruz-Neira, Daniel J Sandin and Thomas A DeFanti. *Surround-screen projection-based virtual reality: the design and implementation of the CAVE*. In SIGGRAPH, ACM Proc., pages 135–142, 1993. (Cited on pages 3, 14, 93 and 117.)

[Cunningham 2011] Douglas Cunningham and Christian Wallraven. Experimental design: From user studies to psychophysics. AK Peters, Ltd., 2011. (Cited on pages 51, 101 and 102.)

[Cutler 1997] Lawrence D Cutler, Bernd Fröhlich and Pat Hanrahan. *Two-handed direct manipulation on the responsive workbench*. In Proceedings of the 1997 symposium on Interactive 3D graphics, pages 107–114. ACM, 1997. (Cited on pages 11 and 40.)

[Czernuszenko 1997] Marek Czernuszenko, Dave Pape, Daniel Sandin, Tom DeFanti, Gregory L. Dawe and Maxine D. Brown. *The ImmersaDesk and Infinity Wall projection-based virtual reality displays*. ACM SIGGRAPH Computer Graphics, vol. 31, no. 2, pages 46–49, 1997. (Cited on pages 14 and 25.)

[Debevec 1996] Paul E Debevec, Camillo J Taylor and Jitendra Malik. *Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach*. In SIGGRAPH, ACM Proc., pages 11–20, 1996. (Cited on page 89.)

[Dipietro 2008] Laura Dipietro, Angelo M. Sabatini and Paolo Dario. *A Survey of Glove-Based Systems and Their Applications*. IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 38, no. 4, pages 461–482, 2008. (Cited on page 14.)

[Dritschel 1992] Barbara H Dritschel, JMG Williams, Alan D Baddeley and I Nimmo-Smith. *Autobiographical fluency: A method for the study of personal memory*. Memory & Cognition, vol. 20, no. 2, pages 133–140, 1992. (Cited on pages 87, 88 and 96.)

[Dubois 2000] B Dubois, A Slachevsky, I Litvan and B Pillon. *The FAB A frontal assessment battery at bedside*. Neurology, vol. 55, no. 11, pages 1621–1626, 2000. (Cited on page 97.)

[Eisemann 2008] Martin Eisemann, Bert De Decker, Marcus Magnor, PhilippeBekaert, Edilson de Aguiar, Naveed Ahmed, Christian Theobalt and Anita Sellent. *Floating Textures*. Computer Graphics Forum, Eurographics Proc., vol. 27, no. 2, pages 409–418, 2008. (Cited on pages 89 and 90.)

[Fitts 1954] Paul M Fitts. *The information capacity of the human motor system in controlling the amplitude of movement.* Journal of experimental psychology, vol. 47, no. 6, page 381, 1954. (Cited on pages 18, 58 and 75.)

[Folstein 1975] Marshal F Folstein, Susan E Folstein and Paul R McHugh. *“Mini-mental state": a practical method for grading the cognitive state of patients for the clinician*. Journal of psychiatric research, vol. 12, no. 3, pages 189–198, 1975. (Cited on pages 96 and 97.)

[Fröhlich 2000] Bernd Fröhlich, Henrik Tramberend, Andrew Beers, Maneesh Agrawala and David Baraff. *Physically-Based Manipulation on the Responsive Workbench*. In IEEE VR 2000, 2000. (Cited on page 13.)

[Furukawa 2009] Yasutaka Furukawa and Jean Ponce. *Accurate, Dense, and Robust Multi-View Stereopsis*. IEEE Trans. PAMI, vol. 32, no. 8, pages 1362–1376, 2009. (Cited on pages 89 and 91.)

[Gardiner 1998] John M Gardiner, Cristina Ramponi and Alan Richardson-Klavehn. *Experiences of remembering, knowing, and guessing*. Consciousness and Cognition, vol. 7, no. 1, pages 1–26, 1998. (Cited on page 96.)

[Ghez 1991] Claude Ghez, Wayne Hening and James Gordon. *Organization of voluntary movement*. Current opinion in neurobiology, vol. 1, no. 4, pages 664–671, 1991. (Cited on page 61.)

