Collaboration in Multi-user Immersive Virtual Environment

Weiya Chen




NNT:2015SACLS248

THESE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PREPAREE A
L'UNIVERSITÉ PARIS-SUD

ECOLE DOCTORALE N° 580 STIC
Sciences et Technologies de l'Information et de la Communication

Spécialité de doctorat: **Informatique**

Par :

M. Weiya Chen

Titre de la thèse :

Collaboration in Multi-user Immersive Virtual Environments

Thèse présentée et soutenue à Orsay, le mardi 15 décembre 2015

Composition du jury :

Président :	Michel Beaudouin-Lafon	Professeur (Université Paris-Saclay)
Rapporteurs :	Thierry Duval	Professeur (Télécom Bretagne)
	Betty Mohler	Chercheur (Max Planck Institute)
Examinateurs :	Michel Beaudouin-Lafon	Professeur (Université Paris-Saclay)
	Victoria Interrante	Professeur (University of Minnesota)
Directeur de thèse :	Patrick Bourdot	Directeur de Recherche (LIMSI-CNRS)
Invité :	Samir Otmane	Professeur (Université d'Evry-Val-d'Essonnes)

Groupe RV&A VENISE
LIMSI-CNRS
B.P. 133
91403 Orsay Cedex, France

STIC - ED580
UFR Sciences - Université Paris-Saclay
Bat 650 PCRI bureau 417 - Rue Noetzlin
91403 Orsay Cedex, France

Acknowledgement

It has been more than six years that I live in France, more specifically, in the suburban areas of Paris. Six years ago, I was a student facing a whole new world, curious of everything that is different from the east.

Now as I'm at the point of finishing my PhD work, I'd like to take a moment to have a small break and look back on my life in France and my research in Virtual Reality. Living in Paris is an amazing experience, here you can find different facets of life, joy, sorrow, love, thought, etc. I feel their love for life by counting the number of cheese and wine, I see their bravery by seeing how they face death and terror.

It was great to do research at VENISE team, to meet people sharing the same passion. First, I'd like to thank Patrick Bourdot, my PhD advisor, head of VENISE team, for accepting and helping me to conduct my PhD work all along these years. He gave me a lot of useful advices and guidance from paradigm study to experiment design, and he is always available during difficult moments. I also want to thank the two Nicolas of our team: Nicolas Ladeveze and Nicolas Férey for their help and great discussions that I had with them, also Céline Clavel and Julien Nelson, without whom I won't be able to achieve various publications. Of course, I won't forget the great cartoonist of our team - Jean-Marc Vézien, who gave me lots of hardware and English supports, also many useful advices.

Then I want to thank my colleagues and friends in LIMSI lab, Nicolas, Olivier and Jean-Claude of AMIC team, Bénédicte, Sophie and many other staff that helped me one way or another. I also want to thank other big boys at VENISE, Pierre, Mikael, Xavier, Alexandre, Yujiro and Hang, with whom I had amazing time together.

At last, I want to thank my love Tianqiao, for supporting me all the time, and taking care of our two babies with my mother while I was not there. I know I owe you a lot and my gratitude is beyond words, this thesis is dedicated to them.

The spring silk worm
will weave its thread until it dies;

The wax candle
will shed tears until it turns to ashes.

- To my family

Abstract

Immersive virtual environments can be used to bring both geographically distributed and co-located users to the same virtual place for collaboration. Compared to remote situations, co-located users collaborate in the same virtual world on top of a shared physical workspace. This collocation allows direct user communication and interaction without computer mediation which facilitates collaborative work.

With the development of multi-user display and tracking technology, classical projection-based immersive setups (e.g. CAVE) can now support group immersion for co-located users by offering individual stereoscopic views without visual distortion. In this context, the coexistence of information from the virtual and real world, especially when users do not share a common spatial reference frame, provides users with a new kind of perceptual and cognitive experience. We are interested in how users perceive and communicate with each other to achieve a shared context for collaboration, and how we can broaden supported collaborative scenarios with more flexible viewpoint control.

This PhD thesis mainly addresses perceptual and cohabitation issues that we identified in the aim of supporting safe and efficient co-located collaboration in immersive virtual environments. First, we conducted a case study to examine how perceptual conflicts would alter user communication and task performance. Second, we concentrated on the design and evaluation of appropriate navigation paradigms to allow individual virtual navigation while solving cohabitation problems in a shared limited physical workspace. At last, based on the results of previous studies, we designed a generic dynamic navigation model which integrates constraints from the physical workspace and also the virtual world to enable co-located collaboration in multi-user immersive systems.

Keywords: Virtual Reality, Co-located Collaboration, Navigation, Perception, Cohabitation.

Résumé

Les Environnements Virtuels Immersifs (EVIs) peuvent être utilisés pour amener des utilisateurs, répartis géographiquement ou co-localisés, à partager un même monde virtuel pour collaborer. Si l'on compare aux situations distantes, les utilisateurs d'une immersion co-localisée collaborent aussi dans le monde virtuel, mais a contrario, partagent physiquement un même espace de travail. Cette co-localisation facilite le travail collaboratif en permettant des communications directes et des interactions sans médiation informatique entre les utilisateurs.

Avec le développement de l'affichage multi-utilisateur et de la technologie de tracking, les dispositifs immersifs classiques basés sur la rétroprojection (ex. CAVE) peuvent offrir maintenant l'immersion pour plusieurs utilisateurs co-localisés en affichant différentes vues stéréoscopiques sans distorsion visuelle pour chacun d'eux. Dans ce contexte, la coexistence de l'information du monde virtuel et réel, en particulier lorsque les utilisateurs ne partagent pas un référentiel spatial commun, offre aux utilisateurs une nouvelle expérience perceptive et cognitive. Dans cette thèse nous nous sommes intéressés à la façon dont les utilisateurs se perçoivent et communiquent entre eux pour atteindre un contexte commun pour la collaboration, et aux moyens permettant d'élargir des scénarios collaboratifs déjà pris en charge dans ce type de dispositifs, basés sur des techniques de contrôle plus flexible des points de vue des utilisateurs.

Cette thèse de doctorat traite donc principalement des problèmes perceptifs et de cohabitation que nous avons identifiés dans l'objectif d'assurer la sécurité et l'efficacité des collaborations co-localisées dans les environnements virtuels immersifs. Tout d'abord, nous avons mené une étude de cas pour examiner comment les conflits perceptifs modifieraient la communication entre les utilisateurs et leur performance. Deuxièmement, nous avons conçu et évalué des paradigmes de navigation appropriés pour permettre la navigation virtuelle individuelle tout en résolvant les problèmes de la cohabitation dans un espace de travail partagé physiquement limité. Enfin, sur la base des résultats de ces travaux, nous avons proposé un modèle dynamique générique qui intègre des contraintes de l'espace de travail physique et aussi ceux du monde virtuel pour gérer la collaboration co-localisée dans les systèmes immersifs multi-utilisateurs.

Mots-clef: Réalité Virtuelle, Collaboration Co-localisée, Navigation, Perception, Cohabitation.

Contents

Introduction

Context and Motivation

Immersive Collaborative Virtual Environment (CVE) is developing as a convergence of research interests from Virtual Reality (VR) and Computer Supported Collaborative Work (CSCW) communities with its capacity to offer high level multi-sensory immersion for networked users. Immersive CVEs provide better support for social human communications and let users feel "being together" in the same virtual world.

Users of immersive CVEs can be geographically distributed or co-located in the same physical workspace. Compared to remote situations, co-located users collaborate in the same virtual world on top of a shared physical workspace. This physical collocation forms a mixed context where user can have direct as well as computer-mediated interaction and communication. While lots of research works focus on supporting remote collaboration by connecting several immersive systems, co-located collaboration for now receives limited attention due to the rareness of multi-user immersive systems.

With the development of multi-user display and tracking technology, classical projection-based immersive setups (e.g. CAVE) can now support group immersion for co-located users by offering individual stereoscopic views without visual distortion. In this context, the coexistence of information from the virtual and real world, especially when users do not share a common spatial reference frame, provides users with a new kind of perceptual and cognitive experience.

This PhD thesis mainly addresses perceptual and cohabitation issues that we identified in the aim of supporting safe and efficient co-located collaboration in multi-user immersive virtual environments. We are interested in how users perceive and communicate with each other to achieve a shared context for collaboration, and how we can broaden supported collaborative scenarios with more flexible viewpoint control. Moreover, as users share the same limited physical workspace, it is necessary to find solutions to properly allocate enough workspace for each user to assure their safety and immersion level while allowing them to navigate in unlimited virtual worlds.

Organization of Thesis

This thesis contains two main parts which are divided into five chapters: Chapter 1 and Chapter 2 present the general background and our contributions to co-located collaboration in terms of concepts, identified issues and relevant research methods, then Chapter 3 to 5 describe in detail our research work consisting of paradigm design and user experiments.

Chapter 1 describes the general context of this thesis by presenting notions and state-of-the-art in Virtual Reality (VR) and Computer Supported Collaborative Work (CSCW), and different topics related to using immersive virtual environments for collaboration.

Chapter 2 concentrates on how co-located users collaborate inside multi-user immersive virtual environments. It begins by presenting multi-user display technology and basic notions about users' spatial organization and virtual navigation, then talks about two major issues identified during the collaboration process and research methodology that we adopted.

Chapter 3 then describes a case study on the perceptual conflicts caused by dual-presence of users during co-located collaboration in a multi-stereoscopic immersive virtual environment. We examined how perceptual conflicts would alter user communication and task performance.

Chapter 4 presents the implementation of Altered Human Joystick along with metrics that we defined to evaluate cohabitation capacity in response to identified cohabitation problems. A series of experiments were carried out to test different combinations of alterations to see their impacts on users' spatial distribution in the physical workspace and navigation experience in the virtual world.

Finally Chapter 5 presents concepts and implementations of a novel navigation model named DYNAMIC. This novel model manages user cohabitation by integrating constraints from physical workspace and allows interactive navigation control by including constraints from the virtual world, all in aim to better support collaboration in multi-user immersive virtual environments.

1

Background Study

Chapter Abstract

This chapter introduces the background of this thesis work by giving a general presentation on notions and current state-of-the-art in the domain of Virtual Reality (VR) and Computer Supported Collaborative Work (CSCW), and also by presenting research questions and existing work on how users collaborate in immersive collaborative virtual environments (CVEs).

Contents

1.1 Introduction

Virtual reality (VR) technology provides users with new interaction possibilities by offering perceptual immersion and real time interaction through sensorimotor interfaces in a virtual space. VR technology has been applied in various domains from industrial product design to training, from education to entertainment. With rapid increasing computation capacity and high-speed computer network, computer-generated virtual environments can be interconnected to achieve group immersion. So both co-located and remote users can be connected to a same virtual environment to accomplish collaborative tasks.

This chapter presents the general background of this thesis by discussing different topics related to collaboration in immersive virtual environments. It is divided into three main sections. The first explores key components of virtual reality technology, from its definition to the current state-of-the-art of advances in technical and social aspects. The second presents research around human collaborative work and how they could be supported by modern computer systems. It also shows issues and challenges of collaborative virtual environments (CVEs), which is a virtual environment shared by networked users. The last section addresses how networked immersive experience changes the way users communicate and interact with each other and summarizes existing work from a user-centered perspective.

1.2 Virtual Reality

1.2.1 Definition

Virtual reality related devices appeared before the term itself was created. The *Sensorama Machine* invented in 1957 and patented in 1962 by Morton Heiling believes to be the first multi-sensorial application (Figure 1.1-left). It is a simulator which provides the illusion of reality using a 3-D motion picture with smell, stereo sound, vibrations of the seat, and wind in the hair to create the illusion[1]. Then in 1968, Ivan Sutherland with the help of his student Bob Sproull created the first virtual reality and augmented reality (partly see-through) headset named "The Sword of Damocles" (Sutherland, 1968) (Figure 1.1-right).

The term "Virtual reality" in its modern usage was coined and popularized by Jaron Lanier (Lanier, 1992) through his company VPL Research in 1980's. Early definitions of virtual reality are often device-oriented which refer to a certain type of sophisticated computer equipment serving as a medium to connect users to a digitally created space. In the meanwhile, researchers began to discuss and establish conceptual framework of virtual reality, switching from device-oriented definitions to user experience based understanding. Especially, the term "presence" or "telepresence" was introduced to describe the generic perception of being in an artificial environment (Sheridan, 1992), which is one of the key notions of VR and will be discussed more thoroughly later on. More elaborated discussions about the history and definition of virtual reality can be found in Rheingold (1991)'s book and Steuer et al. (1995)'s paper.

Nowadays, a relatively complete definition of virtual reality was given by another book of Fuchs et al. (2011): "Virtual reality is a scientific and technical domain that uses computer science and behavioral interfaces to simulate in a virtual world the behavior of 3D entities, which

[1] http://www.mortonheilig.com/InventorVR.html

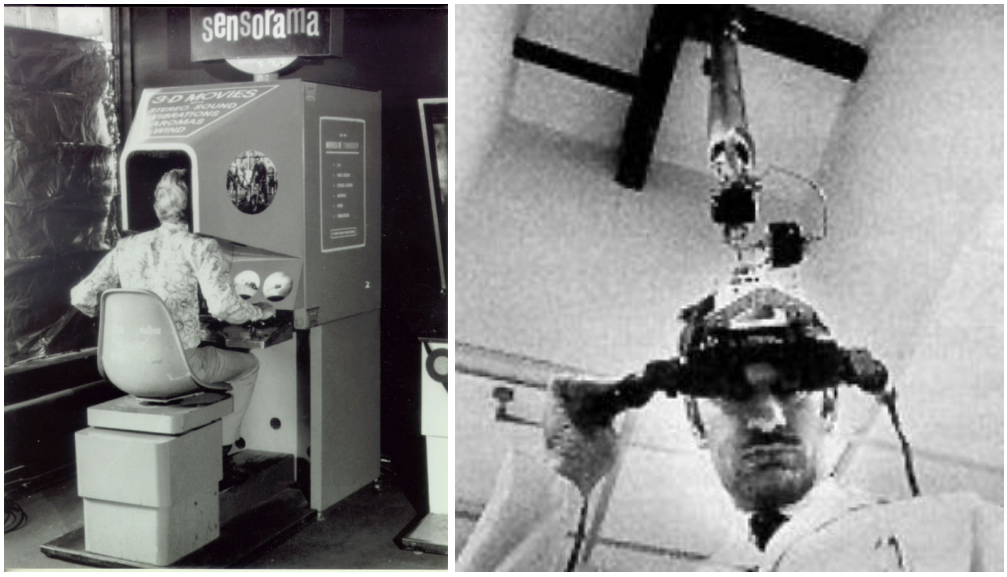


Figure 1.1 – First virtual reality applications: Sensorama Machine by Morton Heiling (left); The first head mounted display by Ivan Sutherland (right).

interact in real time with each other and with one or more users in pseudo-natural immersion via sensorimotor channels."

This definition covers three important aspects of virtual reality:

First, virtual reality is both a technical and scientific domain. On one hand, technological progress makes unthinkable experiences become available and provides a hardware platform on which theoretical research of virtual reality is based. For example, CAVE and Head-Mounted Display bring a higher level of visual immersion so that users can actually "step in" the virtual world and have the sensation of "being at another place". Real-time tracking devices enable motion capture so that users can interact with computers using natural gestures and other types of body movement. On the other hand, researchers make further investigations of the cause and influencing factors of user's subjective feelings using existing devices, and in return, provide guidelines for the development of future virtual reality systems.

Virtual reality is an extension of human computer interaction (HCI) as we move from traditional desktop to 3D user interface, many of its research methods are inherited or inspired by HCI research work. However, unlike HCI which is design-oriented and tries to find efficient, ergonomic and aesthetic interfaces for interaction, virtual reality aims to "remove" this interface and let users interact naturally with the virtual world. Virtual reality groups research efforts from various domains as it relies on different technical science domains (e.g. computer science, robotics and automatics etc.) and also on contributions from human and behavioral science (e.g. cognitive psychology, physiology, neurobiology, etc.).

Second, virtual reality possesses two key characteristics that distinguish itself with other existing technologies related to 3D virtual environments: sensorial immersion and real time interaction. In desktop or console based video games, players enjoy real time interaction in the virtual world alone or with other players. They use input devices like joystick, mouse and keyboard that offer very limited sensorial immersion. On the contrary, 3D movies projected on large screens in the cinema give the audience good visual immersion, but no interaction capability. The emergence of virtual reality technology brings great changes to the way we

are connected to the virtual world. Having the feeling of being immersed in another world and being able to interact in real time in that world as naturally as in the real world is quite appealing and may change completely the game and movie industries. Of course, the application of virtual reality technology is not limited to these domains, it also has great impact on education, product design, training, etc. (Hale and Stanney, 2014).

Last and the most important, it reveals three main components of virtual reality system: the virtual environment, the user, and the behavioral interface allowing them to interact. The "perception, decision and action" loop (Fuchs et al., 2011) describes how users interact with the surrounding virtual world, which is a transposition of the "perception, cognition, action" loop demonstrating man's behavior in a real world. As shown in Figure 1.2, the user acts (vocal commands, gestures and other body movements, etc.) on the virtual world through motor interfaces, and these activities are captured and transferred to the computer system. Then the computer system makes corresponding changes to the virtual world and generates sensorial reactions (images, sound, haptic, etc.) that are transferred to the user via sensorial interfaces.

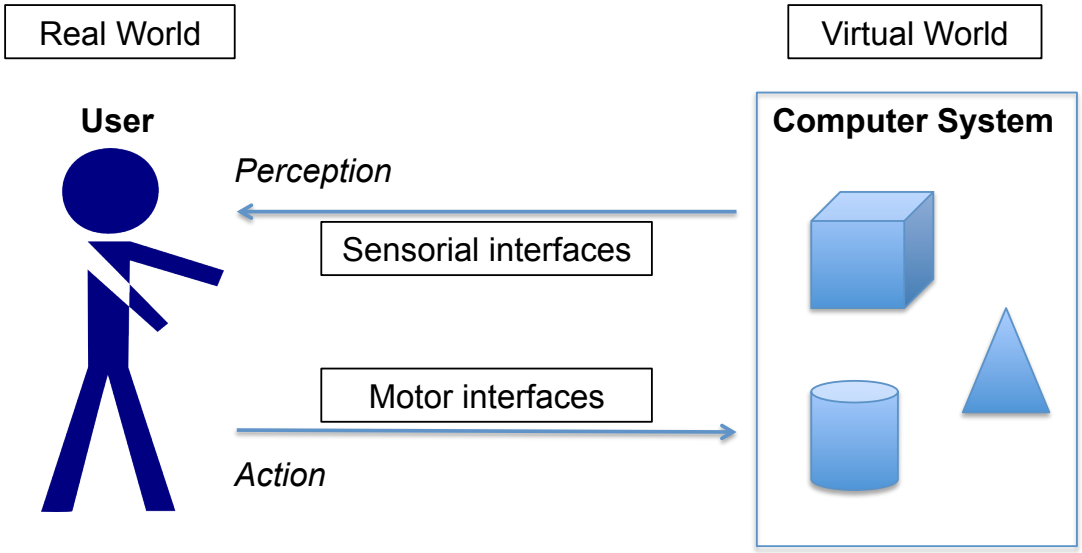


Figure 1.2 – The "perception, decision and action" loop in interactive virtual environments.

Since the virtual world is totally generated by the computer system, two inherent issues need to be solved: the latency and the sensorimotor discrepancies. The latency is the time lag between the user's action on motor interfaces and the perception of the consequences of this action on the virtual environment through sensorial interfaces. This artifact influences every virtual reality application and may be the source of many problems related to user comfort. The sensorimotor discrepancy is another problem for virtual reality applications. With technical progression, we can simulate more complex phenomena and provide more interaction through sensorimotor interfaces in the virtual environment, but the real world offers way more information for us to simulate with current technology and the sensorimotor discrepancies will continue to exist.

Now we will discuss in more details about the three components of virtual reality system mentioned above: the virtual environment, the behavioral interfaces and the human inside.