[Goesele 2010] Michael Goesele, Jens Ackermann, Simon Fuhrmann, Carsten Haubold, Ronny Klowsky, Drew Steedly and Richard Szeliski. *Ambient point clouds for view interpolation*. ACM Trans. on Graphics (TOG), vol. 29, no. 4, page 95, 2010. (Cited on page 89.)

[Gonneaud 2012] Julie Gonneaud, Pascale Piolino, Grégory Lecouvey, Sophie Madeleine, Eric Orriols, Philippe Fleury, Francis Eustache and Béatrice Desgranges. *Assessing prospective memory in young healthy adults using virtual reality*. In Disability, Virtual Reality, and Associated Technologies, Proc. of the 9th Int. Conf., pages 211–218, 2012. (Cited on pages 23, 24 and 86.)

[Griffith 2012] James W Griffith, Jennifer A Sumner, Filip Raes, Thorsten Barnhofer, Elise Debeer and Dirk Hermans. *Current psychometric and methodological issues in the measurement of overgeneral autobiographical memory*. Journal of behavior therapy and experimental psychiatry, vol. 43, pages S21–S31, 2012. (Cited on page 88.)

[Guiard 1987] Yves Guiard. *Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model*. Journal of motor behavior, vol. 19, no. 4, pages 486–517, 1987. (Cited on page 11.)

[Guiard 1999] Yves Guiard, Michel Beaudouin-Lafon and Denis Mottet. *Navigation As Multiscale Pointing: Extending Fitts' Model to Very High Precision Tasks*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '99, pages 450–457, New York, NY, USA, 1999. ACM. (Cited on page 18.)

[Hilliges 2012] Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss and Andrew Wilson. *HoloDesk : Direct 3D Interactions with a Situated See-Through Display*. In CHI '12, pages 2421–2430, 2012. (Cited on pages 11, 13, 36, 37 and 56.)

[Hirota 2003] K. Hirota and M. Hirose. *Dexterous object manipulation based on collision response*. In IEEE VR '03, volume 2003, pages 232–239. IEEE Comput. Soc, 2003. (Cited on pages 12 and 17.)

[Hoffman 2000] Hunter G Hoffman, David R Patterson and Gretchen J Carrougher. *Use of virtual reality for adjunctive treatment of adult burn pain during physical therapy: a controlled study*. The Clinical Journal of Pain, vol. 16, no. 3, pages 244–250, 2000. (Cited on page 16.)

[Holz 2008] Daniel Holz, Sebastian Ullrich, Marc Wolter and Torsten Kuhlen. *Multi-Contact Grasp Interaction for Virtual Environments*. Journal of Virtual Reality and Broadcasting, vol. 5, no. 7, pages 1860–2037, 2008. (Cited on pages 13 and 32.)

[Jacobs 2011] J. Jacobs and B. Froehlich. *A soft hand model for physically-based manipulation of virtual objects*. In IEEE VR 2011. IEEE, 2011. (Cited on pages 13 and 20.)

[Jacobs 2012] Jan Jacobs, Michael Stengel and Bernd Froehlich. *A generalized God-object method for plausible finger-based interactions in virtual environments*. In 3DUI'2012, pages 43–51. Ieee, March 2012. (Cited on pages 7, 13, 20, 32 and 36.)

[Julesz 1971] Bela Julesz. Foundations of cyclopean perception. U. Chicago Press, 1971. (Cited on page 28.)

[Kibby 1998] Michelle Y Kibby, Maureen Schmitter-Edgecombe and Charles J Long. *Ecological validity of neuropsychological tests: focus on the California Verbal Learning Test and the Wisconsin Card Sorting Test*. Archives of Clinical Neuropsychology, vol. 13, no. 6, pages 523–534, 1998. (Cited on pages 105 and 106.)

[Klinger 2013] Evelyne Klinger, Abdelmajid Kadri, Eric Sorita, J-L Le Guiet, Pauline Coignard, Philippe Fuchs, Laure Leroy, Nicolas du Lac, Fabrice Servant and P-A Joseph. *AGATHE: A tool for personalized rehabilitation of cognitive functions based on simulated activities of daily living*. IRBM, vol. 34, no. 2, pages 113–118, 2013. (Cited on page 21.)