1.2.2 Virtual Environment

Virtual environment is a term often used to describe computer-generated synthetic space, similar to virtual world. However, more strictly speaking, as Ellis (1991b) explained, an environment is the theater of human activity, so virtual environment is a human-centered notion, not only refers to the digital data that forms an artificial world. For example, according to Fox et al. (2009): "A virtual environment is a digital space in which a user's movements are tracked and his or her surroundings rendered, or digitally composed and displayed to the senses, in accordance with those movements." This definition emphasizes that a virtual environment is a virtual space that can react to user's movements and change accordingly.

Ellis (1991b) summarized three parts which form a virtual environment: content, geometry and dynamics. The content is often organized as scene graph with clear hierarchical structure and inter relationship between objects. Each object possesses a state vector containing properties such as position, orientation, velocity and color, etc. The geometry is a description of an environmental field of action that has dimensionality, metrics and extent, and the dynamics of an environment are the rules of interaction among its contents describing their behavior, such as physical laws (gravity, object collision, etc.). More details on the description of virtual environment can be found in Ellis (1991a)'s book.

In virtual reality applications, the diversity of origins of the worlds represented in the virtual environment makes virtual reality more than a simple copy of the "reality" that we live in. As summarized by Fuchs et al. (2011), the virtual world could be a simulation of certain aspects of the real world, but also a completely symbolic or imaginary world. The visual representation of objects could vary depending on our need, and spatial, temporal and physical laws may or may not be applied in the virtual world. For example, we can visualize and interact with structure of molecules, flow of fluids under different scales (Férey et al., 2009; Bryson, 1996), we can also change the speed of light to better visualize and understand phenomena of relativity (Doat et al., 2011), or the entire world could just be a figment of imagination of an artist or a science-fiction writer.

1.2.3 Behavioral Interfaces

As mentioned in Fuchs et al. (2011)'s definition of virtual reality, behavioral interface is a term to describe a type of interfaces that connect the user to the virtual environment. Unlike traditional human computer interfaces which act as a communication tool (e.g. keyboard), behavioral interfaces use the motricity or perceptions of human resulting from his/her behavior in the real world to carry out activities in a virtual world. As shown in Figure 1.2, two types of behavioral interface exist: sensorial interfaces are designed to transfer sensory stimuli from computer to the user while the motor interfaces transfer motor responses from the user to the computer, thus we can also call them sensorimotor interfaces. One of the current challenges of virtual reality is to build more efficient and transparent interfaces based on human sensory channels and motor activities.

1.2.3.1 Visual Interface

As the idiom "Seeing is believing" states, the visual sense is nearly the most important sensory channel that humans use to discover the real world, which is also true in virtual reality applica-

tions. The software and hardware advancements in computer graphics constantly improve the quality of real time three-dimensional images, which allows more complex and photo-realistic rendering. The display used as visual interface should have the following characteristics:

- Large field of view (FoV) ("borderless");
- Stereoscopic vision in the entire binocular field of vision;
- High graphic resolution (pixel density), currently most widely used is 1080p FullHD display, but displays with higher resolution are on the way;
- Head (even eye) tracking compatible. Since we can move our head and eyes to look into different directions, it is necessary to assure that user always has perspective-correct view of the virtual scene. This can be achieved by using tracking devices[2]

A lot of display systems exist nowadays which satisfy all or part of the characteristics listed above, from large immersive rooms and wall displays to table-sized screen, and even portable screens mounted on user's head (see Figure 1.3). They vary in size, form, display technology and aimed applications. Only CAVE (Figure 1.3d) and HMD (Figure 1.3e) are designed to provide fully visual immersion. Other displays, on the contrary, can not assure a user to always stay visually inside the virtual world, but they can offer good visual immersion and are better suited for data visualization and 3D object interaction in certain cases. There are also some special display solutions not mentioned above that can provide individual stereoscopic views of 3D models such as the volumetric display (Grossman and Balakrishnan, 2008) and holographic display (Lucente, 1997), but currently they can only be applied to non-immersive context with contents that are difficult to be changed in real time.

1.2.3.2 Audio Interface

Audio cues are very important for us to perceive events happen in the real world, similarly, the virtual world would be much less appealing if there is no sound. A virtual object would be more "realistic" if there is a corresponding spatialized sound combined with its visual representation displayed in stereoscopy. For example, a virtual telephone displayed in front of a user with a ringtone coming exactly from that location would make the user believe that the telephone is ringing. 3D audio sometimes has more contributions to the immersion level because it allows users to perceive objects behind them and also objects that are beyond their vision scope. A virtual environment filled with ambiance noises and spatialized sound coming from different virtual entities can enhance user's feeling of "being in that world".

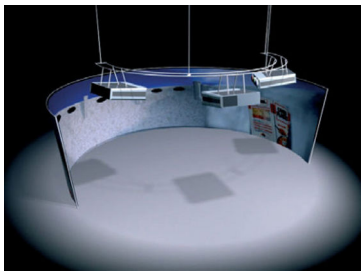
Audio interface has two types of roles: (1) to capture and transfer the sound made by the user (speech, hand clapping, etc.) to the computer system; (2) to transfer audio stimuli from the virtual environment (and from other users connected to the same VE) to the user. The implementation of binaural audio interface managing spatialized sound requires quite complex hardware and software setup which involves lots of research and engineering efforts (Begault et al., 1994).

Currently three solutions exist: first, ambisonic sound can be reproduced by a group of loud-speakers situated around the user. This kind of setup can be used to simulate ambient noises or

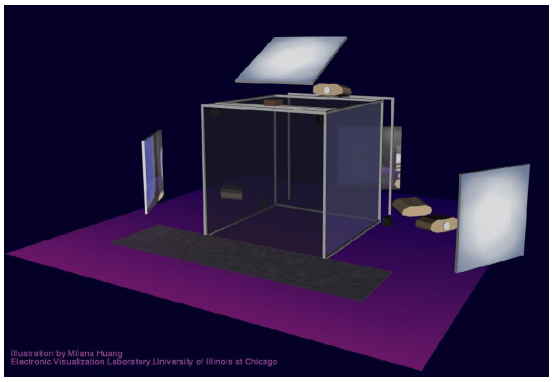
[2] see section 1.2.3.3.

(a) Workbench formed by two separate screens (LIRIS-CNRS).

(b) Large image wall composed of high resolution touch screens (WILDER, in|situ|, LRI).



(c) The curved panoramic projection screen i-Cone™ (Simon and Göbel, 2002).



(d) CAVE™: Cave Automatic Virtual Environment (Cruz-Neira et al., 1992).

(e) Head-Mounted Display (HMD): Oculus DK2 © Oculus VR LLC.

Figure 1.3 – Different types of visual interfaces.

audio sources coming from places farther than the perimeter of the loudspeaker group; second, based on Head-Related Transfer Function (HRTF) (Kistler and Wightman, 1992) which is a response that characterizes how an ear receives a sound from a point in space, binaural sound can be reproduced for multiple users by equipping each of them a stereo headphone combined with head tracking; at last, we can use a matrix of micro loudspeakers to reproduce physically the acoustic field based on Wave Field Synthesis (Verheijen, 1998), which is independent with listener's position. The headphone-based binaural sound is a more portable and lightweight solution while the other two methods require more complex hardware setup and are more suitable for large acoustic rooms or theaters. Moreover, the Wave Field Synthesis method can not be combined with retro-projection based immersive display as the loudspeaker matrix behind the screen would occlude projected images.

1.2.3.3 Tracking Interface

The goal of tracking interface is to capture and transfer user's motor information as input for the computer system to update the virtual world and to generate proper sensorial stimuli (visual, audio, etc.) for the user. Tracking an entity in 3D space requires six degrees of freedom (DoF) information: three for position in form of a vector (x, y, z) and three for orientation in form of Euler angle $(\theta x, \theta y, \theta z)$.

In many virtual reality applications, it is sufficient to track user's head and dominant hand to respectively ensure correct visual rendering according to user's viewpoint and to allow basic

interaction (e.g. object selection) in the virtual world. However, full-body movement tracking (motion capture) is increasingly demanded to enrich interaction scenarios and to enhance the level of immersion. Moreover, when multiple users are inside the same virtual environment, full-body tracking combined with decent avatar representations largely facilitate social human communication (which will be further discussed in section 1.4.2) by transferring subtle non-verbal cues (postures and gestures) among users. Recently, eye tracking technology has received a lot of attentions for its potential benefits in many domains (Duchowski, 2007) and it will become an important part of tracking interface for virtual reality systems.

Tracking interface can be implemented with different physical principals, including mechanical, electromagnetic, optical or acoustic methods (Meyer et al., 1992). Although no single technology works for all purposes, certain methods work quite well for specific applications (Welch and Foxlin, 2002). The new trend now is to combine different tracking technologies into one product to get better tracking quality, e.g. the hybrid suit from ART[3] is a combination of optical, mechanical and magnetic trackers. Video based tracking devices using computer vision methods, such as kinect and leap motion (Figure 1.4), are also getting more popular because of their low cost and easy setup procedure.

(a) The Kinect (version 2) © Microsoft. (b) Leap motion controller © Leap Motion.

Figure 1.4 – Video-based tracking interfaces.

1.2.3.4 Haptic Interface

The word *haptic* coming from Greek means "pertaining to the sense of touch". Haptic technology was first developed for tele-operation task so that users can enhance the remote control of machines and devices. Now more and more virtual reality applications integrate haptic interface so users can "tele-operate" objects in the virtual world. With the help of haptic feedback, users can not only perceive virtual objects' different properties (shape and texture, etc.) by touch, but also interact physically with them (e.g. to push a virtual forklift (Martin et al., 2012)). Being able to imply the sense of touch in the interaction within virtual environment is a huge step forward in aim of creating fully immersive experience in the virtual world, though lots of technical issues remain to be solved.

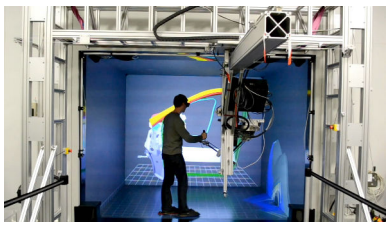
Generally speaking, haptic technology aims to recreate the sense of touch by applying two kinds of stimuli to the user: tactile feedback and proprioception feedback. Tactile sensation is formed from several modalities including pressure, skin stretch, vibration and temperature, while proprioception feedback concerns the feeling of having physical contact between different parts of human body and the virtual environment. Many haptic devices on the market provide force feedback and vibration for part of the human body, especially the hands and arms, for example, electromechanical haptic arms shown from Figure 1.5a to 1.5c have different workspace

[3] http://www.ar-tracking.com/products/motion-capture/hybrid-suit

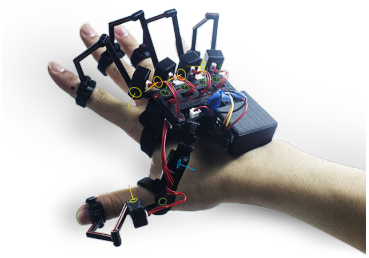
size and accessible force range to provide from desktop-based to room-sized haptic interaction. Some glove-shaped (Figure 1.5d) or string-based (Figure 1.5e) devices equipped with finger-level motors can support even finer force feedback to the user for actions like object grasping and manipulating with fingers. However, there seems still no easy solution to provide full-body haptic feedback and tactile experience with relatively complex objects, so sometimes real objects (called props or tangible interfaces) are introduced into the virtual environment for passive haptic feedback in specific applications (Figure 1.5f).

(a) PHANTOM® Omni haptic arm © Sensable.

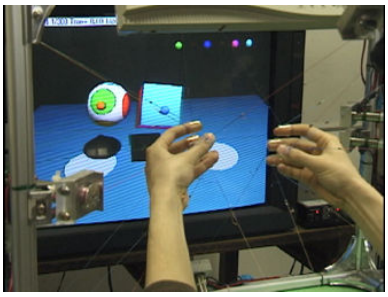
(b) The Virtuose 6D haptic arm © Haption.



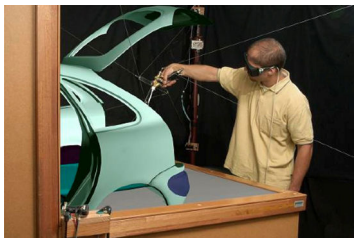
(c) Scale 1™ room-sized haptic solution © Haption.



(d) Dexmo® wearable mechanical exoskeleton © Dexta Robotics.



(e) SPIDAR: string-based force feedback device (Sato, 2002).



(f) The putty gun served as prop offering passive haptic feedback (Ortega and Coquillart, 2005).

Figure 1.5 – Different types of haptic interfaces.

1.2.3.5 Other Interfaces

Apart from major sensory channels like vision, audio and the sense of touch, the human sensorimotor system is way more complex and all senses could contribute to the feeling of presence in the virtual world. For example, Matsukura et al. (2011) designed a multi-sensorial field (MSF) display which can generate air flow and odor vapors to simulate odor distribution in the virtual world, and Narumi et al. (2011) developed a "Pseudo-gustation" method to change the perceived taste of food by changing its appearance and scent. However, integrating all types of behavioral interfaces into a single virtual environment remains a challenge because of the complexity and compatibility issues of different hardware and software solutions.

1.2.4 Human in Virtual Environments

As we discussed before, virtual reality is a human-centered research domain. Technical advancement concerning the modeling of virtual world and the design of behavioral interfaces all serve to provide a better feeling of "being in another world" (presence) for the human in virtual

environments. So to understand human activity and sensation at perceptual and cognitive levels is a key to the success of virtual reality system.

1.2.4.1 3D Interaction

The interaction between the user and 3D virtual world begins once he/she is inside the virtual environment, and these human activities can be divided into some basic behaviors named Virtual Behavioral Primitives (VBP) Fuchs et al. (2011). VBPs can be grouped into five categories:

- observation;
- moving;
- acting;
- communicating with others;
- application control.

In the list above, except for application control, these activities are very similar to what people practice in the real world. Observation is our first triggered action after we dived into a virtual environment, which is a relatively passive action. The only interaction part is that when user turns head or has ocular movements (captured with eye tracking), the system will adapt the visual and audio rendering accordingly. Moving or navigation in the virtual world is also a basic interaction for the user to accomplish various tasks like object searching or transporting, way finding, sightseeing, etc. Besides natural walking, many virtual navigation metaphors are developed to effectuate more efficient virtual viewpoint control under different specific conditions (more details to be presented in Chapter 2). Regarding "acting", which is a vague notion, is further broke into object selection, manipulation and symbolic input by Bowman et al. (2004). The way we interact with virtual object is still quite different than with real objects due to technical limitations (e.g. lack of detailed and robust haptic feedback), thus many interaction metaphors are proposed to enable other forms of interaction (Hand, 1997).

1.2.4.2 Presence

Sensorial immersion offered by immersive virtual environments (IVEs) can give user the illusion of "being there" or the feeling of presence (Heeter, 1992), which is a central concept of virtual reality. Presence, or telepresence, is used to describe user's subjective feeling of being immersed in a virtual environment, while the term "immersion" is a product of technology that facilitates the production of multimodal sensory stimuli to the user (Slater et al., 1994; Bystrom et al., 1999). Presence in virtual reality is a psychological state relying on sensorimotor illusion, which is different than the presence feeling at cognitive level that we have in dreams or by reading a book.

Many researchers share this general definition of presence, but there are still nuances in the explanations and interpretations of the definition.

Slater et al. (1994) suggest that presence is assessed by the subjects as their sense of "being there", the extent to which they experienced the virtual environments as more the presenting

reality than the real world in which the experiment was taking place, and the extent to which the subject experienced the virtual environments as places visited rather than images seen. Then Kim and Biocca (1997) describe presence as the product of two factors: (1) "the arrival" or the feeling of being "there" in the virtual environment, and (2) "the departure" or the feeling of not being there "there" in the physical environment. "The arrival" (or one's involvement in a virtual environment) occurs when an individual concentrates his/her energy and attention onto a stimulus and the events happening in the virtual environment, thus permitting the augmentation of the degree of involvement or of presence. Similarly, Witmer and Singer (1998) relate presence in part to the concept of attention: presence may vary across a range of values that depends in part on the allocation of attentional resources. They think that both involvement and immersion are necessary for experiencing presence. Lombard and Ditton (1997) have attempted to offer another explanation of the concept of presence as they define presence as the perceptual illusion of non-mediation, which focuses on the transparency of behavioral interfaces.

Studies have shown that the level of presence has not only a pronounced effect on user's task performance (d'Angelo et al., 2008), but also an impact on the social relationship between collaborators (Slater et al., 2000). To step further towards a more complete virtual reality system that could invoke higher level of presence for the user, it is important to identify factors that contribute to the formation of presence and to establish related evaluation models. Being an ongoing research topic, existing models for presence measurement are restricted to subjective rating through questionnaires (Usoh et al., 2000; Witmer and Singer, 1998). The use of physiological measures for presence evaluation has been attempted (Meehan et al., 2002), but we still need more follow-up studies to design a complete evaluation model based on physiological indicators. A detailed analysis of influencing factors of presence and a taxonomy for presence measurement methods are presented by Schuemie et al. (2001).

1.2.4.3 Cybersickness

Cybersickness is a polygenic (Kennedy and Fowlkes, 1992) and troublesome problem with current virtual reality technology. Cybersickness, or simulator sickness (they may be different according to Stanney et al. (1997)), is the tendency for some users to exhibit symptoms that parallel symptoms of classical motion sickness both during and after being immersed in virtual environments. Short-term symptoms of cybersickness have been identified after repeated studies (Lawson et al., 2002), but currently we still know little about its long-term effects. It is essential to understand the causes for cybersickness and to find ways to eliminate it for better use of virtual reality technology.

Researchers have tried to identify factors that are susceptible to cause cybersickness when using a virtual environment. For example, Rich and Braun (1996) examined the relationship between sensory compatibility and cybersickness symptoms, while So et al. (2002) emphasized that the visual complexity of the virtual scene should also be taken into consideration. Stanney et al. (2002) gives us a better global view that addresses different perceptive factors that are susceptible to cause cybersickness. Technical issues such as lags (Pausch and Crea, 1992), flickers (Harwood and Foley, 1987), and individual factors like gender (Biocca, 1992) and age (Reason and Brand, 1975), all have some influences on the severity of cybersickness symptoms. A more complete list of primary factors that contribute to the cause of cybersickness was provided by LaViola (2000). He also gave an interesting discussion on three conflicting theories that try to explain the occurrence of cybersickness.

Currently, to evaluate the severity of cybersickness, the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993) remains the reference tool. SSQ groups all symptoms around three main factors: nausea (nausea, stomach awareness, etc.), oculomotor (eyestrain, blurred vision, etc.) and disorientation (dizziness, vertigo). The development of various physiological measurements also begins to show interesting correlation between cybersickness symptoms and various physiological indicators (Kim et al., 2005; Min et al., 2004; Sugita et al., 2008). However, further research efforts are required to establish a valid cybersickness evaluation model based on physiological indicators. The study of physiological responses of human body when exposed to virtual environments will also help us to better understand the cause of cybersickness.

1.2.4.4 Workspace Management

When interacting in an immersive virtual environment, user's sensory channels are partly blocked by the computer system. However, this does not change the fact that the user is still in the real world, thus he/she is constrained by limits of the physical workspace, for example, the user can not cross real walls or screen displays. The implication of these real world constraints on the use of immersive virtual environments is a major issue of this manuscript and will be discussed in Chapter 2).

1.2.5 Summary

This first part of Chapter 1 gives a brief introduction of virtual reality around its three main components: virtual environment, behavioral interface and the user. Key notions like presence and cybersickness, as well as different hardware implementations of virtual reality system are presented, so they can be used directly later on in this manuscript.

1.3 Computer Supported Collaboration

The development of communicational technology changes the way people work together, for example, telephone and telegraph made the exchange of information much faster than paper-based media. Later on, with the boom of the Internet, emails, forums, video conferences create a tighter link among remote collaborators. However, networked collaboration still faces a lot of difficulties in terms of communication and coordination. For example, how a group of coders can contribute to the same project remains a non-trivial job despite using versioning tools like SVN[4] or Git[5].