[Koons 1994] David B. Koons and Carlton J. Sparrell. *Iconic: speech and depictive gestures at the human-machine interface*. In Conference Companion on Human Factors in Computing Systems, CHI '94, pages 453–454, New York, NY, USA, 1994. ACM. (Cited on pages 3, 10 and 118.)

[Krüger 1994] Wolfgang Krüger, Christian-A. Bohn, Bernd Fröhlich, Heinrich Schüth, Wolfgang Strauss and Gerold Wesche. *The Responsive Workbench*. IEEE Computer Graphics and Applications, vol. 14, pages 12–15, 1994. (Cited on pages 3, 14, 40 and 117.)

[Lacquaniti 1992] F Lacquaniti, NA Borghese and M Carrozzo. *Internal models of limb geometry in the control of hand compliance*. The Journal of neuroscience, vol. 12, no. 5, pages 1750–1762, 1992. (Cited on page 61.)

[Latoschik 1998] M.E. Latoschik, M. Frohlich, B. Jung and I. Wachsmuth. *Utilize speech and gestures to realize natural interaction in a virtual environment*. In Industrial Electronics Society, 1998. IECON '98. Proceedings of the 24th Annual Conference of the IEEE, volume 4, pages 2028 –2033 vol.4, aug-4 sep 1998. (Cited on pages 3, 10 and 118.)

[Legault 2013] Isabelle Legault, Rémy Allard and Jocelyn Faubert. *Healthy older observers show equivalent perceptual-cognitive training benefits to young adults for multiple object tracking*. Frontiers in psychology, vol. 4, pages 1–7, 2013. (Cited on pages 7 and 25.)

[Massie 1994] Thomas H Massie and J Kenneth Salisbury. *The phantom haptic interface: A device for probing virtual objects*. In Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems, volume 55, pages 295–300. Chicago, IL, 1994. (Cited on page 12.)

[McGee 2000] Jocelyn S McGee, Cheryl Van der Zaag, J Galen Buckwalter, Marcus Thiébaux, Andre Van Rooyen, Ulrich Neumann, D Sisemore and Albert A Rizzo.

Issues for the assessment of visuospatial skills in older adults using virtual environment technology. CyberPsychology & Behavior, vol. 3, no. 3, pages 469–482, 2000. (Cited on pages 2, 21, 25 and 116.)

[Meehan 2001] Michael Meehan. *Physiological reaction as an objective measure of presence in virtual environments*. PhD thesis, University of North Carolina at Chapel Hill, 2001. (Cited on page 16.)

[Miller 1984] Edgar Miller. *Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology*. British Journal of Clinical Psychology, vol. 23, no. 1, pages 53–57, 1984. (Cited on page 88.)

[Moehring 2010] Mathias Moehring and Bernd Froehlich. *Enabling functional validation of virtual cars through natural interaction metaphors*. In Virtual Reality Conference (VR), 2010 IEEE, pages 27–34. IEEE, 2010. (Cited on pages 10, 11, 20 and 40.)

[Moehring 2011] M. Moehring and B. Froehlich. *Natural Interaction Metaphors for Functional Validations of Virtual Car Models*. IEEE TVCG, vol. 17, no. 9, pages 1195–1208, 2011. (Cited on pages 10, 12, 17, 20, 36 and 58.)

[Nancel 2011] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis and Wendy Mackay. *Mid-air Pan-and-zoom on Wall-sized Displays*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pages 177–186, New York, NY, USA, 2011. ACM. (Cited on page 18.)

[Nguyen 2014] Quan Nguyen and Michael Kipp. *Orientation Matters: Efficiency of Translation-rotation Multitouch Tasks*. In Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pages 2013–2016, New York, NY, USA, 2014. ACM. (Cited on pages 18, 19, 78 and 83.)

[O'Hagan 2002] R.G. O'Hagan, A. Zelinsky and S. Rougeaux. *Visual gesture interfaces for virtual environments*. Interacting with Computers, vol. 14, no. 3, pages 231 – 250, 2002. (Cited on pages 7, 9 and 17.)

[Optale 2010] Gabriele Optale, Cosimo Urgesi, Valentina Busato, Silvia Marin, Lamberto Piron, Konstantinos Priftis, Luciano Gamberini, Salvatore Capodieci and Adalberto Bordin. *Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study*. Neurorehabilitation and neural repair, vol. 24, no. 4, pages 348–357, 2010. (Cited on page 25.)

[Ortega 2007] Michael Ortega, Stephane Redon and Sabine Coquillart. *A Six Degree-of-Freedom God-Object Method for Haptic Display of Rigid Bodies with Surface Properties*. IEEE TVCG, vol. 13, no. 3, pages 458–469, May 2007. (Cited on page 12.)

[Pair 2006] Jarrell Pair, Brian Allen, Matthieu Dautricourt, Anton Treskunov, Matt Liewer, Ken Graap and Greg Reger. *A virtual reality exposure therapy application for Iraq war post traumatic stress disorder*. In Proc. IEEE VR '06, pages 67–72, 2006. (Cited on pages 14 and 22.)

[Plancher 2010] Gaën Plancher, Valerie Gyselinck, Serge Nicolas and Pascale Piolino. *Age effect on components of episodic memory and feature binding: A virtual reality study*. Neuropsychology, vol. 24, no. 3, page 379, 2010. (Cited on pages 23 and 24.)

[Plancher 2012] G Plancher, A Tirard, V Gyselinck, S Nicolas and P Piolino. *Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and Alzheimer's disease: Influence of active and passive encoding*. Neuropsychologia, vol. 50, no. 5, pages 592–602, 2012. (Cited on pages 24 and 104.)

[Plancher 2013] Gaën Plancher, Julien Barra, Eric Orriols and Pascale Piolino. *The influence of action on episodic memory: a virtual reality study*. The Quarterly Journal of Experimental Psychology, vol. 66, no. 5, pages 895–909, 2013. (Cited on pages 24 and 86.)

[Ponto 2013] Kevin Ponto, Michael Gleicher, Robert G Radwin and Hyun Joon Shin. *Perceptual calibration for immersive display environments*. Visualization and Computer Graphics, IEEE Transactions on, vol. 19, no. 4, pages 691–700, 2013. (Cited on pages 45 and 73.)

[Prachyabrued 2012] M. Prachyabrued and C.W. Borst. *Visual interpenetration tradeoffs in whole-hand virtual grasping*. In 3DUI, pages 39–42. IEEE, 2012. (Cited on pages 13, 32 and 37.)

[Pugnetti 1998] Luigi Pugnetti, Laura Mendozzi, Elizabeth A Attree, Elena Barbieri, Barbara M Brooks, Carlo L Cazzullo, Achille Motta, F David Rose and C Psychol. *Probing memory and executive functions with virtual reality: Past and present studies*. CyberPsychology & Behavior, vol. 1, no. 2, pages 151–161, 1998. (Cited on page 21.)

[Ramstein 1994] Christophe Ramstein and Vincent Hayward. *The pantograph: a large workspace haptic device for multimodal human computer interaction*. In Conference companion on Human factors in computing systems, pages 57–58. ACM, 1994. (Cited on page 12.)

[Regenbrecht 2002] Holger Regenbrecht and Thomas Schubert. *Real and illusory interactions enhance presence in virtual environments*. Presence: Teleoperators and virtual environments, vol. 11, no. 4, pages 425–434, 2002. (Cited on page 100.)

[Riva 1997] Giuseppe Riva. Virtual reality in neuro-psycho-physiology: Cognitive, clinical and methodological issues in assessment and rehabilitation. IOS press, 1997. (Cited on page 20.)

[Rizzo 1998] Albert A Rizzo, J Galen Buckwalter, Ulrich Neumann, Carl Kesselman and Marcus Thiebaux. *Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities*. CyberPsychology & Behavior, vol. 1, no. 1, pages 59–78, 1998. (Cited on page 21.)