In 1984, Irene Greif of MIT and Paul M. Cashman of Digital Equipment Corporation organized a workshop attended by people from various domains interested in using technology to support people in their work. In this workshop the term Computer Supported Cooperative Work (CSCW) was first mentioned (Grudin, 1994), then it becomes a research domain attracting world-wide interests. CSCW studies how computer systems could enable a group of people connected by network to work together efficiently, or as stated by Carstensen and Schmidt

[4] http://subversion.apache.org
[5] https://git-scm.com

(1999): "how collaborative activities and their coordination can be supported by means of computer systems". More details about the history, state-of-the-art and research issues of CSCW can be found in the book of Beaudouin-Lafon (1999).

1.3.1 Characteristics of Collaborative Work

We can learn many things from observation of people working and collaborating in the real world to get some insights into the nature of collaborative work (Churchill and Snowdon, 1998). Here are some general points to summarize characteristics of collaborative work:

Collaboration is human-centered The number and profile of people involved in the task define the general organization for collaboration. For example, small groups are likely to work together in real time with more flexible schedules, communicate informally and share information with minimal cost, while large groups need much more efforts for coordination and more adapted communication technology. The profile of collaborators also matters, if one person is way more experienced than others, it is better to have a leader-follower mode so everyone works around this leader. However, if working capacity is homogeneously distributed and especially, each one has his/her own expertise, it is better to work with equal responsibility for the task.

Collaboration is task-oriented The nature of the task defines the spatial and temporal constraints for collaboration, or we can say that sometimes people need to collaborate because of some spatial or temporal issues and the task can not be done without combining efforts from different actors. For example, tasks like painting a wall or repairing a car together require people to be co-located, other tasks like operating a TV live show, or having a video conference need synchronous actions. Moreover, complex tasks often consist of different components with inherent logical links and the task can only be accomplished with a certain procedure.

In some collaborative tasks people play similar roles, e.g., two workers try to move a heavy object (*piano movers' problem*), two pilots control the landing of a plane, etc. Whilst in other tasks collaborators can also have distinct roles with different associated competence (Pouliquen-Lardy et al., 2014), e.g. the relationship between trainer and trainee, field operator and online assistant, etc.

Awareness is the base of coordination Awareness is one's knowledge of task related activities, especially activities of others. As stated by Dourish and Bellotti (1992), awareness is an "understanding of the activities of others, which provides a context for your own activity". This "context" allows an individual to evaluate his/her actions with respect to group goals and work progress. In addition, awareness may also refer to the knowledge of the state of a task related object (e.g. the history and current state of the shared document for group editing task) and the working atmosphere (whether there is an emergency, whether people are under pressure, etc.).

A shared context is essential A shared context allows group members to "stay on the same page" (shared understanding), which largely facilitates their communication and coordination. We can call it a *common ground* (Clark and Brennan, 1991) or *team situation awareness* (Salas et al., 1995) which is based on the situation awareness of each individual. In conventional

situations, a shared context is naturally established by a shared physical and social space. For example, two workers are in the same room, there will not be confusion when one points to a table and says "Let's move it to the corner". When people are distributed and connected by computer network, it is essential for a successful collaborative system to maintain a shared context by sharing artifacts and activities.

Multiple viewpoints are helpful A complex task should have multiple representations, each offering a different point of view of the problem or focusing on a specific subtask. In certain cases, one individual may require multiple representations to reflect different aspects of their task, whilst in other cases people with different profiles and skills may require tailored representations to provide information specific to their tasks. Taking an example from Churchill and Snowdon (1998), people taking part in the architectural design review of a building might only want to see features relating to their specialty in detail. So an electrician might only want to see detailed wiring plans, but not necessarily the plans for the plumbing, except in cases where there was a potential conflict. There are more reasons to vary the representation of a task in a broader social context, for example, companies involved in a same project often need to keep their sensitive data or workflow from others. So when designing computer supported collaborative system, "WYSIWIS" (What You See Is What I See) (Stefik et al., 1987) may not be a proper choice in many cases.

Transition between shared and individual activities is required Collaborative scenarios usually contain different coupling phases (Gutwin and Greenberg, 1998; Lissermann et al., 2014): *closely-coupled phase* to solve problems in group and *loosely-coupled phase* to tackle individual subtask. So collaborative work involves the interleaving of individual and group effort, which requires explicit communication and coordination between collaborators. For example, in a car assembly task, workers may need to go to storehouses and fetch a certain component for change from time to time, and then come back to the assembly work. This task contains both individual object-searching and closely coupled interactions for the assembly work. Isenberg et al. (2012) further identified a series of eight different collaboration styles and activities that participants adopted during a working session around an interactive table. So a collaborative system should manage group workspace as well as personal workspace, and allow active transition between shared and individual activities.

These characteristics help us to understand the process of collaborative work and offer a guideline for the design of computer supported collaborative system. Okada (2007) proposed a multi-layered hierarchical framework which shows a clear picture of different elements contributing to the collaboration and a structural link between aforementioned characteristics. As shown by Figure 1.6, the model consists of four layers and each layer is based on the layer below: spatial and temporal coexistence enables awareness of others' activities, which then allows the exchange of views and opinions, the sharing of knowledge and information, and the distribution of work and operations. At the highest level, the sharing of activities enables the final collaboration between multiple actors with the balance between assertions and cooperations (negotiation).

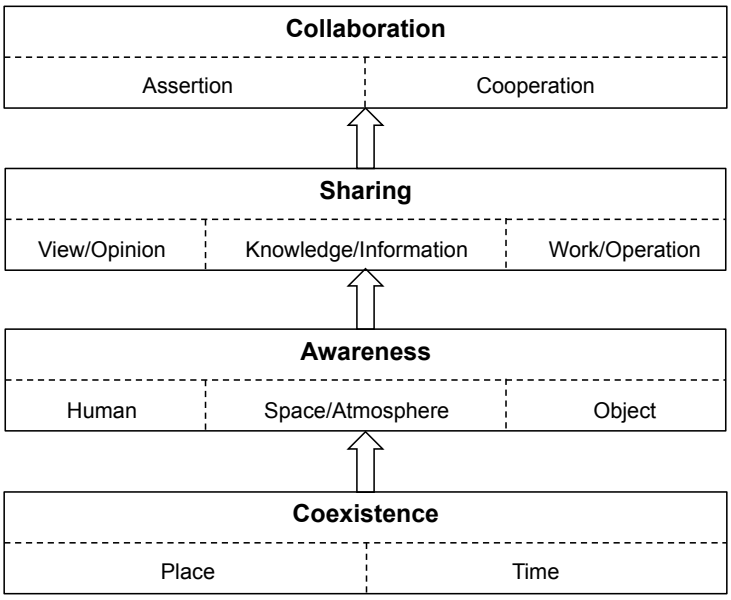


Figure 1.6 – A hierarchical collaboration model from Okada (2007).

1.3.2 Groupware

Being the central interest of CSCW, groupware is a type of computer-based systems designed to support a group of people to work together. Ellis et al. (1991) defined groupware as the following: "computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment".

Groupware implementations can be grouped by a conceptual time-space matrix (Johansen, 1988; Ellis et al., 1991) regardless of involved technology. As shown in Figure 1.7, this matrix has a temporal dimension (whether users work the same time or asynchronously) and a spatial dimension (whether users are co-located or geographically distributed).

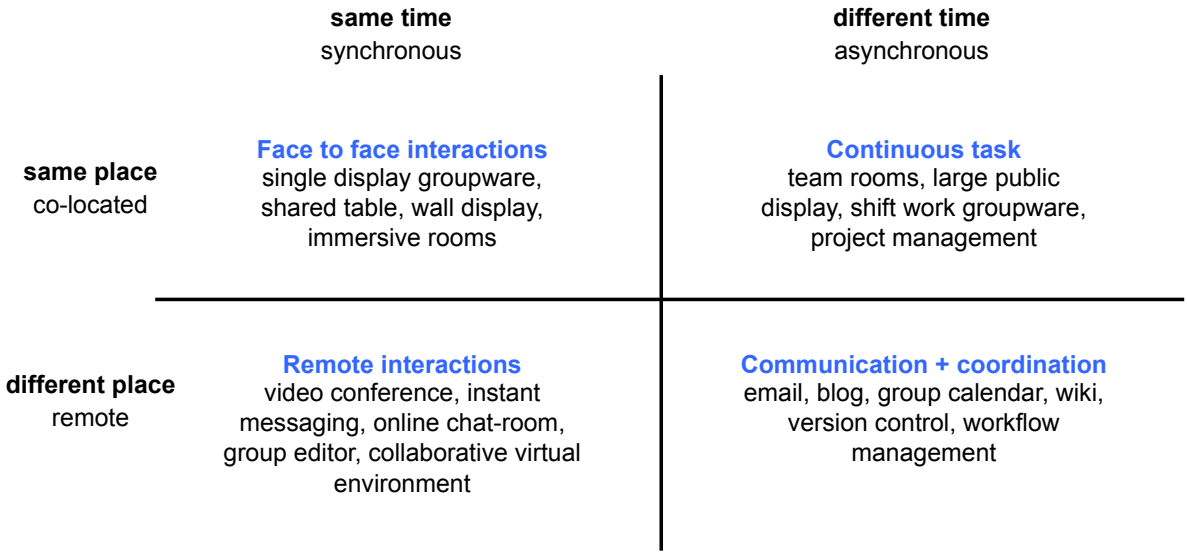


Figure 1.7 – The time-space matrix for groupware classification.

This typology of groupware based on time-space matrix is not aimed to provide strict rules to put every groupware application into one of the categories. Sometimes a given device can

support both synchronous and asynchronous communications (e.g. two groups of users work on a wall display with results left by a previous group), and the used technology does not imply geographic distance between users (e.g. forwarding an email to a co-worker in the same room). However, it shows an overview of existing CSCW systems depending on the context of a system's use. It also offers a guideline for groupware design concerning temporal and spatial constraints. For example, synchronous applications need management of concurrent resource access and live communication between users, and groupware connecting remote users need to carefully choose the network architecture (server-client, peer-to-peer, etc.) to be used and to take into consideration the network latency.

Another taxonomy by (Grudin and Poltrock, 2012) divides collaborative activities into three functional categories: communication, information sharing and coordination. The accomplishment of a collaborative task often requires all three types of activities. For example, when several authors of a book use a group editor to finish their work together, they need to have a shared version of the document and a clear assignment of different unfinished parts to corresponding authors, and also real-time communications by means of instant messages or audio meetings and asynchronous communications by leaving comments and annotations.

1.3.3 Collaborative Virtual Environment

1.3.3.1 Definition

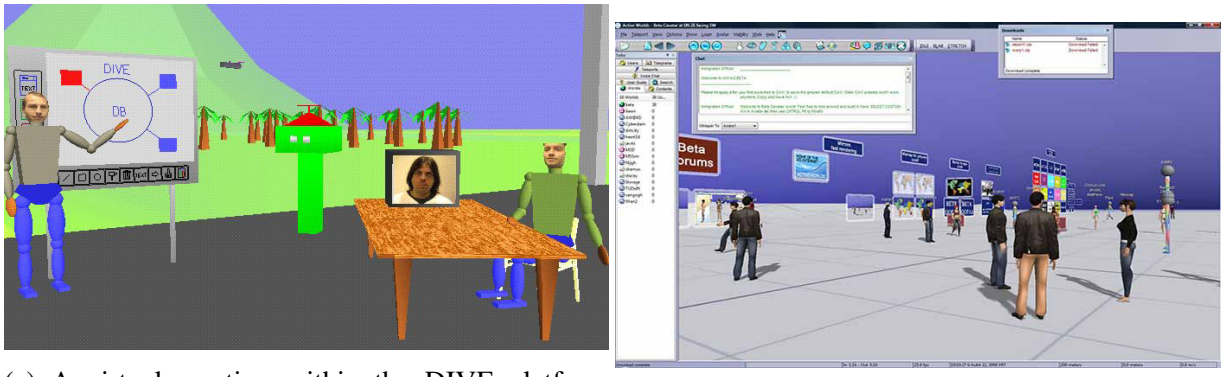
A Collaborative Virtual Environment (CVE) refers to a virtual environment that enables multiple users to interact and to achieve collaborative tasks. CVEs provide a potentially infinite, graphically realized digital landscape within which multiple users can interact with each other and with simple or complex data representations (Churchill and Snowdon, 1998).

CVE differs from traditional groupware applications (e.g. email, instant messages, video conference, etc.) in that, instead of connecting people from different locations, CVEs create a virtual world and "put" users and task related information directly in that world. This virtual world naturally provides users with a common spatial and social context for collaboration. CVEs are also different from other content sharing groupware such as group editors and group calendars because CVEs are multi-function and multi-purpose platforms that can support many forms of communication both for routinized and highly flexible tasks. Users have more degrees of control over the communication with others (free to navigate and encounter people) and more types of interactions with digital objects inside a malleable virtual space.

A lot of CVE applications exist from early prototypes (e.g. DIVE (Carlsson and Hagsand, 1993), NPSNET (Macedonia et al., 1994) etc.) appeared during 1990's (Figure 1.8a), to more developed systems like MASSIVE-3 (Greenhalgh et al., 2000) and ANTS (López and Skarmeta, 2003), etc. The general framework of CVE is gradually stabilized and standardized since then. Now mature CVE systems are widely applied and become the major trend in the game industry, there are enormous commercial products like online communities (e.g. Active Worlds[6]) (Figure 1.8b) and collaborative games like MMORPGs[7] which attract millions of users (Brown and Bell, 2004). More information about the history and current status of CVEs are presented by Joslin et al. (2004).

[6] https://www.activeworlds.com

[7] Massively Multiplayer Online Role-Playing Games



(a) A virtual meeting within the DIVE platform (Carlsson and Hagsand, 1993).

(b) The Cavatar world from Active Worlds Inc.

Figure 1.8 – Some examples of existing CVE applications.

1.3.3.2 Issues and Challenges

The emergence of CVEs comes along with technical issues due to the complexity of such system. Supporting rich social interaction in densely populated virtual worlds requires addressing a variety of technical challenges. Here is a list of technical issues that need to be addressed inspired by the work of Benford et al. (2001) and Joslin et al. (2004):

- Scene management: how to merge events and keep data consistency among different sites, how to segment a virtual world into smaller sections (Kazman, 1993);
- Network topology: choose among server-client, peer-to-peer or hybrid structure to reduce traffic loads and improve data distribution efficiency;
- Compression: how to compress the data to be transmitted knowing that real time interaction between multiple users can generate huge amount of data, especially with sophisticated virtual human (Capin et al., 1998);
- Personal viewpoint: design software model to support "subjective" view on the shared world (Smith, 1996);
- Human computer interface: using virtual reality technology to convey more information for natural interaction and social communication;
- etc.

Human factors are also crucial to the design and effective use of CVEs as a new type of social community. Here we group research questions around four topics:

- Social conventions: are the social conventions in the real world still applicable in the shared virtual world (Becker and Mark, 1998)? What has or has not changed?
- User embodiment: how the virtual representations (simple object, robot, humanoid avatar, etc.) influence their communication? How to design the body image of a user to provide information such as identity, activity, availability, mood and many other factors (Benford et al., 1995)?

- Awareness: how to be aware of other people's intention and perceptual capabilities (e.g. size of field of view) in the virtual world (Benford et al., 1994)?
- Social communication: how social communications are supported in CVEs compared to cases in real world (Bailenson and Yee, 2006)?

Otto et al. (2006) also give a similar summary of factors that influence collaboration in CVEs by grouping them into different levels:

- Application factors: task design, usability, performance, workflow ,etc.
- Human factors: presence, social human communication, etc.
- Technology factors: immersion, field of view (FoV), user interface, data distribution, etc.

1.3.4 Summary

This second part of Chapter 1 summarizes characteristics of collaborative work in the real world and their implications for the design of computer supported collaboration systems (groupware). CVE as a flexible and multi-function platform attracts research interests from both CSCW and virtual reality communities. Although there are still many technical and human-related issues to be solved, CVEs already show great potential for supporting efficient and seamless collaboration in a rich social context.

1.4 Immersive Collaborative Virtual Environment

As stated in the previous section, CVE is developing as a convergence of research interests within CSCW and VR communities. On the one hand, most existing virtual reality systems (e.g. CAVEs or HMD-based systems) can offer high level of multi-sensory immersion for a single user (except multi-user systems that we will talk about in Chapter 2), so it is interesting to make connexions between these (remote) immersive systems by CVE technology to enable group immersion. On the other hand, with the progress of telecommunication technology and computer science (computer graphics, high performance computing, etc.) in recent years, many low-level issues of CVEs are already solved or we can say, are no longer the bottleneck for CVE usage. However, desktop-based CVEs with traditional human-computer interface have still limited support for social human communications compared to face-to-face interaction in the real world, and the lack of sensorial feedback makes it difficult for users to feel "being in the virtual world" and impairs their task performance (Narayan et al., 2005; Nam et al., 2008), so it is important to introduce behavioral interfaces for CVEs to convey more social cues for human communication and to improve the immersion level.

As a consequence, more immersive CVEs are implemented with various hardware and software solutions. With the "perception, decision and action" loop as shown in Figure 1.2, we can apply this loop in a multi-user situation and extend it to illustrate the conceptual framework of immersive CVEs (Figure 1.9). In this shared virtual world, users' "perception, decision and action" loops are interconnected as one user's action could affect other users' perception of the virtual world.

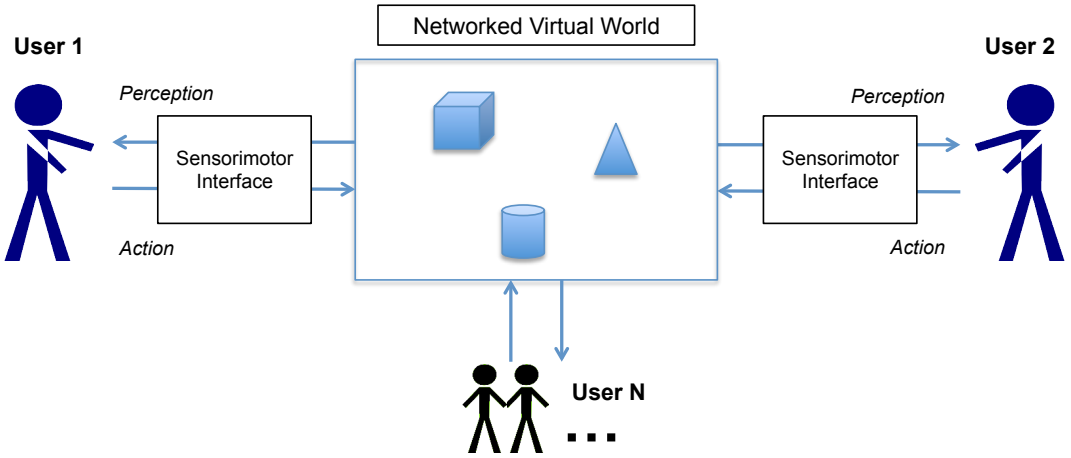


Figure 1.9 – The conceptual framework of immersive CVE designed for the multi-user collaboration.

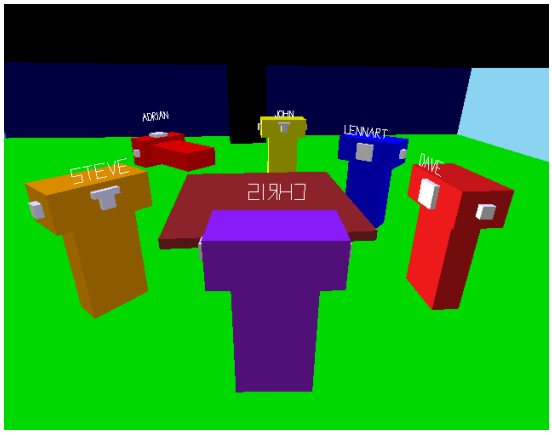
This new networked immersive experience not only changes the way people interact with objects in the virtual world, but also has impacts on how users communicate and interact with each other. The following part of this section will present in more details about different research questions and existing work around CVEs and immersive CVEs from a user-centered perspective.