[Rizzo 2005] Albert "Skip" Rizzo and Gerard Jounghyun Kim. *A SWOT analysis of the field of virtual reality rehabilitation and therapy*. Presence: Teleoperators and Virtual Environments, vol. 14, no. 2, pages 119–146, 2005. (Cited on page 20.)

[Robert 2002] P. H Robert, S Clairet, M Benoit, J Koutaich, C Bertogliati, O Tible, H Caci, M Borg, P Brocker and P Bedoucha. *The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment*. International journal of geriatric psychiatry, vol. 17, no. 12, pages 1099–1105, 2002. (Cited on page 97.)

[Robert 2003] Philippe H Robert, Stéphane Schuck, Bruno Dubois, Jean Pierre Olié, Jean Pierre Lépine, Thierry Gallarda, Sylvia Goni and Sylvie Troy. *Screening for Alzheimer's disease with the short cognitive evaluation battery*. Dementia and geriatric cognitive disorders, vol. 15, no. 2, pages 92–98, 2003. (Cited on page 97.)

[Rose 2005] F David Rose, Barbara M Brooks and Albert A Rizzo. *Virtual reality in brain damage rehabilitation: review*. CyberPsychology & Behavior, vol. 8, no. 3, pages 241–262, 2005. (Cited on pages 2, 20 and 116.)

[Schubert 2001] Thomas Schubert, Frank Friedmann and Holger Regenbrecht. *The experience of presence: Factor analytic insights*. Presence: Teleoperators and virtual environments, vol. 10, no. 3, pages 266–281, 2001. (Cited on pages 17, 100 and 103.)

[Seidl 2011] Ulrich Seidl, Ulrike Lueken, Philipp A Thomann, Josef Geider and Johannes Schröder. *Autobiographical memory deficits in Alzheimer's disease*. Journal of Alzheimer's Disease, vol. 27, no. 3, pages 567–574, 2011. (Cited on page 88.)

[Shum 2007] Heung-Yeung Shum, Shing-Chow Chan and Sing Bing Kang. Image-based rendering. Springer, 2007. (Cited on pages 87 and 89.)

[Slater 1995] Mel Slater, Martin Usoh and Anthony Steed. *Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality*. ACM Transaction on Computer-Human Interaction, vol. 2, no. 3, pages 201–219, 1995. (Cited on page 16.)

[Slater 2009a] M. Slater. *Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments*. Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 364, no. 1535, pages 3549–3557, 2009. (Cited on pages 2, 15, 36, 55, 58 and 116.)

[Slater 2009b] Mel Slater, Pankaj Khanna, Jesper Mortensen and Insu Yu. *Visual realism enhances realistic response in an immersive virtual environment*. Computer Graphics and Applications, IEEE, vol. 29, no. 3, pages 76–84, 2009. (Cited on pages 16 and 87.)

[Snavely 2006] Noah Snavely, Steven M. Seitz and Richard Szeliski. *Photo tourism: exploring photo collections in 3D*. ACM Trans. on Graphics (TOG), vol. 25, no. 3, pages 835–846, 2006. (Cited on page 91.)

[Stich 2011] Timo Stich, Christian Linz, Christian Wallraven, Douglas Cunningham and Marcus Magnor. *Perception-motivated interpolation of image sequences*. ACM Trans. on Applied Perception, vol. 8, no. 2, pages 11:1–11:25, February 2011. (Cited on page 89.)

[Stoelen 2010] Martin F Stoelen and David L Akin. *Assessment of Fitts' Law for quantifying combined rotational and translational movements*. Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 52, no. 1, pages 63–77, 2010. (Cited on pages 18, 78 and 83.)

[Sturman 1989] D. J. Sturman, D. Zeltzer and S. Pieper. *Hands-on interaction with virtual environments*. UIST '89, pages 19–24, 1989. (Cited on pages 3, 4, 7, 9, 12, 13, 14, 117 and 118.)

[Sturman 1994] David J. Sturman and David Zeltzer. *A Survey of Glove-based Input*. IEEE Comput. Graph. Appl., vol. 14, no. 1, pages 30–39, January 1994. (Cited on page 14.)

[Sutherland 1968] Ivan E Sutherland. *A head-mounted three dimensional display*. In Proceedings of the December 9-11, 1968, fall joint computer conference, part I, pages 757–764. ACM, 1968. (Cited on pages 2, 3, 7, 13 and 117.)

[Thompson 2007] Edmund R Thompson. *Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS)*. Journal of Cross-Cultural Psychology, vol. 38, no. 2, pages 227–242, 2007. (Cited on page 100.)

[Tsirlin 2010] Inna Tsirlin, Eve Dupierrix, Sylvie Chokron, Theophile Ohlmann and Sabine Coquillart. *Multimodal virtual reality application for the study of unilateral spatial neglect*. In Proc. IEEE VR, pages 127–130, 2010. (Cited on page 23.)

[Ullmann 2000] Thomas Ullmann and Joerg Sauer. *Intuitive Virtual Grasping for non Haptic Environments*. In Pacific Graphics '00, pages 373–381, 2000. (Cited on pages 11, 12 and 32.)

[Van der Linde 2002] Richard Q Van der Linde, Piet Lammertse, Erwin Frederiksen and B Ruiter. *The HapticMaster, a new high-performance haptic interface*. In Proc. Eurohaptics, pages 1–5, 2002. (Cited on page 12.)

[Vangorp 2013] Peter Vangorp, Christian Richardt, Emily A. Cooper, Gaurav Chaurasia, Martin S. Banks and George Drettakis. *Perception of Perspective Distortions in Image-Based Rendering*. ACM Trans. Graph. (SIGGRAPH Conference Proceedings), vol. 32, no. 4, pages 58:1–58:12, July 2013. (Cited on page 90.)

[Viaud-Delmon 2000] Isabelle Viaud-Delmon, Yuri P Ivanenko, Alain Berthoz and Roland Jouvent. *Adaptation as a sensorial profile in trait anxiety: a study with virtual reality*. Journal of anxiety disorders, vol. 14, no. 6, pages 583–601, 2000. (Cited on pages 17, 74 and 100.)

[Wang 2009] Robert Y. Wang and Jovan Popović. *Real-time Hand-tracking with a Color Glove*. In ACM SIGGRAPH 2009 Papers, SIGGRAPH '09, pages 63:1–63:8, New York, NY, USA, 2009. ACM. (Cited on page 15.)

[Watson 1988] David Watson, Lee A Clark and Auke Tellegen. *Development and validation of brief measures of positive and negative affect: the PANAS scales.* Journal of personality and social psychology, vol. 54, no. 6, page 1063, 1988. (Cited on page 100.)

[Weiss 2004] Patrice L Weiss, Debbie Rand, Noomi Katz and Rachel Kizony. *Video capture virtual reality as a flexible and effective rehabilitation tool*. Journal of neuroengineering and rehabilitation, vol. 1, no. 1, page 12, 2004. (Cited on page 21.)

[Weiss 2006] Patrice L Weiss, Rachel Kizony, Uri Feintuch and Noomi Katz. *Virtual reality in neurorehabilitation*. Textbook of neural repair and neurorehabilitation, vol. 2, pages 182–197, 2006. (Cited on pages 2, 20 and 116.)

[Wexelblat 1995] Alan Wexelblat. *An approach to natural gesture in virtual environments*. ACM Trans. Comput.-Hum. Interact., vol. 2, no. 3, pages 179–200, September 1995. (Cited on page 36.)

[Whitman 1970] Richards Whitman. *Stereopsis and stereoblindness*. Experimental Brain Research, vol. 10, no. 4, pages 380–388, 1970. (Cited on page 28.)

[Wilson 2008] Andrew D. Wilson, Shahram Izadi, Otmar Hilliges, Armando Garcia-Mendoza and David Kirk. *Bringing Physics to the Surface*. In ACM UIST '08, pages 67–76, 2008. (Cited on page 13.)

[Woods 2005] Bob Woods, A Spector, C Jones, M Orrell and S Davies. *Reminiscence therapy for dementia*. Cochrane Database Systematic Reviews, vol. 2, 2005. (Cited on page 86.)