1.4.1 User Representation

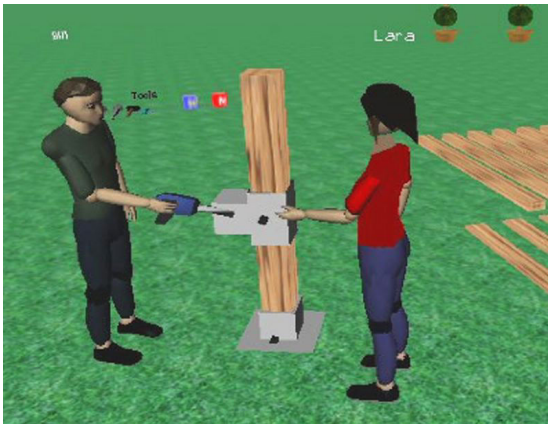
As discussed in section 1.3.3.2, users need virtual representations in the shared virtual space in order to be perceived by others. The virtual representation of users, or can be called user embodiment as explained by (Benford et al., 1995), "concerns the provision of users with appropriate body images so as to represent them to others (and also to themselves) in collaborative situations." Information conveyed by such representation could be position, identity, activity, availability and many other factors (Thalmann, 2001). In an immersive virtual environment, this embodiment plays a crucial role for user interaction and communication (Slater and Usoh, 1994), it is also helpful (sometimes indispensable) for users to get self-related information. For example, a user equipping an HMD perceives the virtual world from a first-person viewpoint, but the perception of his/her real body is completely blocked so a virtual body image can help to "recreate the missing body" (Lok et al., 2003; Mohler et al., 2010). Moreover, according to the task that the users need to achieve in the virtual context, social representation of self may be provided through dedicated virtual clothings and/or virtual tools.

The virtual representation of a user is usually a graphic entity, although sometimes a spatialized audio source is also sufficient for certain tasks. Three dimensional graphic recreations of human body (i.e. avatar) can be in extremely different forms, both in terms of morphology and photorealism (rendering style and Levels of Detail (LoD)) (Garau, 2006). We can have from simple humanoid robots to highly detailed realistic virtual humans. For example, as shown by Figure 1.10a, users in MASSIVE-1 system are embodied in T-shaped robots with their names on top, they can easily get each other's information as position, orientation, identity and availability (avatars of off-line users are lying down) (Greenhalgh and Benford, 1995). Avatars as used by Roberts et al. (2004) can convey more non-verbal information such as gestures and postures (Figure 1.10b). Users can also be represented by live video streams shown on 3D win-

dows (billboards) (Hayashi et al., 2007) (Figure 1.10c). Another interesting method developed by Ogi et al. (2001) introduces an animated video avatar based on live video capture to avoid mesh-based avatar design. This 2.5 dimensional video avatar is captured by a set of cameras so the avatar can be seen from a certain range of directions (Figure 1.10d).

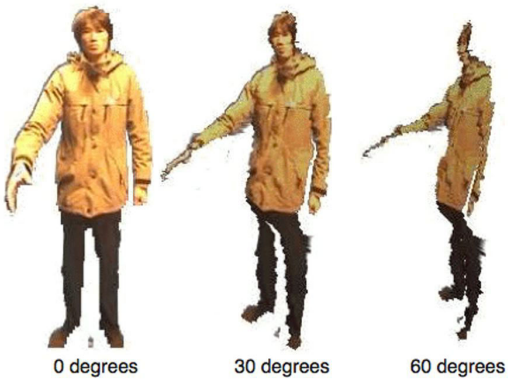


(a) Users represented by T-shaped robots in MASSIVE-1 (Greenhalgh and Benford, 1995).



(b) Simple avatars with higher level of detail (Roberts et al., 2004).

(c) Users having a round table meeting through a set of video streams shown on 3D billboard (Hayashi et al., 2007).



(d) Live video texture based 2.5 dimensional video avatar seen from various directions (Ogi et al., 2001).

Figure 1.10 – Some examples of user's virtual representation in CVEs.

In immersive CVEs (as well as desktop CVEs), high fidelity virtual humans are not always preferred against simple avatars. First, as human body is extremely complex which contains hundreds of muscles and joints, a virtual model replicating exactly the human body in real time is both challenging and costly in terms of computational and network resources. So generally we only need avatars that are "good enough" for the given task depending on the trade-off between fidelity and efficiency. Second, the resemblance between the real user and his/her avatar is not always found to be beneficial due to the uncanny valley (Mori, 1970; Mori et al., 2012), which has led to lots of discussions and research works in the fields of robotics and computer animation. At last, as collaborative works are task-oriented, certain tasks proceeded in non-realistic virtual worlds (e.g. scientific data visualization, imaginary artistic world) do not necessarily require virtual humans. User representations are not limited to humanoid avatars and can be any kind of "beings" or objects according to the context of the virtual world, sometimes realistic virtual humans can even be "distractors" for the ongoing task.

Overall, a lot of technical issues still present regarding avatar modeling, animation and

motion capture technology if we look for realistic virtual humans inside immersive CVEs. Magnenat-Thalmann and Egges (2006) give a good review of research work on interactive virtual humans in real-time virtual environments.

1.4.2 User Communication

Social human communication (SHC) contains four primary elements: verbal and non-verbal communication, references to objects and references to the environment (Burgoon et al., 1994). Bolt (1980)'s "put-that-there" command gives a good illustration of these four elements despite the fact that he was talking to a computer: he was speaking (verbal communication) while pointing (non-verbal communication) to an object (reference to objects) and "there" refers to a certain place in the virtual environment.

When working with CVEs, references to objects and environment can be easily understood as all users share the same objects and environment. Regarding verbal and non-verbal communications, while it is relatively easy to enable direct verbal communication (spatialized audio requires more complex setup) between users, it is complicated to convey non-verbal communications through network.

Recent research shows increasing importance of non-verbal communications during networked collaborative work (Guye-Vuillème et al., 1999; Roberts et al., 2004) including gestures (Dodds et al., 2011), postures (Normand et al., 2012), eye contact (Bailenson et al., 2002; Garau et al., 2003) and facial expressions (Boker et al., 2009). With desktop-based CVEs, we can play pre-recorded animations to carry out non-verbal communication as implemented in many multi-player games. However, this method is cumbersome and all personal information can not be conveyed. Immersive CVEs allow users connected by network to "step into each other's world" and provide the closest resemblance of co-location compared to other tele-collaboration technologies (Wolff et al., 2007). In immersive CVEs, non-verbal cues such as gestures and postures, can be better supported by motion tracking technology combined with real time avatar animations.

Social scientists also start to use immersive CVEs as a tool to study human behavior in face-to-face communication by introducing biases in the virtual simulation. For example, Ennis et al. (2010) investigated human sensitivity to audio mismatches and visual desynchronization using motion capture data (Figure 1.11).

1.4.3 User Interaction

During closely-coupled collaboration, user interaction mainly concerns object co-manipulation among multiple users. Co-manipulation means that multiple users can act on the same object simultaneously, which belongs to level 3 cooperation according to the classification given by Margery et al. (1999). To achieve co-manipulation, the system needs to manage concurrent access to an object by combining inputs from multiple users which often involve multimodel instructions (Martin et al., 2011). As we presented in section 1.2.4.1, the way that users perform actions in the virtual world can be similar or very different from how we interact in the real world, metaphors are often needed to enable efficient interactions.

Two general solutions exist to support object co-manipulation: we can allocate the control

Figure 1.11 – Avatars animated by motion capture data for behavioral study (Ennis et al., 2010).

of different attributes to different users, i.e. to separate degrees of freedom (Pinho et al., 2002) (e.g. one moves the object while another person changes its color); if we need to modify the same attribute, for certain type of attributes (e.g. position, size, etc.) we can take the average value of all the inputs (Ruddle et al., 2002), or we can use sophisticated metaphors for real time concurrent manipulation, such as the SkeweR (Duval et al., 2006), 3-Hand Manipulation Technique (Aguerreche et al., 2009), etc.

Various research work prove that visual immersion (Schroeder et al., 2001; Roberts et al., 2003; Narayan et al., 2005) and multi-sensory feedback (especially haptic feedback) (Nam et al., 2008; Oguz et al., 2010) tend to provide users with higher level of presence and improve their task performance on object co-manipulation based on comparisons between desktop CVEs and immersive (or partially immersive) CVEs. Slater et al. (2000) also find that immersion has an impact on the social relationship between group members: the person who uses immersive display tended to emerge as the leader among users connected to the same CVE by desktop interfaces.

1.4.4 From Presence to Copresence

Users in an immersive virtual environment can have the feeling of presence ("being there") while a group of users connected to the same immersive CVE can experience copresence ("being there together") (Slater et al., 2000). Copresence differs from *social presence* as the latter is a much broader concept which refers to the individual's experience of being with another person (not limited to CVEs) (Schroeder, 2002).

Copresence itself can also have different implications and can be interpreted as mode of being with others, or as sense of being with others (Zhao, 2003). It is difficult to separate copresence in the sense of co-immersiveness from copresence in the sense of doing things together (Schroeder, 2002) as the task duration and amount of interactions during collaboration seem to influence the feeling of being together with other people. Existing studies (Slater and Steed, 2002; Garau, 2003) find a link between copresence and avatar fidelity in terms of appearance and behavior: even minimal behavioral cues can enhance the perceived quality of social interaction in VEs, however, the consistency between the fidelity of the avatar's behavior

and its appearance is essential. Bailenson et al. (2005)'s experiment showed that copresence was lowest when there was a large mismatch between the appearance and behavioral realism of an embodied agent, though the "uncanny valley" was not discussed due to the limited test conditions.

Several other concepts, such as mutual awareness, connected presence, engagement and virtual togetherness, have similar meaning with copresence. Detailed descriptions and discussions on these concepts can be found in Schroeder (2006)'s review which facilitates understanding of group behavior in networked virtual environments.

1.4.5 Summary

This last section presents different aspects of collaboration in CVEs from user-centered perspective: how users are presented and perceived in virtual environments, how they communicate and interact to achieve a common goal. Being the combination of CVEs with virtual reality technology, immersive CVEs provide users with novel experience in terms of perception and interaction which makes a further step towards seamless collaboration among networked users.

1.5 Conclusion

This chapter contextualizes the research by presenting an overview of relevant notions for collaboration in virtual environments, especially in immersive systems where users interact through behavioral interfaces. With technical progress of CVEs and VR interfaces, more immersive CVEs support user groups for collaborative work. Novel interaction paradigms are needed to ensure users' mutual awareness and to facilitate user communication and interaction, also to keep their level of immersion in the virtual environments.

When multiple users are physically situated in the same immersive system for co-located collaboration, the coexistence of real and virtual information create a more complicated mixed context. The next chapter will concentrate on co-located collaboration in immersive virtual environments with the driving motivation, addressed issues, and the approach taken in this thesis.

2

Co-located Immersive Collaboration

Chapter Abstract

This chapter concentrates on how co-located users collaborate inside multi-user immersive virtual environment and it is divided into three main parts. The first presents technical aspects related to multi-user immersive displays and summarizes existing methods to separate images for different users. The second describes basic notions about users' spatial organization and how users navigate in the virtual environment. Then the last part presents two major issues that we identified during the collaboration process and proposes research methodology that we adopted to solve these issues.

Contents

2.1 Introduction

When several users work in a multi-user immersive virtual environment for collaborative tasks, they share a virtual world on top of the same physical workspace. This physical collocation forms a mixed context which lies in the middle of Milgram et al. (1995)'s reality-virtuality continuum where user can have direct as well as computer-mediated interaction and communication. While lots of research works focus on supporting remote users to work efficiently together via immersive CVEs, co-located collaboration in the same immersive virtual environment for now receives limited attention. This is mainly due to the rareness and relatively high-cost of multi-user immersive systems, in contrast to non-immersive multi-user systems such as interactive walls and tables that are already widely studied in the CSCW community (Scott et al., 2003; Inkpen et al., 2005).

Here we focus on co-located collaboration in immersive multi-user systems. Given the characteristics of collaborative work presented in chapter 1, we are mainly interested in following research questions:

- How users perceive each other and achieve a shared context for collaboration?
- How to intelligently manage users' viewpoints to support various types of collaborative tasks that require different spatial configuration of users in the virtual world (side-by-side or far away from each other)?
- How to manage each user's workspace and spatial relationship with other users?
- How to allow fluent transition between shared and individual activities for each user?

Below we begin by presenting multi-user immersive displays designed to offer individual stereoscopic views for multiple users. Then we talk about different concepts related to spatial arrangement of users in the common physical workspace and navigation in the virtual world. At last we talk about issues that we identified during collaborative tasks and research methods that are applied to study these issues.

2.2 Multi-user Immersive Display

Immersive display constitutes a main component of immersive virtual environment as the visual sense plays a predominant role in user's perception process. As mentioned in previous chapter (section 1.2.3.1), immersive displays differ from traditional displays in that they have large field of view, high resolution, stereoscopic images associated with viewpoint tracking technology.

2.2.1 Stereoscopy

Stereoscopy is the production of the illusion of depth by presenting a pair of 2D images showing two perspectives that both eyes naturally have in binocular vision. Then the brain merges the two images into a "single" one to achieve stereopsis (Blake and Sekuler, 2006). Besides stereoscopy, other cues (e.g. object occlusion, linear perspective, etc.) can also help to determine

relative distances and depth in a perceived scene, but stereoscopy remains the most effective factor that provides users with instant depth perception.

To achieve stereoscopic vision, we should first generate a pair of images depending on user's head position and orientation, as well as the interpupillary distance (IPD) (Dodgson, 2004). Then we need to provide two images separately for each eye. Existing separation methods can be classified into three categories:

- Head-mounted displays: we can put two small screens in front of the eyes, or a single screen displaying two images side by side to show only the desired image to each eye (Figure 1.3e);
- Auto-stereoscopic screen: auto-stereoscopic display separate images at the screen level by using the lenticular lenses or parallax barrier to assure that each eye of the user sees different pixel columns which correspond to two different images (Perlin et al., 2000);
- Eyeglasses separation: images can be separated passively by colorimetric differentiation or polarized glasses, or actively by rapidly alternating shuttered glasses.

Each type of stereoscopic display has its advantages and inconveniences. HMD is the most straightforward way to provide two different images for the eyes and allows fully visual immersion of the user (the real world is completely blocked from user's eyes except for see-through HMDs (Schmalstieg et al., 2002)). However, various technical limitations like limited resolution and field of view enhance user's visual fatigue and restrain the wide use of HMDs. Unlike HMDs that are heavy to carry, eyeglasses based technology combined with large immersive projection display (e.g. image wall, CAVE, dome) gives user a more comfortable stereo experience. Auto-stereoscopic screens do not require users to equip any glasses or headgear, but have limited work range - users should stay at predefined positions (at a certain distance from the screen) to correctly perceive stereo images.

Recent technical developments largely improved the usability of all three types of stereoscopic display. HMDs now are lighter and less expensive with more compact design, along with better resolution, larger field of view. The improvements of immersive projection technology (IPT) (Bullinger et al., 1997) makes projection-based systems with eyeglasses a widespread solution both in academic institutes and industries, and they are specially useful to create large immersive virtual environment for group immersion by the combination of multiple projectors, such as CAVEs. Auto-stereoscopic screen combined with user tracking can now allow the tracked user to pass through different viewing areas without discontinuity (Kooima et al., 2010), but still requires complex hardware and software setup. So HMDs and projection-based systems become the major platform to support immersive or partially-immersive experience.

2.2.2 Visual Distortion

Stereoscopic images are generated from a single location - the center of projection (CoP) (Banks et al., 2009). A user can properly perceive the displayed 3D content on a projection screen when viewed from the CoP. So in an immersive virtual environment, user's viewpoint is captured by head-tracking to serve as the CoP to generate appropriate images in real time as the user moves physically.

While HMDs provide personal stereoscopic images for a single user, projection-based systems are often shared by a group of people for team work, as shown by Figure 2.1. In such multi-user situation, only the tracked user receives correct stereo images from the CoP, other co-located users (followers) share the same stereo images intended for the tracked user and participate passively in the collaborative task (Bayon et al., 2006). When users are close enough to the tracked user, they can get a relatively faithful representation of the virtual environment, but as the distance increases, displacement from the CoP results in increasingly inappropriate stereo cues and distorted perception of the virtual space.

Figure 2.1 – A group of people collaborating in a CAVE for data visualization (Pollock et al., 2012).

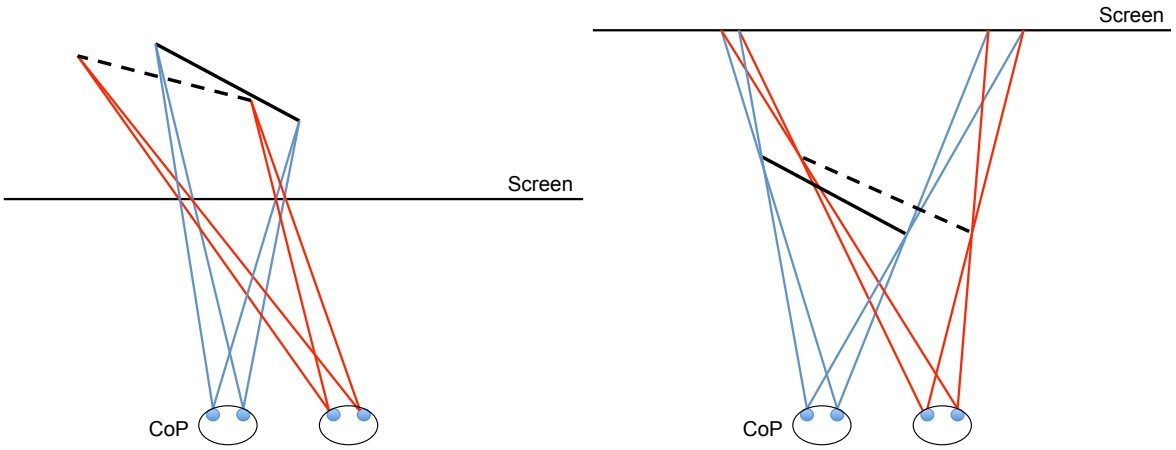


Figure 2.2 – Illustration of visual distortion by perceiving the right view from a location other than the CoP. The black line represents the correct view of object perceived from CoP and the dotted line corresponds the distorted view observed from another location.

Visual distortion caused by displacement from the CoP can be predicted by a ray-intersection model (Burton et al., 2012). As shown by Figure 2.2, no matter the virtual object is situated behind or in front of the screen, a lateral displacement from the CoP will cause deformation and shift of the perceived object. Similar effects can be observed by moving forward or backward with respect to the CoP.

A series of studies have been conducted to assess the influences of such visual distortion. When viewing monocular displays from locations displaced from the CoP, the spatial judgments remain relatively acceptable. Vishwanath et al. (2005) investigated the mechanism underlying this perceptual invariance by studying the perceived shapes of pictured objects viewed from various locations and they find that invariance is achieved through the awareness of the 2D picture plane. However, when it comes to stereo images, Banks et al. (2009)'s experiment indicates that human viewers of stereo pictures are unable to compensate for incorrect viewing position. This result is confirmed by follow-up studies: judgments of angles are distorted after leftward and rightward displacement from the CoP (Burton et al., 2012) and judgments of object depth are distorted after forward and backward displacement from the CoP (Pollock et al., 2012), although the magnitude of these distortions is consistently less than predicted by the ray-intersection models. More studies on visual distortions in stereoscopic systems can be found in articles by Woods et al. (1993); Held and Banks (2008); Ponto et al. (2013).

2.2.3 View Separation

To better support co-located use of immersive virtual environments, we need to provide perspective-correct (distortion free) individual stereoscopic views for each user. One direct way is to use personal displays such as HMDs (Salzmann and Froehlich, 2008) or see-through HMDs (Schmalstieg et al., 2002), and another solution is to introduce adaptations to existing immersive displays which already support group immersion.

Bolas et al. (2004) categorize solutions for displaying multiple images in a common area:

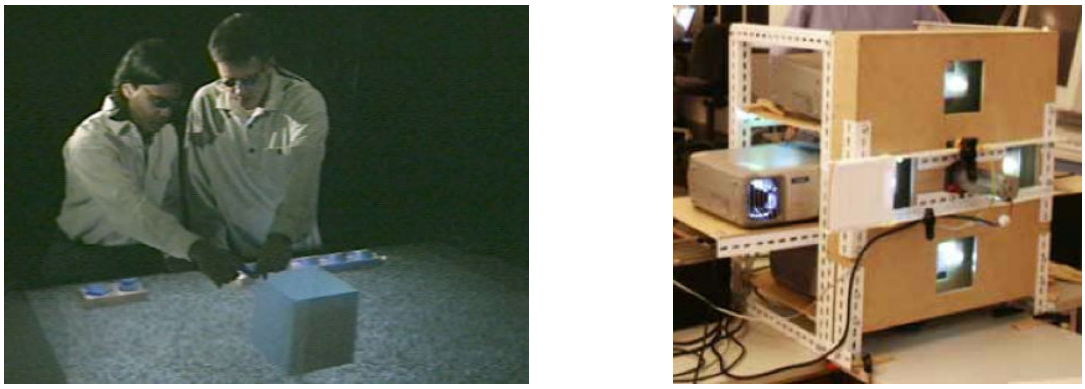
- Spatial barriers use the display's physical configuration and user placement to block users from seeing each other's view.
- Optical filtering involves systems that filter viewpoints using light's electromagnetic properties, such as polarization or wavelength.
- Optical routing uses the angle-sensitive optical characteristics of certain materials to direct or occlude images based on the user's position.
- Time multiplexing solutions use time-sequenced light and shutters to determine which user sees an image at a given point in time.

Except the first solution, the three other options are the same technologies that are used to separate images for the left and right eye of a single user as presented in section 2.2.1. Multi-user systems are often build with mixed solutions from these categories, here is a description of existing systems that provide independent stereo images for different users inside immersive or semi-immersive environment.

2.2.3.1 Image Separation

Typically we can create separate image channels for different users by adding shutters and/or optical filters in front of projectors combined with synchronized counterparts in front of user's eyes.

The two-user Responsive Workbench developed by Agrawala et al. (1997) relies purely on a time-multiplexing method which displays four different images in sequence on a CRT projector at 144Hz, thus each eye of a user views the virtual scene at 36Hz (Figure 2.3a). This workbench is the first demonstration of a two-user stereoscopic system, but the time-multiplexing method largely reduces projection time for each eye (low brightness) and users suffer from image flicker and crosstalk. Blom et al. (2002) then extended this active shuttering method to multi-screen systems like CAVEs. (Froehlich et al., 2004) further studied active shuttering technology by testing two kinds of shutters on the projector side with a range of shuttering frequencies (Figure 2.3b). The mechanical shuttering delivers higher brightness and less cross talk, but does not extend as easily to more than two users as liquid crystal (LC) shutters because of the required rotation speed and size of the disc. They tested LC shutters from 140Hz to 400Hz and found that users did not perceive flicker above a refresh rate of 200Hz, but a frequency higher than 320Hz would result in very dark images.



(a) The two-user Responsive Workbench (Agrawala et al., 1997).

(b) Two-user active separation with four projectors (Froehlich et al., 2004).

Figure 2.3 – User separation by time-multiplexing with active shutters.

(a) Wall display for two users (Fröhlich et al., 2005).

(b) The six-user projection display (Kulik et al., 2011).

Figure 2.4 – User separation by active time-multiplexing combined with polarization.

Another solution is to combine time-multiplexing with polarization filters. In 1999, Barco developed the "Virtual Surgery Table"[1] which provides two users with stereoscopic images by

[1] http://www.barco.com/en/Products/Compact-Multi-User-Projection-Table.aspx/

differently polarizing the output of two active stereo projectors. Then Fröhlich et al. (2005) extended this shuttered display to support up to four users with eight shuttered liquid-crystal display (LCD) projectors (Figure 2.4a). In this setup, images for different users are separated by active shutter glasses while the separation of the images for the left and right eye is ensured by passive polarized filters. This approach shows better performance in terms of perceived flicker, brightness of each view and crosstalk compared to purely active shuttering method.

Fröhlich et al. (2005) also summarized three main parameters can be considered to evaluate the quality of a multi-stereoscopic projection system:

- Brightness per view;
- Static and dynamic crosstalk;
- Perceived flicker, which depends on the shutter frequency, the video rate of the projector and brightness.

In 2011, Kulik et al. (2011) developed a projection-based stereoscopic display for six users by using six customized digital light processing (DLP) projectors running at 360Hz for a single screen, which results in 60Hz per user (Figure 2.4b).

Dodgson (2005) provided an introduction and overview of auto-stereoscopic multi-view displays. Users in this kind of multi-view system get perceive 3D objects from his/her own point of view without tracking or eyeglasses, but only inside a limited zone depending on different properties of the display. Another issue is that with increasing number of views (e.g. up to 256 views by Takaki and Nago (2010)), generating images in real time for dynamic interaction would be a challenge.

2.2.3.2 Spatial Separation

Instead of creating image channels, we can also take advantage of the spatial property of the display by assigning different screens or parts of a single screen to different users. For example, the "Protein Interactive Theater" (PIT) (Arthur et al., 1998) uses two orthogonal screens and each user looks at only one of the screens (Figure 2.5a). The IllusionHole (Kitamura et al., 2001) uses a circular mask on top of a tabletop projection. By looking through the mask, users positioned around the table see their individual stereo images shown in different areas of the screen (Figure 2.5b). Other systems like the Virtual Showcase (Bimber et al., 2006) and Joint Space Station (Mulder and Boscker, 2004) use similar mirror-based display to support multiple users. These desktop-based systems are often designed to accomplish specific collaborative tasks (e.g. 3D object visualization) and provide limited workspace with inherent spatial constraints.

When using larger wall display or CAVE, users can have head-tracked individual monoscopic (Maksakov et al., 2010) or stereoscopic views (Schulze et al., 2012) on different parts of the display. When they share the same part of the screen, sophisticated algorithms are applied to recalculates images based on an averaged viewpoint depending on positions and orientations of all tracked users. These software-based solutions allow co-located collaboration in immersive virtual environment without additional hardware setup, although the reduced visual distortions could still be disturbing for certain tasks that require precise spatial operations (e.g. object co-manipulation).

(a) The PIT system for two-user collaboration (Arthur et al., 1998).

(b) The IllusionHole for tabletop collaboration (Kitamura et al., 2001).

Figure 2.5 – User separation by using different screens or different parts of the same screen.

2.2.3.3 Other Methods

There are also many other display solutions offering individual stereoscopic views like the volumetric display (Grossman and Balakrishnan, 2008) and holographic display (Lucente, 1997). However, for now it is still difficult to extend these displays to provide large-scale visual immersion.

Another interesting multi-user display is the omni-stereo display (Simon et al., 2004) such as AVIE (McGinity et al., 2007) and i-Cone (Simon and Göbel, 2002) which provides good support for immersive visualization with large user group, but theoretically users need to stay at predefined positions to get good perspective. Simon (2007) compared usability and interaction performance between multi-viewpoint images and head-tracked stereo display. Results showed that for certain tasks that users do not need to move physically (e.g. ray-casting selection and in-hand object manipulation), multi-viewpoint images can produce similar or even better performance than fully head-tracked interaction.

2.2.4 Summary

This section gives a general presentation of multi-user immersive display from the initial motivation (to provide distortion-free stereoscopic view for each user) to various implementations.

Among all presented multi-user display technologies, the active & passive method which combines time-multiplexing with polarization filters is the most effective and cost-efficient way to build multi-user immersive virtual environment. There are multiple reasons: first, it completely eliminated visual distortion due to observation from another position than the CoP; second, it can be easily applied to different shapes of large-scale immersive systems like walls, CAVEs or domes without imposing strong spatial constraints on user's position and orientation with respect to the screen(s), users can move inside a relatively large physical workspace; at last, compared to purely time-multiplexing separation, it provides high brightness, low crosstalk and less flicker stereo images.

2.3 Spatial Organization

2.3.1 Stage

In most VR systems, there is a spatial reference frame that maps user's physical position in the tracking space to a corresponding location in the virtual world, so that the computer can use this position to render proper images in correspondence with user's point of view. In practice, the spatial reference frame is often expressed as the virtual coordinates of the center of physical workspace defined by the screen and/or tracking device configuration. Here we borrow some definitions from the Immersive Interactive Virtual Cabin (IIVC) model (Figure 2.6) that can help to describe user's workspace:

- The *stage* is a virtual representation of the physical workspace.
- The *conveyor* is the integration frame of the stage into the virtual world. This conveyor has its own position, orientation and scale in the virtual world coordinate system, and the stage is linked to the conveyor with position, orientation, and scale offsets.

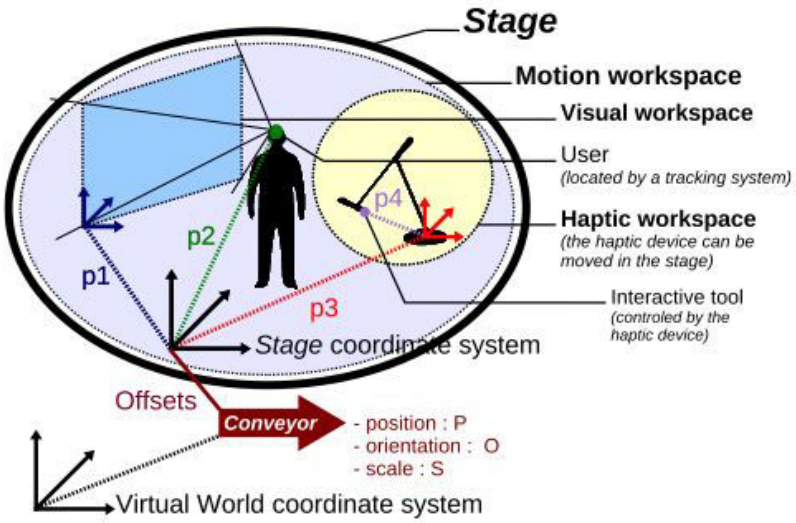


Figure 2.6 – The IIVC model: the conveyor carries the stage with its workspaces in the virtual world (Fleury et al., 2010).

So with the stage model, the stage center becomes the spatial reference frame to correctly locate not only the user, but also other devices (e.g. props) in the virtual world. For example in Figure 2.7, from left to right, user remains at the same position in the physical workspace while the stage center is set to different location in the virtual world, the user will have different virtual viewpoints accordingly.

2.3.2 Spatial Consistency

In a multi-user virtual environment, each user has a corresponding stage in the virtual world. The multi-user display system is able to render completely different stereoscopic images for each user depending on the configuration of users' stage centers.

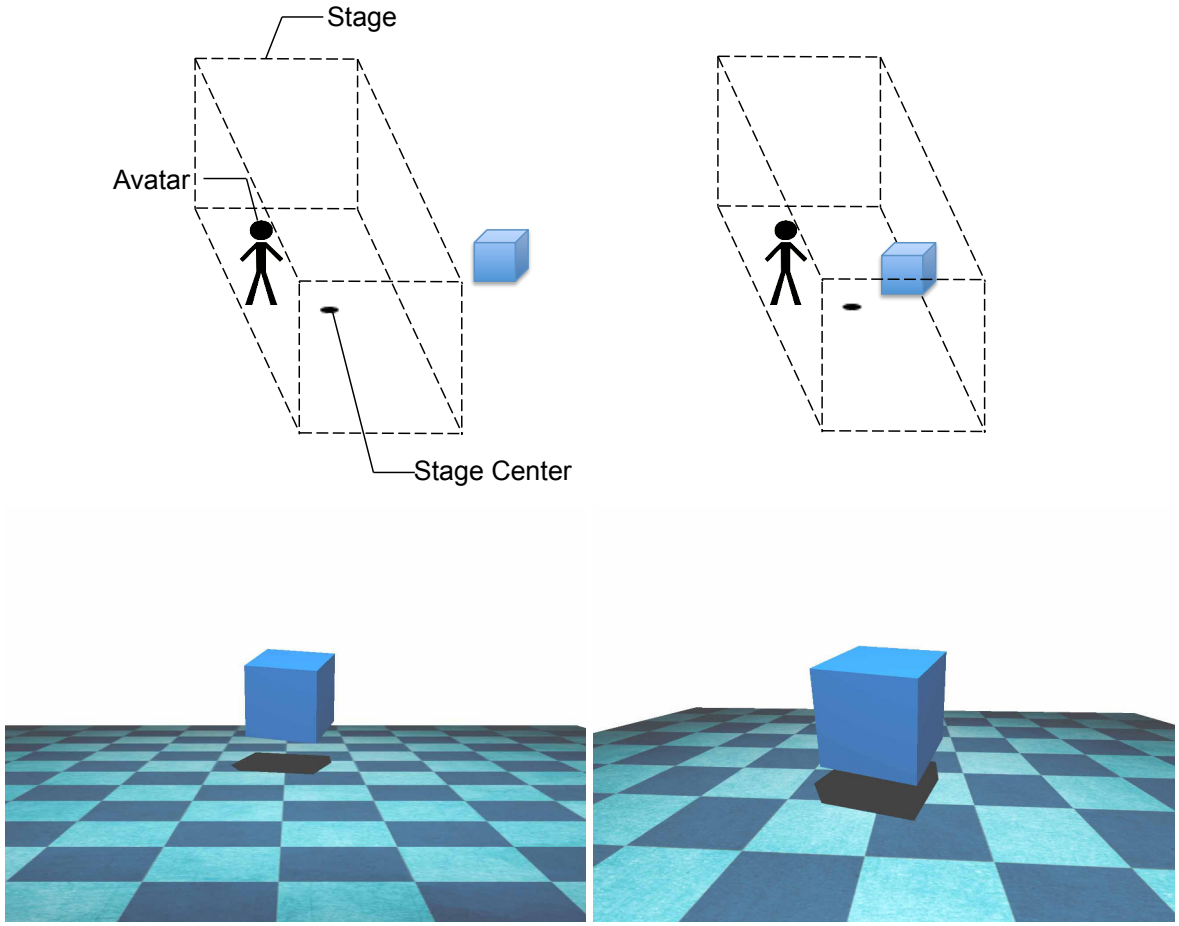


Figure 2.7 – Illustration of the impact of moving spatial reference frame on user's viewpoint location in the virtual scene.

When all stage centers are strictly superimposed, the spatial relationship between users in the virtual world will be consistent with their relative spatial distribution in the real workspace. In an immersive multi-user virtual environment, this spatial consistency allows users to perceive a virtual object at exactly the same physical location from corresponding viewpoints (Figure 2.8 top). This particular situation is quite similar to cases when users collaborate around physical objects in the real world.

Otherwise, each user will no longer perceive the same object at the same physical location. What a user perceives totally depends on the configuration of spatial reference frame in the virtual world. For example in bottom part of Figure 2.8, user B's stage center is shifted to the left regarding the one of user A, so user B sees a cube in front of him/her while user A considers the same cube to be on user B's left.

2.3.3 Collaborative Mode

The distribution of stage centers has a direct influence on how users interact and communicate with each other, so collaboration in multi-user immersive virtual environment can be divided into two modes depending on whether the spatial consistency is maintained among all the users.

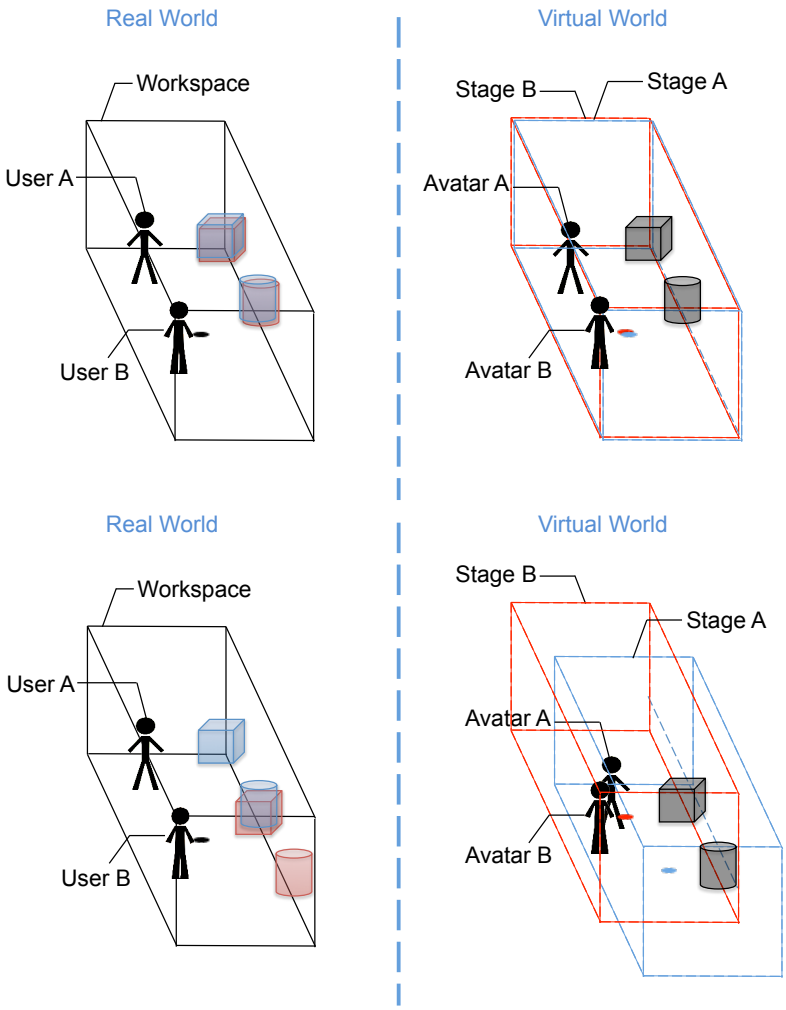


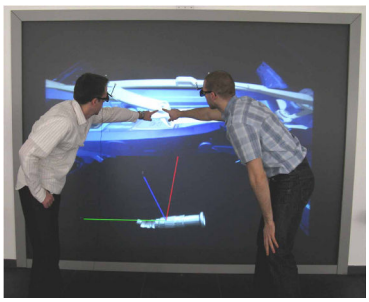
Figure 2.8 – Distribution of users' stage centers: consistent mode (above) and individual mode (below).

2.3.3.1 Consistent Mode

In this mode, all users' virtual spaces are consistent with the common physical workspace, which results in a shared spatial understanding of the virtual environment. In this context, users can communicate spatial information similarly as in the real world. For example, one can speak to another "Pass me the book on your left", or point to objects or directions by deictic gestures (Salzmann et al., 2009b) (Figure 2.9a). Another advantage of consistent mode is that users can make use of tangible devices (props) to get passive tactile feedback for object co-manipulation (Aguerreche et al., 2009; Salzmann et al., 2009a) (Figure 2.9b) or for other specific tasks like two-user driving test (Salzmann and Froehlich, 2008), etc.

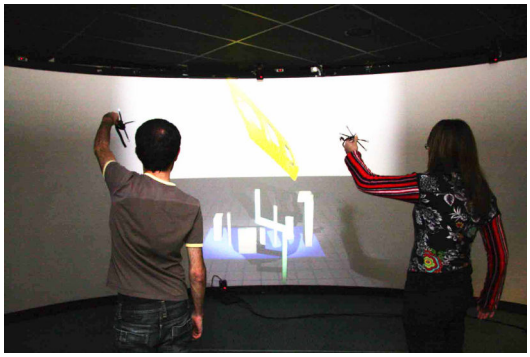
In projection-based immersive system, we can consider that users' physical bodies are directly "integrated" into the virtual environment. Social cues that facilitate human communication (e.g. gestures, postures, facial expressions and gaze direction, etc.) can be conveyed without computer mediation. This is a big advantage compared to remote situations where users communicate through embodied avatars.