[Yeh 2009] Shih-Ching Yeh, B. Newman, M. Liewer, Jarrell Pair, A. Treskunov, G. Reger, B. Rothbaum, J. Difede, J. Spitalnick, Rob McLay, T. Parsons and A. Rizzo. *A Virtual Iraq System for the Treatment of Combat-Related Posttraumatic Stress Disorder*. In Proc. IEEE VR, pages 163–170, 2009. (Cited on page 23.)

[Yu 2012] Insu Yu, Jesper Mortensen, Pankaj Khanna, Bernhard Spanlang and Mel Slater. *Visual Realism Enhances Realistic Response in an Immersive Virtual Environment-Part 2*. Computer Graphics and Applications, IEEE, vol. 32, no. 6, pages 36–45, 2012. (Cited on pages 16 and 105.)

[Zhai 1993] Shumin Zhai. *Investigation of feel for 6DOF inputs: isometric and elastic rate control for manipulation in 3D environments*. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 37, pages 323–327. SAGE Publications, 1993. (Cited on page 9.)

[Zhai 1995] Shumin Zhai. *Human performance in six degree of freedom input control*. PhD thesis, University of Toronto, 1995. (Cited on page 8.)

[Zhai 1998] Shumin Zhai and Paul Milgram. *Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices*. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 320–327. ACM Press/Addison-Wesley Publishing Co., 1998. (Cited on page 17.)

Gestes et Manipulation Directe pour la Réalité Virtuelle Immersive

Résumé:

La réalité virtuelle est une technologie qui voit ses applications s'étendre à de nombreux domaines (médical, automobile, etc.).

Cette thèse se place dans le contexte des espaces virtuels complètement immersifs, et a pour but d'étudier les effets des deux principaux types d'interfaces proposés (manette avec 6 degrés de liberté, et système de suivi de doigts) sur l'expérience des utilisateurs, dans le cadre de la manipulation d'objets 3D. Nous nous intéressons à des paramètres tels que la facilité d'utilisation, la sensation d'immersion, la rapidité et la précision offertes... Pour cela, nous proposons des expériences évaluant ces paramètres à travers des tâches dont le succès est mesurable, et qui ne sont pas spécifiques à un domaine.

Dans une première étude, nous nous intéressons aux tâches complexes d'ordre général, faisant appel à des compétences requises dans les manipulations quotidiennes, telles que le fait d'attraper, de relâcher, de translater, de tourner et de maintenir en équilibre des objets tout en se déplaçant. Nous affinons ensuite notre étude en observant les effets de ces interfaces sur les mouvements eux-mêmes, en les décomposant en degrés de liberté individuels et groupés. Enfin, nous testons l'applicabilité de notre système de manipulation directe dans le cadre d'une étude préliminaire sur l'utilisation de la réalité virtuelle pour le traitement de la maladie d'Alzheimer.

Ces études analysent les propriétés de ces interfaces dans le but de fournir des indications aidant au choix de l'interface la plus appropriée pour des applications futures.

Mots-clés: réalité virtuelle, espaces immersifs, manipulation directe, interfaces naturelles, évaluation d'interfaces.

Gestures and Direct Manipulation for Immersive Virtual Reality

Abstract:

Virtual reality is a technology with applications in numerous fields (medical, automotive, etc.).

This thesis focuses on immersive virtual spaces, and aims at studying the effects of the two major types of interfaces proposed (6 degree of freedom flystick, and finger-tracking system) on the user experience, and specifically for 3D object manipulation. We are interested in parameters such as ease of use, sense of presence, speed and precision offered. To do so, we design experiments to evaluate these parameters via tasks with measurable success, and which are not field specific.

In a first experiment, we study complex general purpose tasks, combining skills required in everyday manipulations, such as grabbing, releasing, translating, rotating, and balancing objects while walking. We then refine our study by observing the effects of those interfaces on the movements themselves, by decomposing them into individual and grouped degrees of freedom. Lastly, we evaluate the applicability of our direct manipulation system in the context of a preliminary study on the use of virtual reality for the treatment of Alzheimer's disease.

These studies analyze the properties of these interfaces to provide guidelines to the choice of the most appropriate interface for future experiments.

Keywords: virtual reality, immersive spaces, direct manipulation, natural interfaces, interface evaluation.