The consistent mode is initially motivated by efforts to bring distortion-free stereoscopic view for each tracked user, and for a long time it is considered to be the default mode for co-



(a) Direct user interaction with deictic gestures (Simon, 2005).

(b) A windshield assembly task using tangible interface (Salzmann et al., 2009a).



(c) Object co-manipulation with flysticks (Aguerreche et al., 2010).

(d) MalCoMIICs demo for car assembly task (Martin et al., 2011).

Figure 2.9 – Co-located collaborations in consistent mode.

located collaboration. This consistent mode suits well for closely-coupled collaborative tasks (Simon, 2005; Aguerreche et al., 2010; Martin et al., 2011) (Figure 2.9c and 2.9d) that do not require users to be at or travel to different virtual places.

This spatial constraint restricts collaborative scenarios that can be supported in multi-user immersive virtual environment, though several interaction techniques have been proposed to get around spatial constraints and to extend collaborative scenarios in consistent mode, such as the bent pick ray (Riege et al., 2006) and the see-through techniques (Argelaguet et al., 2010).

2.3.3.2 Individual Mode

As opposite to consistent mode, the individual mode loosens constraints applied to users' stage centers. This mode is complementary to the consistent mode and largely broadens collaborative scenarios that can be supported in the co-located use of immersive virtual environment. Collaborative scenarios can be extended in the following ways:

- Spatial distribution of users in the virtual world becomes more flexible:
  - Users are distributed at remote virtual places;
  - Users can directly see each other, but their distance in the virtual world is larger than what the available physical workspace can offer;
  - Users are close to each other around an object of interest, but from different perspectives that they can get in consistent mode;

- Artificial control of users' viewpoints:
  - To provide user with a third-person view or a god view of the virtual scene;
  - To enable multi-scale observation (zoom on a particular zone);
  - To quickly share one's first person view to others, or to exchange viewpoints between users (Lopez et al., 2014).

The individual mode can be very useful, even indispensable in some cases. For example, when two users working face-to-face in a projection-based immersive system, objects situated between them can not be correctly perceived since images on the screen are occluded by user's body. In this case individual mode can be applied to allow users to be virtually face-to-face while being side-by-side physically.

In fact, individual mode puts users in an intermediate state between remote collaboration through CVE where all interactions are mediated by the computer network, and consistent co-located collaboration where users can interact the same way as they do in the real world. In individual mode, the physical collocation still allows direct user communication in verbal and non-verbal form (e.g. one wants another to stop moving a table, he or she can simply say "stop" or show a stop hand gesture). However, since users' spatial relationship (relative position and orientation) in the virtual world is no longer constrained by the one in the physical workspace, direct user interaction and communication involving spatial information may become confusing (e.g. the same virtual object will appear at different spatial locations for users who do not share the same spatial reference frame). In this case, embodied avatars as those used in remote situations are often needed to allow coherent user interactions.

2.3.3.3 Mode Switching

Multi-user immersive virtual environments should support both consistent and individual collaborative modes to cover as more collaborative scenarios as possible, and also to integrate advanced viewpoint control that facilitate user communication and understanding of the situation.

The transitions between the two modes are achieved by splitting or merging users' stages through automatic or guided virtual navigation and viewpoint control. Here we summarize different ways to activate mode switching:

- User initiated commands (e.g. button, gesture or vocal command, etc.);
- Automatic transitions (i.e. embedded in the story line of the task);
- Event-driven design (e.g. triggered when users enter or leave a certain zone in the virtual world).

2.4 Virtual Navigation

The individual mode in multi-user immersive virtual environment supposes that users may navigate separately in the virtual world. Actually, navigation is a basic type of interaction for users

to accomplish various tasks in the virtual world. This is the reason why it is time to discuss existing navigation techniques of different nature that allow users to travel through large virtual space.

Navigation methods used in immersive virtual environments are in general different with desktop-based ones. At least two requirements should be met when designing navigation technique in such context:

- Desktop-based input devices like mouse and keyboard are not preferred as they are not compatible with behavioral interfaces;
- Users should have access to an infinite virtual world while staying in a limited physical workspace.

Below we present a taxonomy of existing navigation techniques by their control law and a virtual vehicle model that we propose as a conceptual model to design rate control navigation techniques. Then we discuss how co-located users navigate in the same immersive virtual environment.

2.4.1 Taxonomy

2.4.1.1 Position Control

Natural walking is considered to be the most intuitive way to explore the virtual environment (Ruddle and Lessels, 2009). However, due to the limited size of available physical workspace, we need additional controls to enable infinite walking in restricted real workspace. Both hardware solutions like various locomotion devices (e.g. treadmills (Iwata and Yoshida, 1999)) and software solutions (e.g. redirection (Peck et al., 2008), resetting (Williams et al., 2007) and scaling techniques (Interrante et al., 2007)) are proposed to tackle the space limitation. Other metaphors like walk-in-place (Razzaque et al., 2002) and WIM (World-In-Miniature) (Stoakley et al., 1995) are also interesting alternatives.

2.4.1.2 Rate Control

Unlike previous techniques, rate control techniques activate virtual navigation by giving a velocity to user's stage to change constantly its position and orientation in the virtual world. Users can have the sensation of navigating (self-motion illusion or vection (Riecke and Feuereissen, 2012)) without moving physically. Actually the navigation can be controlled by information coming from different sources, for example, various input devices like joystick, haptic arm (Martin et al., 2012) or even specific locomotion devices (Marchal et al., 2011). With video cameras or optical tracking systems, users can specify the navigation velocity by motion tracking data of the hand (camera-in-hand (Ware and Osborne, 1990)) or head movements (Bourdot and Touraine, 2002), and also gestures (Konrad et al., 2003) or postures (von Kapri et al., 2011). Bowman et al. (2004) named this kind of virtual navigation techniques steering metaphors which are often relatively easy to implement and can provide efficient and flexible control of virtual navigation.

2.4.1.3 Mixed Control

Some navigation metaphors such as the bubble technique (Dominjon et al., 2005) and the magic barrier tape (Cirio et al., 2009) combine both position and rate control in order to enable infinite navigation within restricted real workspace. Position control is used within the physical workspace and then rate control is applied to the virtual vehicle to move further in the virtual world. Cirio et al. (2012) summarized several metaphors for safe navigation in a restricted cubic workspace. Moreover, Fleury et al. (2010) proposed a general model to integrate physical workspace into the virtual world and make the user aware of the physical environment in different ways.

2.4.2 Virtual Vehicle

Rate control navigation techniques can be modeled with a virtual vehicle model. The concept of vehicle was originally used by Bourdot and Touraine (2002) to describe the whole workspace of a user. With this vehicle model, the navigation is always effectuated based on the center of the physical workspace. Later on, in the virtual cabin model of Fleury et al. (2010), a separate reference frame called "conveyor" was introduced to allow navigation control with an offset to the center of the workspace.

In this thesis, we redefined the virtual vehicle, which behaves like a special type of conveyor that is superimposed with user's avatar in the virtual world, so from user's first person viewpoint, the vehicle is always situated underneath user's current location, or we can say that the user is on board his/her personal vehicle in the virtual world in order to navigate. As illustrated in Figure 2.10, the user navigates as the physical vehicle moves forward in the virtual world (the stage is "carried" by the vehicle). The vehicle has a constant orientation offset with the stage, but the position offset changes as the user moves in the real environment. In this way the rotation is always centered on user's avatar position no matter his/her location with respect to the physical workspace.

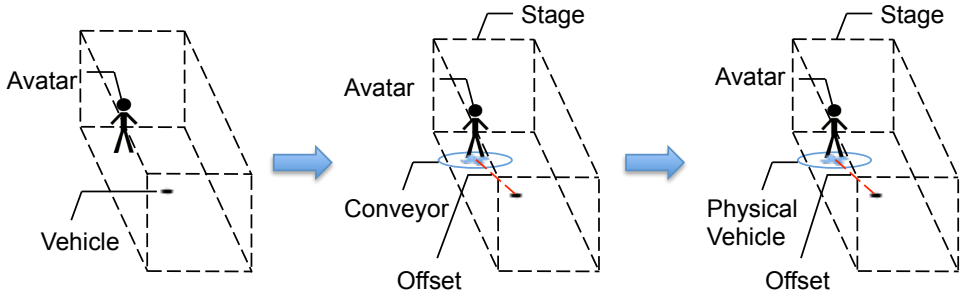


Figure 2.10 – Evolution of the virtual vehicle concept.

We name it as vehicle instead of conveyor mainly for two reasons: first, rate control navigation techniques provide similar navigation experience as we drive or pilot vehicles in the real world; second, we can reproduce some "realistic" vehicle-based navigation behaviors (e.g. collisions, frictions) by transforming the reference frame used for rendering to a virtual entity possessing physical properties.

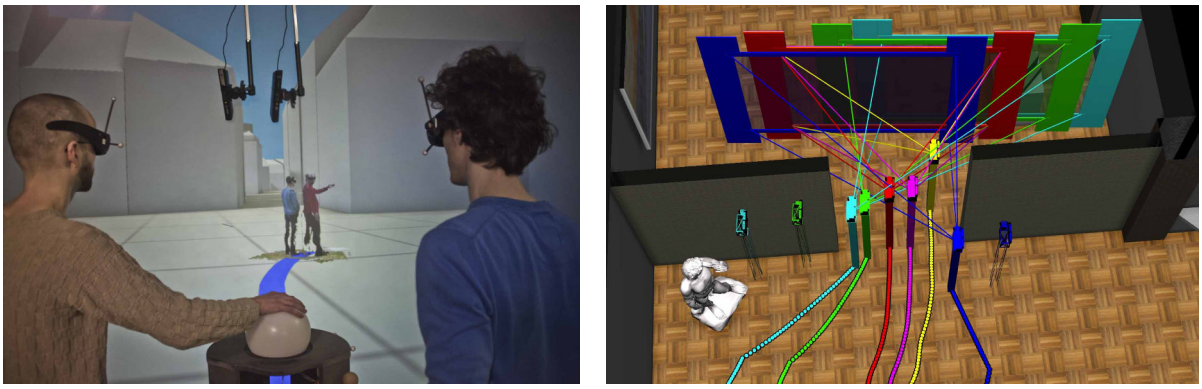
As a summary, the virtual vehicle provides a standard abstraction that facilitates the design of rate control navigation techniques, we just need to concentrate on how to control the vehicle.

This conceptual model also makes it easier to understand user's velocity perception within immersive displays. When the user can move in the physical workspace, the navigation velocity perceived by the user is the sum of the virtual vehicle's velocity and user's velocity in the real workspace:

$$\overrightarrow{v_{perceived}} = \overrightarrow{v_{vehicle}} + \overrightarrow{v_{user}} \tag{2.1}$$

2.4.3 Navigation in Multi-user IVE

In a multi-user immersive virtual environment, existing studies on co-located collaboration focused mainly on closely-coupled collaboration (e.g. co-manipulation) in consistent mode. In this situation, virtual navigation is often disabled. When users need to change the place for collaboration, navigation is conducted in a leader-follower mode that one user controls the navigation for the whole group. Spatial consistency is always maintained (Beck et al., 2013) (Figure 2.11a) or temporarily disabled (Kulik et al., 2011) in certain cases to avoid virtual obstacles (Figure 2.11b).



(a) Group navigation controlled by the user on the left using a trackball (Beck et al., 2013).

(b) Users' viewing frustums are shifted to avoid passing through the wall (Kulik et al., 2011).

Figure 2.11 – Navigation of co-located users controlled by a leader.

The virtual vehicle along with the IIVC model can be easily adapted to manage multiple co-located users in immersive virtual environments. Each user has his/her own copy of stage and virtual vehicle (Figure 2.12) which allows individual navigation. Group navigation can also be achieved by merging all the stages and giving the control of all vehicles to a leader.

In individual mode, individual navigation can be achieved by providing independent navigation metaphors for each user that gives users full control on their own virtual locations. However, using navigation metaphors based on users' physical movements may be problematic as users share the same physical workspace. Thus multi-user navigation models with individual navigation control should comply with user cohabitation constraints that will be discussed in the following section.

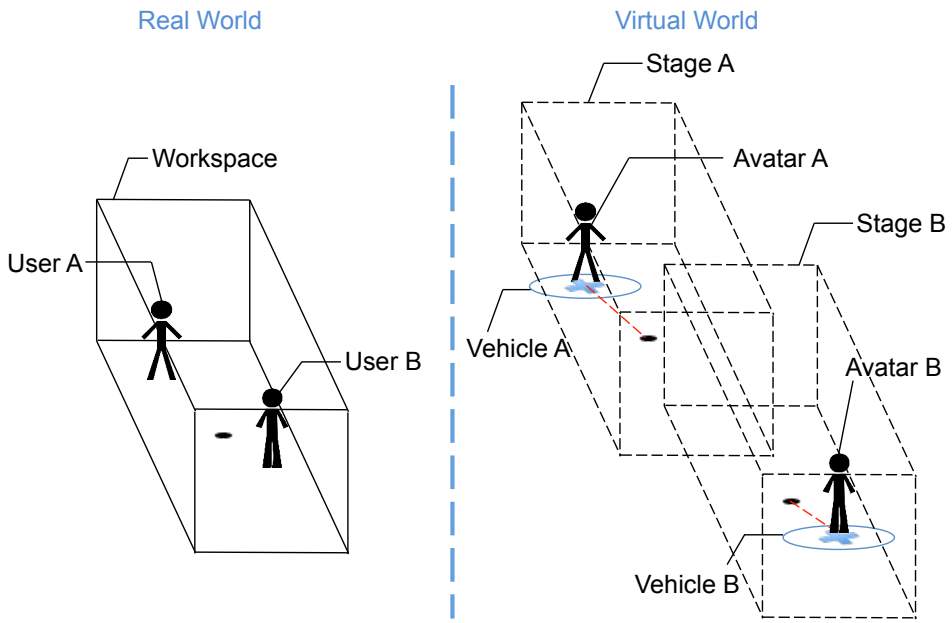


Figure 2.12 – Illustration of virtual vehicle based navigation.

2.5 Addressed Issues

As presented above, co-located collaboration in IVE contains two different collaborative modes: the consistent mode and individual mode. In consistent mode, users have a distribution of virtual spaces that are coherent with their physical workspace, which offers a similar interaction experience as in real world. On the contrary, users in individual mode have independent viewpoints, which broadens supported collaborative scenarios with more flexible viewpoint control.

However, due to the inconsistency between users' spatial distributions in the virtual world and real workspace, the individual mode provides users with a new kind of perceptual immersion and related cognitive experiences as they need to handle information from both real world and the virtual scene at the same time. We identified two main categories of spatial issues: perceptual conflicts during user communication and interaction, and user cohabitation in limited physical workspace. These issues should be studied if we want to validate the use of individual mode for co-located collaboration in IVEs.

2.5.1 Perceptual Conflicts

Communications between collaborators are usually conducted in a multimodal way (Paggio and Jongejan, 2005) including verbal and non-verbal modalities (Ennis et al., 2010; Dodds et al., 2011), especially deictic gestures to refer to objects or places. In individual mode, this kind of spatial interactions can be restored by introducing embodied avatars just like in remote situations - each user interacts temporally with other users' avatars instead of the real user in his/her own "copy" of virtual world. As a consequence, multimodal interactions in individual mode lead to two kinds of perceptual conflicts:

First, in projection-based immersive systems, users sharing the same display may sometimes enter each other's field of view when they work with embodied avatars. The simultaneous per-

ception of a user and his/her avatar leads to a special experience that we called "dual-presence". The dual-presence offers conflicting visual cues because the perceived real user and corresponding avatar share some common properties in terms of appearance and body movement (animated by real-time motion capture).

For example, in a two-user collaborative scenario (Figure 2.13), user A is pointing at a cube next to him/her for user B to delete it. User B may be troubled by seeing both the real user A and his/her avatar doing the pointing gesture, and it may be more confusing when there happens to be another virtual object (e.g. a cylinder) in the pointing direction of real user A.

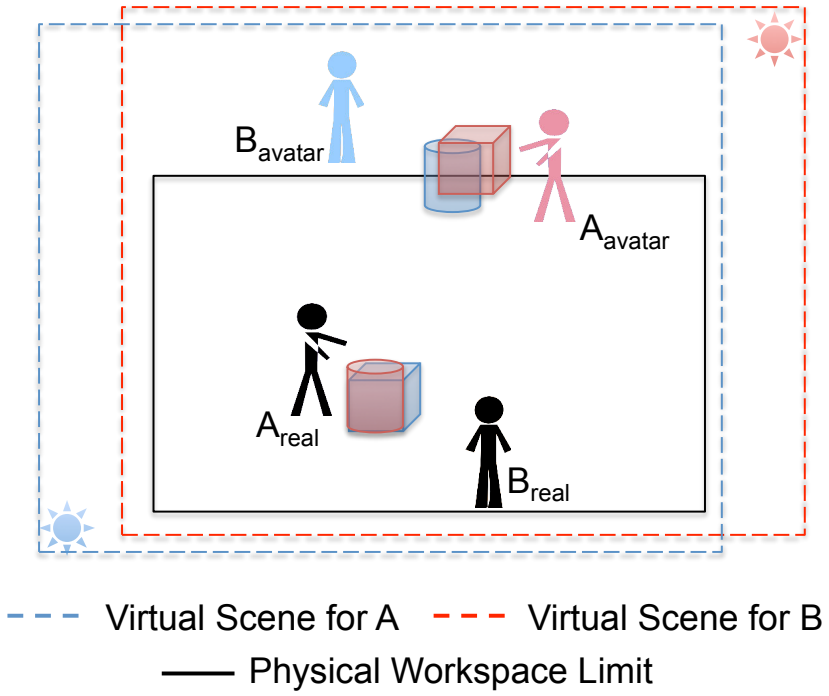


Figure 2.13 – Illustration of dual presence with a two-user scenario, two users are virtually face-to-face while standing side-by-side. User B simultaneously perceives that the real user A and A's avatar is pointing at different virtual objects.

Second, in individual mode, users dialogue with each other while interacting via embodied avatars. Since co-located users talk to each other directly without computer mediation, hearing someone from a different location than his/her visual representation (i.e. avatar) results in a visual-auditive conflict. Still in a two-user example (Figure 2.14), user B is looking at user A's avatar in front, but the audio source (i.e. real user A) is on his/her left.

Although according to Spence (2013)'s review, spatial coincidence does not represent a general constraint on multi-sensory integration in humans, but would be a much more task-dependent phenomenon. However, in a multi-user IVE, this visual-auditive conflict could be amplified due to the flexibility of virtual environment. For example, users could be far away from each other in the virtual world and they still hear each other from a close distance.

These two categories of perceptual conflicts are interrelated since they all depend on the distance between a user and his/her avatar. We need to study the usability of individual mode for co-located collaboration by investigating users' reaction to dual-presence and visual-auditive conflicts. The effects of such perceptual conflicts will influence choices on the design of co-located collaborative immersive systems. For example, whether we should prevent users from

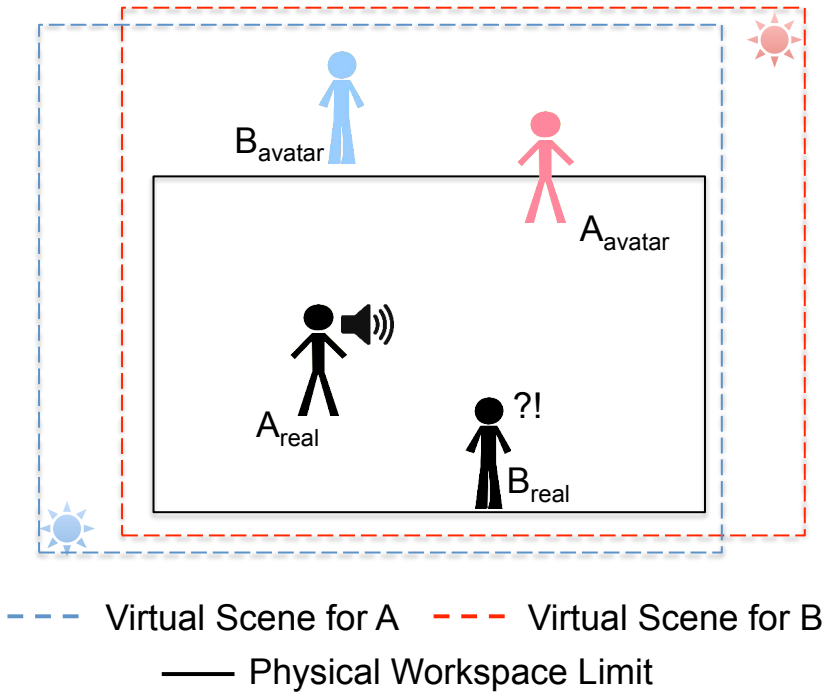


Figure 2.14 – Illustration of visual-auditive conflicts with a two-user scenario. User B experiences visual-auditive spatial conflicts when focusing on user A's avatar in front while talking to the real user A situated at a different location.

seeing the real users to avoid dual-presence, whether we need headphones to map users' speech to avatars' positions by 3D audio, etc.

2.5.2 User Cohabitation

Users in a multi-user IVE are subjected to two kinds of spatial constraints: users are inside a limited physical workspace, and they share the same workspace. These two constraints lead to different aspects of user cohabitation management: how to manage the spatial relationship between active users and the workspace boundaries while keeping good visual immersion? How to allocate personal workspace for users so that they won't collide or occlude each other?

Immersive virtual environment has the power to bring users to an artificial world by blocking the perception of the real world. In such situation users often forget boundaries of the physical workspace due to visual immersion, so when they move around in such system they risk to run into walls (with HMDs) or screens (projection-based display) which could endanger both the user and the device. Moreover, visual immersion is limited by functional tracking area and the screen configuration (for projection-based display). For example, in large projection-based immersive systems with non-closed displays such as large image wall or a 3-wall CAVE, visual immersion is limited to the display area (Figure 2.15).

When multiple users share the same immersive device for co-located collaboration in individual mode, they may run into collision when they move around without paying attention to the others. In projection-based immersive systems, one's visual perception of the virtual world could be disrupted by another if the latter appears to be in the field of view of that user due to body occlusion (Figure 2.16).

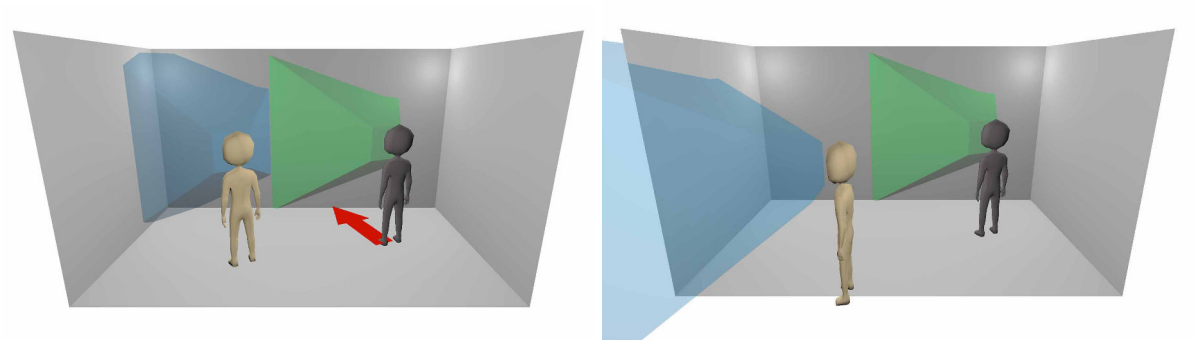


Figure 2.15 – Illustration of workspace boundary for translation (left) and looking direction (right) in a multi-stereoscopic 3-wall CAVE.

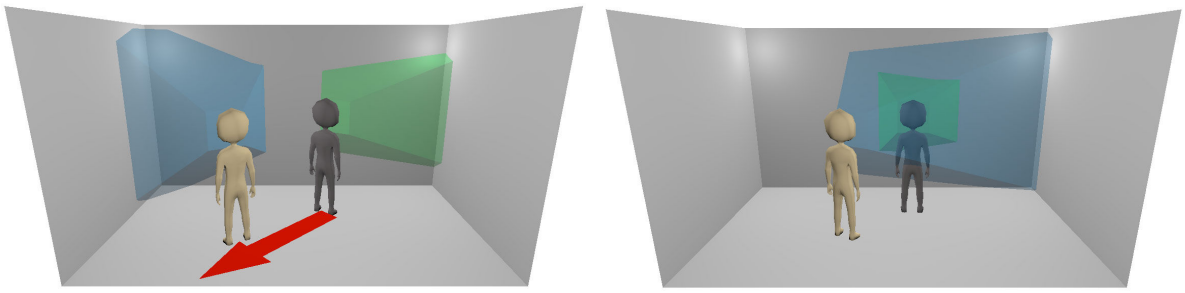


Figure 2.16 – Illustration of collision (left) and occlusion (right) between co-located users in a multi-stereoscopic 3-wall CAVE.

Depending on the choice of display technology and screen configuration, not all aforementioned issues are present in all types of immersive systems. Here Table 2.1 divides all cohabitation issues into different categories: some are inherent issues of the individual collaborative mode, others are more system-dependent, i.e. depending on the type and configuration of used immersive display. For example, user occlusion merely appears in projection-based immersive system, and perceiving "empty screen" is only possible with non-closed immersive display.

Table 2.1 – Cohabitation Issues in Multi-user Immersive Systems.

	Workspace-related	User-related
Generic	Collision with workspace border (e.g. screen wall of CAVE, walls of the room when wearing HMD) or outside tracking zone	Collision between users
System-dependent	Seeing non-projected area (non-closed systems like image walls or open-CAVE)	User occlusion (projection-based systems)

For safe and efficient use of multi-user IVEs, new paradigms are needed to enable individual navigation in infinite virtual space while keeping users' level of immersion and safety.

2.6 Approach

To assess users' reactions to certain stimuli (e.g. perceptual conflicts) or the performance of novel paradigms (e.g. paradigms for navigation or object manipulation) in virtual environments, experiment-based user evaluation inherited from HCI research has become a mainstream method. Besides formative user experiment, there are many other evaluation methods such as cognitive walkthrough, heuristic expert evaluation, interview, post-hoc questionnaire and so on, each can help us to gather some information from a different angle with certain cost (Bowman et al., 2002).

In the context of this thesis, our goal is to study spatial issues related to individual collaborative mode in order to design a valid and efficient framework for co-located collaboration in IVE. All experiments discussed below are conducted in projection-based immersive system in which all spatial issues that we mentioned are present. The main experimental platform named EVE (Evolutive Virtual Environment), is a four-screen multi-user immersive CAVE system built with the active & passive separation technique (see Appendix B). It serves as an experimental tool to study issues related to user interaction and communication for co-located collaboration in projection-based multi-user immersive virtual environment.

The first step is to study effects and influencing factors of perceptual conflicts due to the introduced avatar representation for co-located collaboration. We want to examine how perceptual conflicts generated by dual-presence would alter users' communication and their task performance. In a two-user case study, participants receive instructions from an experimenter to accomplish an object-picking task, we artificially created different levels of perceptual conflicts by varying the distance between the real experimenter and her avatar, and also by changing modalities involved in the instructions (verbal and/or gestural). A post-hoc questionnaire was designed to collect users' subjective feelings towards dual-presence and their understanding of the instructions. This experiment is presented in detail in Chapter 3.

Second, we concentrate on the design and evaluation of appropriate navigation paradigms to allow individual navigation while solving user cohabitation problems in a progressive way. The paradigm that we propose is based on a Human Joystick metaphor controlled by head tracking (Bourdot and Touraine, 2002). First we evaluated this metaphor by a comparative experiment (Chen et al., 2013) to confirm its advantages over traditional navigation solutions (e.g. joystick). Then we created an Altered Human Joystick paradigm by adding additional controls to manage users' relationship with system borders and to minimize inter-user interferences.

At last, by taking into account feedbacks of previous studies, we continue to make improvements on the human joystick paradigm in the aim of providing a generic framework for co-located collaboration which integrates constraints from the physical workspace and virtual world into the navigation control as described in Chapter 5.

2.7 Conclusion

The research described in this thesis consists of investigating how to enable individual mode to better support co-located collaboration in multi-user IVEs. The first part of this chapter outlines motivations and techniques to create immersive systems for co-located collaboration, and the second part presents notions around spatial organization of users' workspaces and how

we navigate in IVEs. Then the last part summarizes issues related to spatial aspects of co-located use of IVE, it also presents research efforts that we made to study these issues and to propose solutions accordingly.

In following chapters, we present separately our studies on perceptual and cohabitation issues that we identified in the aim of building a general framework for co-located collaboration in IVE that allows both consistent and individual modes.

3

Experiment on Perceptual Conflicts

Chapter Abstract

This chapter presents a formative user experiment combined with post-hoc questionnaire to investigate users' reactions and subjective feelings towards perceptual conflicts caused by dual-presence in a two-user case study.

Contents

3.1 Introduction

In a multi-user IVE, spatial consistency is no longer maintained when users enter individual collaborative mode. User interaction involving spatial information can be restored by introducing embodied avatars. However, multimodal interactions in individual mode could lead to two kinds of perceptual conflicts: visual conflicts induced by dual-presence and visual-auditive conflicts due to the shift of audio source.

These two kinds of perceptual conflicts may be the source of misunderstanding and disturbance during co-located collaboration. In this chapter, we concentrate on effects and influencing factors of perceptual conflicts caused by dual-presence. The second type of perceptual conflicts (i.e. visual-auditive conflicts) is not in the scope of this chapter, but will be subject of future studies.

This chapter presents a formative user experiment combined with post-hoc questionnaire to investigate user's reactions and subjective feelings towards perceptual conflicts caused by dual-presence in a two-user case study (Chen et al., 2014). We want to see how users work with their partners when the latter are both physically and virtually (via avatar) present, and especially when multimodal instructions are used to communicate spatial informations.

3.2 Experimental Aims and Expectations

When multiple users collaborate inside the same projection-based immersive system in individual mode, one user may perceive the real body and his/her avatar the same time. We would like to know in the first place, how users receive and handle this dual visual information. If we do not tell users to interact with the avatars, will they be misled by the physical presence of others? Even choose to work with the real person regardless the existence of the avatar?

We also want to observe to what extent perceptual conflicts induced by the dual-presence influence the efficiency of user communication. Configurations of users' spatial reference frames and their relative positions result in different levels of distance between one's avatar and body. It is interesting to see whether the proximity of these two representations brings stronger perceptual conflicts. Moreover, user communication involves multimodal interaction between users, so we would like to examine the correlation between the level of perceptual conflicts and the modality implied in the communication process. For example, whether users are more disturbed with deictic gestures or verbal commands facing the dual-presence.

Finally, as a first study of perceptual issues related to dual-presence, besides assessing effects of different variables that we define in the experiment, it is useful to collect users' subjective feelings regarding their experiences with the dual-presence of their partners during co-located collaboration, which will provide us more insights for future user studies and collaborative system design.

3.3 Experimental Design

3.3.1 Task

We designed a collaborative scenario where two users were situated in a virtual room of about the same size as the immersive setup (3-wall CAVE) with some floating crowns (Figure 3.1). The task was for a participant to pick one of the crowns according to the instructions given by a guide. The guide was played by an experimenter who was physically co-located with the participant and also had an embodied avatar in the virtual world.

Figure 3.1 – Front view of the virtual scene used for the experiment.

The embodied avatar was designed to match the experimenter's profile like the height, the clothes and the face (without facial expression) (Figure 3.2a). In order to make the avatar follow user's body movements, we used an optical motion capture system to map the experimenter's upper body motions to her avatar. A flystick was also provided for the experimenter to trigger the timer for task performance measurements (see Figure 3.2b). No avatar was provided for the participants since the task only involved one-way communication and we did not need to study the visual perception of the experimenter. The participants wore a pair of tracked shutter glasses to see stereoscopic images and a tracker on one hand to enable direct interaction with virtual targets (crowns) by hand.

We intentionally put participants in extreme situations where dual-presence were always present. For example, there was always a crown on the right of the real experimenter and another one on the right of her avatar from participant's point of view when the instruction was "Go to the crown on my right" Figure 3.3). Participants should choose between these two forms of collaborator in order to know which target to pick. We could extract their choice of collaborator by recording the chosen target ID. In the meanwhile, instructions were given in different forms (deictic gestures and/or verbal commands) in order to observe the influence of perceptual conflicts on different sensory channels (visual and auditory).

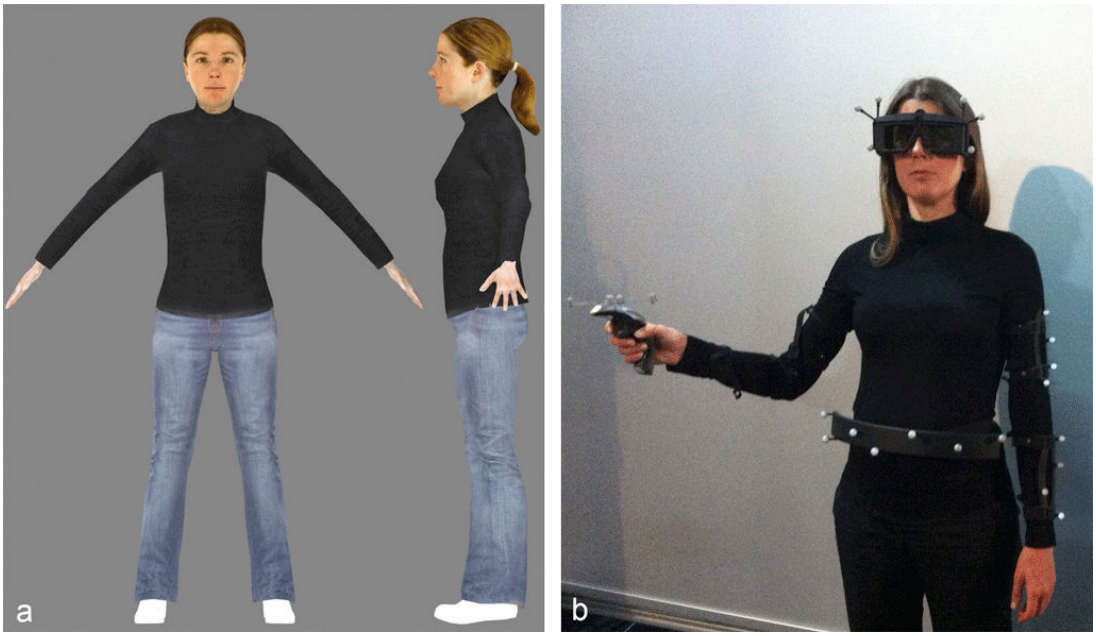


Figure 3.2 – a) The avatar of the experimenter used in the experiment. b) The experimenter wore a motion-tracking suit to map upper body movements to the avatar.

3.3.2 Participants

27 participants (23 men and 4 women) joined our experiment. The average age is 26 years (standard deviation 4.8 years). A large number (23) of the participants were people with a computer science background. All participants filled out a background questionnaire, which was used to gather demographic information such as their level of computer skills, experience with video games and with virtual reality devices, and some other personal statistics (e.g. age, dominant hand, etc.). Among these participants, 26 were right-handed and one person was left-handed. All participants used their dominant hand to perform the designation task and were naive with respect to the experimental setup and purpose of the experiment.

3.3.3 Independent Variables

In this experiment, we manipulated two independent variables to study their impact on participants' choice of collaborator and task performance. Here is a detailed description of these independent variables:

3.3.3.1 Delocalization Angle

As shown in Figure 3.4a, at the beginning of each trial, the participant should stand at a given point (the white square at the center of the circle) to wait for the experimenter's instruction. Around this starting point, seven predefined positions were distributed on a half circle with a radius of two meters symmetrically from the left to the right of the CAVE workspace (smaller white squares). For each trial, the experimenter and her avatar were located on one of these seven positions which were equidistant to participant's standing point, so the delocalization level between the real collaborator and her avatar can be measured by the angle between two

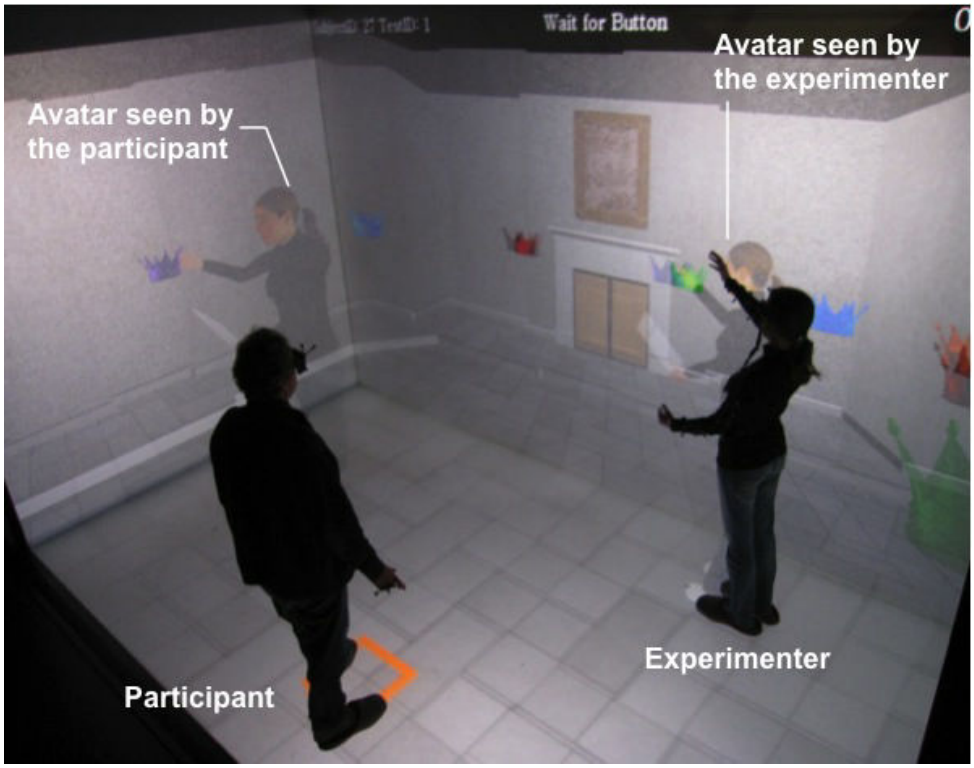


Figure 3.3 – The 3D view of the participant and that of the experimenter were superimposed on the multi-stereoscopic display. The participant faced the dual-presence of the experimenter and her avatar, while from the experimenter's viewpoint, the avatar was co-localized with her physical body.

segments started from the center of the circle (where the participant stood) and ended respectively at the positions of the experimenter and her avatar (see Figure 3.4b). This delocalization angle has seven different values from 0° to 180° with an interval of 30° given the configuration of the chosen seven locations.

3.3.3.2 Instruction Type

Interaction between users usually involves verbal and gestural modalities. In order to inspect their impact on co-located collaboration with dual-presence, we defined three deictic instruction groups: verbal-only instruction, gesture-only instruction, and multimodal (gesture and verbal) instruction. Each instruction group was then divided into two sub-groups to take into consideration the possible left-right effect. Finally we got six different instruction groups as indicated in Table 3.1. To be more precise, the target object (the crown) indicated by the experimenter was the nearest one (e.g., "Go to the crown on my left" indicated the closest crown on the left side of the experimenter when there were more than one presented target). To make sure that ambiguities were present no matter the actual spatial configuration, there was always one crown on the left-hand side and one on the right for each of the seven predefined positions of the experimenter.

We defined the combination of the experimenter's position and her avatar's position as the spatial configuration of the trial. Since there were seven possibilities for both positions (Figure 3.4(a)), this resulted in 49 different spatial configurations in total. Due to time limitation,

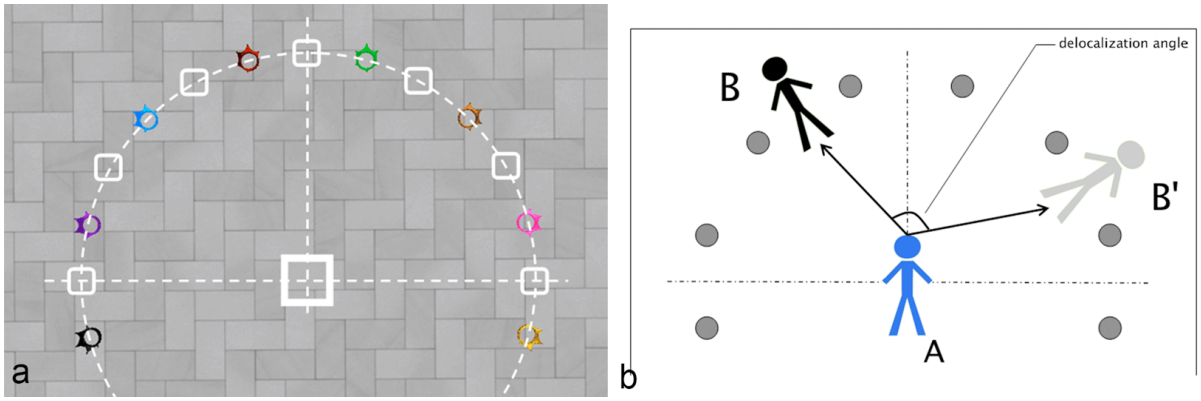


Figure 3.4 – a) The top view of the virtual scene indicates the configuration of the virtual room and predefined positions for participants and their collaborator. b) An example of the delocalization angle due to the separation of the real experimenter and the associated avatar.

Table 3.1 – All six types of deictic instructions given by the experimenter for the object-picking task.

No.	Type	Example
1	Gesture-only	Point to the crown on the left.
2	Gesture-only	Point to the crown on the right.
3	Verbal-only	“Go to the crown on my left.”
4	Verbal-only	“Go to the crown on my right.”
5	Multimodal	“Go to the crown on my left” (point to the left).
6	Multimodal	“Go to the crown on my left” (point to the left).

there were too many spatial configurations for each participant to test all of them, so we distributed these configurations among all the participants with each person testing a subset and covered all the conditions at a global level. We grouped the spatial configurations by the delocalization angle that they formed, and each participant passed one configuration for each delocalization angle with all six types of instructions. In order to offset the interaction affect between the delocalization level and the instruction type, all the trials for each participant were presented in a randomized order.

Finally, in this experiment, participants moved directly in the virtual scene through natural walking with a scale of one to reach the target object. This choice was taken in order to make the evaluation independent with any virtual navigation paradigm to avoid warping the results because some participants may be more familiar with virtual interactive navigation techniques than the others.

3.3.4 Procedure

First, each participant was invited to sign an informed consent and then an overview of the experiment was provided. This experiment was conducted in three main phases for each participant:

- Presentation stage: We introduced the main procedure of the experiment, the nature of instructions and the duration of each step to the participant. A pre-questionnaire about cybersickness was filled out.

- Learning stage: Six randomized trials were given for the participants to get familiar with the experimental setup.

- Experimental stage: The participant accomplished collaborative tasks under the instructions of the experimenter. The experimental stage for each participant was the same except each one of them tested a subset of all the spatial configurations. Each participant tested seven spatial configurations with all six types of instructions, so in total 42 trials. The working scenario was the following:

  At the beginning of trial, the virtual scene was covered by a blue curtain (virtual one). We set up the curtain and asked the participant to lower the head in order to avoid the participant from seeing the experimenter and her avatar before the trial began. The participant should stay at his/her departure position and look at his/her feet until the curtain disappeared. Once the experimenter reached the required position indicated by the square on the floor (Figure 3.4(a)), the system then put the experimenter's avatar at another required position to form a delocalization angle (from 0° to 180°) and hid the curtain.

  The experimenter read the instruction showed on a piece of virtual paper in her left hand and then indicated the target (one of the crowns) the participant should travel to by giving verbal and/or gestural instruction, and then started the timer by pushing a button on the flystick. Then a beep was delivered by a loudspeaker to inform the participant to respond. Participants were not supposed to start before the beep, otherwise the trial was not counted.

  The system recorded the time when the participant left the initial position as reaction time (A participant had the right to give up and pass to the next trial before he/she reached a certain target). The system also stored the time from the moment the participant left the initial position until he/she reached a target (correct or not) as execution time.

  The selected crown ID (from 1 to 8) was recorded in order to know the participant's choice of reference frame. When the experimenter and the avatar were not superimposed, they would lead the participant to different targets depending on their actual positions and the given instruction. So we recorded the selected crown ID to see which target was chosen by the participant, and then we could determine whether the participant followed the real experimenter, or her avatar, or neither of them (considered as a failure). If the participant chose to give up, the trial was also considered as a failure of the task. When the task was finished, the participant came back to the initial position and waited for the next trial.

- Questionnaire stage: First, a questionnaire was filled by a participant before and after the experiment to evaluate the level of cybersickness (see section 3.3.5.2). Then we tested the participants with a detailed subjective questionnaire for the collaborative task (see Appendix C) and finally another questionnaire to get participants' personal information.

3.3.5 Measures

3.3.5.1 Performance Measures

For each trial, we recorded reaction time (noted as RT), execution time (ET) and the total time (TT, sum of the reaction and execution time) to finish the task. Moreover, we recorded participant's choice of collaborator for each trial using the crown ID.

3.3.5.2 SSQ Pre-assessment and Post-assessment

The questionnaire used for cybersickness evaluation is Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993). The SSQ was administered before and after exposure to the virtual environment. The SSQ is a 16-item measure in which participants report symptoms on a scale of 0 to 3 (0 = None, 3 = Severe). Three types of symptoms are assessed: oculomotor dysfunctions (O) (eyestrain, blurred vision, difficulty in focusing), mental disorientation (D) (difficulty in concentrating, confusion, apathy), and nausea (N) (including vomiting). Unit scores (O, D, N) are weighted scores. The SSQ is a widely used measure of simulator sickness that has been shown to be a valid measure of this construct in VR research (Cobb et al., 1999).

3.3.5.3 Subjective Measures

Data were collected using a subjective evaluation questionnaire with a 4-point rating scale, ranging from 0 (do not agree at all), 1 (do not agree), 2 (neutral), 3 (agree), to 4 (fully agree). We regrouped the items into three categories according to three measured dimensions. Nine items measured the participant's confidence on the success of the task, 11 items estimated the participant's understanding of instructions, and 15 items measured participant's reaction to the dual-presence of the experimenter.

In order to avoid directing participants in an affirmative direction (Dillman, 2000), all categories contained indicative items (e.g., "I chose targets randomly.") and counter-indicative items (e.g., "I easily identified the targets."). As an indication of how well a set of items measures a latent construct, we used Cronbach's alpha. In other words, Cronbach's alpha indicates whether a set of items is a homogeneous set that covers the meaning of the theoretical construct. The higher the Cronbach's alpha, the more reliable the generated scale. The American Psychological Association considers a questionnaire as acceptable when the alpha coefficient is above 0.7. We computed Cronbach's alpha for the dimensions related to the feeling of confidence, the understanding of the instruction and the reaction of dual-presence. The values were satisfactory (between 0.71 and 0.82).

3.4 Hypotheses

H1 Participants would naturally interact with the avatar of the experimenter rather than the real person to accomplish collaborative task in a multi-user immersive virtual environment.

H2 It should take participants longer time to accomplish the task when the real experimenter and the avatar are close to each other than when they are far away (the proximity of the two forms of collaborator should result in stronger visual conflicts for the participants).

H3 Participants should be more efficient with verbal-only instructions than with gestural or multimodal instructions, as the dual-presence are enhanced with gestural commands.

3.5 Results

The results presented in this section were considered statistically significant when $p < 0.05$. All the analyses were performed with Statistica 9.

3.5.1 Collaborator Choice

First we gathered all trials to see which collaborator the participants chose as a reference frame for the object-picking task. Among all trials, 4.9% were recorded as failure (either the participant gave up or he/she picked a wrong object) and 13.6% of all trials could not give us useful information on participants' choices because in these trials the experimenter and her avatar were superposed (delocalization angle was 0°). Then in 50.2% of all trials participants chose the target indicated by the avatar, and in 31.3% of all trials participants picked the target indicated by the real experimenter.

Chi-square (χ^2) test on all successful trials (81.5%) revealed that there was no association between the delocalization angle and the choices made by the participants. Chi-squared (χ^2) test also showed that there was no association between the instruction type and participants' choices. In order to have a better understanding of these results, we regrouped trials by participant and then counted for each participant how many times one chose the real experimenter and how many times the avatar. We found that most participants had an a priori choice of reference frame. This result, observed on the behavior of the participants during the task phase, was also confirmed by their answers to the subjective questionnaire for statements like "I only interacted with the virtual assistant".

Based on this result, we decided to divide the participants into subgroups to get a closer look on the impact of individual characteristics on their behavior. To define a criterion for clustering, we used the k-means method that took each participant's number of choices for the real experimenter and the avatar as input. The best estimate using DBI (Davies-Bouldin index) as metric was obtained for a number of clusters k = 3 (tested for $2 < k < 27$). Thus we created three groups: the Real, Virtual and Mixed groups with respectively 9, 16 and 2 participants (see Figure 3.5). The Real group corresponded to the participants who mainly chose to interact with the real experimenter and the Virtual group for those who mainly relied on the avatar. The Mixed group contained participants who kept changing their collaborator choice throughout different trials during the experiment.

From Figure 3.6, we can get a global view of the difference between the three groups by counting the number of referential choice transitions in each trial. A referential choice transition occurs when a participant modifies his/her reference frame (Real to Virtual, or vice versa). For

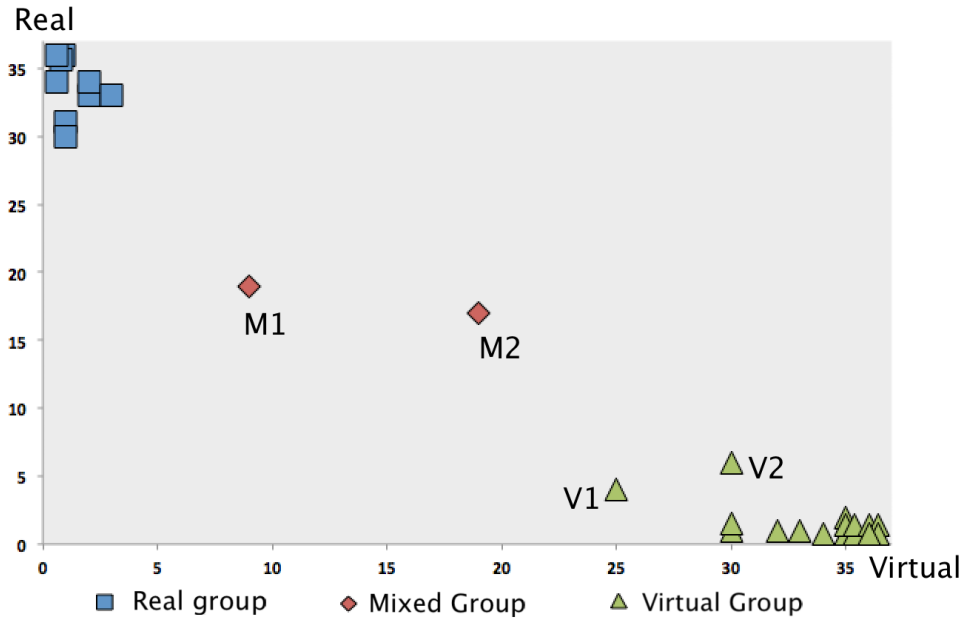


Figure 3.5 – The distribution of participants according to their number of choices on the virtual (avatar) and real (the experimenter) reference frames.

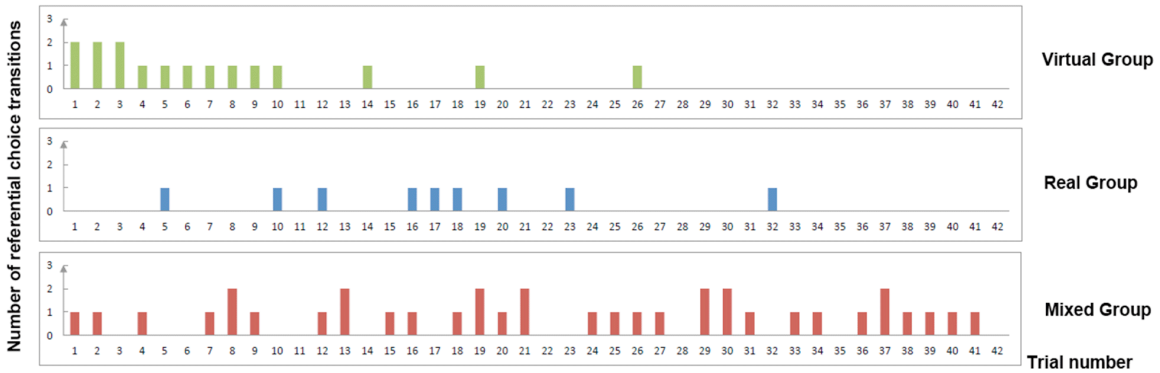


Figure 3.6 – Distribution of the number of collaborator choice transitions for each group.

each trial, the number of transitions corresponds to the number of participants having such transition.

In the Virtual and Real groups, participants chose, from time to time, a reference frame which was not relevant for their group: the experimenter for the Virtual group, her avatar for the Real group. In these groups, the number of transitions was quite low if we compare it with the maximum value that it could reach (group size). For the Mixed group, we chose as baseline the virtual referential, so that what was shown in the graph was the number of participants who switched from the avatar toward the experimenter.

Anyway, since there were only two participants in the Mixed group, this group was not included for the comparison between groups in the following analyses.

3.5.2 Performance

We applied the Kolmogorov-Smirnov test to verify if variables satisfy normality assumptions. All three behavioral variables failed the test: Reaction time (RT), K-S d = 0.21467, $p < 0.01$; Execution time (ET) K-S d = 0.17533, $p < 0.01$; Total time (TT) K-S d = 0.16585, $p < 0.01$. Therefore we used a non-parametric Kruskal–Wallis ANOVA test with delocalization angle and instruction type as inter-trial factors. For Post-hoc analyses, we used multiple comparisons of mean ranks for all groups.

3.5.2.1 Global Analyses

Regarding task efficiency (RT – Reaction Time, ET - Execution Time, TT – Total Time), first we examined the influence of delocalization angle (Table 3.2) and we found a main effect on the ET (H (6, N=1079) = 20.51355, p = 0.002) and TT (H (6, N=1079) = 18.68352, p = 0.004). Post-hoc comparisons showed that participants spent more time to execute the task with angle 180° than with 0°, 30° and 90°. For TT, participants spent more time to complete the task with angle 180° than with angle 0° (Figure 3.7).

Table 3.2 – Mean value and standard deviation of reaction time (RT), execution time (ET) and total time (TT) for each delocalization angle.

Delocalization Angle	Reaction Time (millisecond)		Execution Time (millisecond)		Total Time (millisecond)	
(degree)	Mean	SD	Mean	SD	Mean	SD
0	1482	1062	1592	607	3074	1263
30	1779	1543	1646	755	3425	1905
60	1570	966	1698	784	3268	1271
90	1633	1188	1599	629	3232	1520
120	1764	1746	1606	613	3370	1949
150	1946	1901	1852	1000	3798	2336
180	1770	1508	1891	972	3661	1754

Second we checked the relationship between task efficiency and instruction type, and we found a main effect on RT (H (2, N=1078) = 86.98551, p = 0.00001) and TT (H (2, N=1079) = 41.48673, p = 0.00001). Post hoc comparisons showed that gesture-only instructions required the most time and the results were intermediate with multimodal instructions (Figure 3.8).

3.5.2.2 Inter-group Analysis

We measured the influence of delocalization angle and instruction type in each group separately to see if they had an effect on a certain population.

For the Real group, we got a main effect of instruction type on RT (H (2, N=364) = 27.10793, $p < 0.000001$) and TT (H (2, N=364) = 10.5341, p = 0.005). Post hoc comparisons showed that

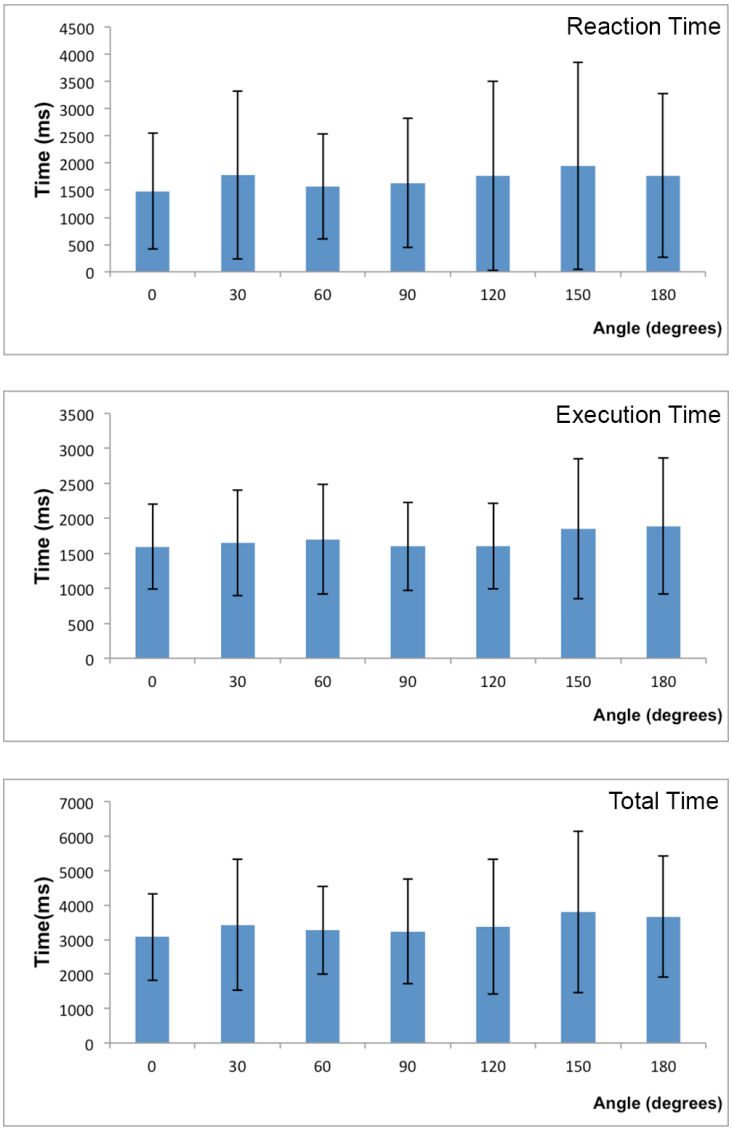


Figure 3.7 – Mean value and standard deviation of reaction time (RT), execution time (ET), and total time (TT) for each delocalization angle.

participants spent more time to react and to finish the task with gesture-only instructions than with the other two types of instruction while multimodal instructions were at a medium level.

For the Virtual group, a main effect of instruction type was also found on RT ($H(2, N = 640) = 60.62227$, $p < 0.000001$) and TT ($H(2, N=641) = 32.36614$, $p < 0.000001$). Post hoc comparisons showed the same result as for the Real group.

3.5.3 Subjective Data

3.5.3.1 Global Analyses

We gathered participants' subjective feelings as a complement to performance measures. First we used the SSQ to see possible occurrence of cybersickness after the exposure to the experimental perceptual conflicts. Then we extracted some descriptive results (e.g., confidence

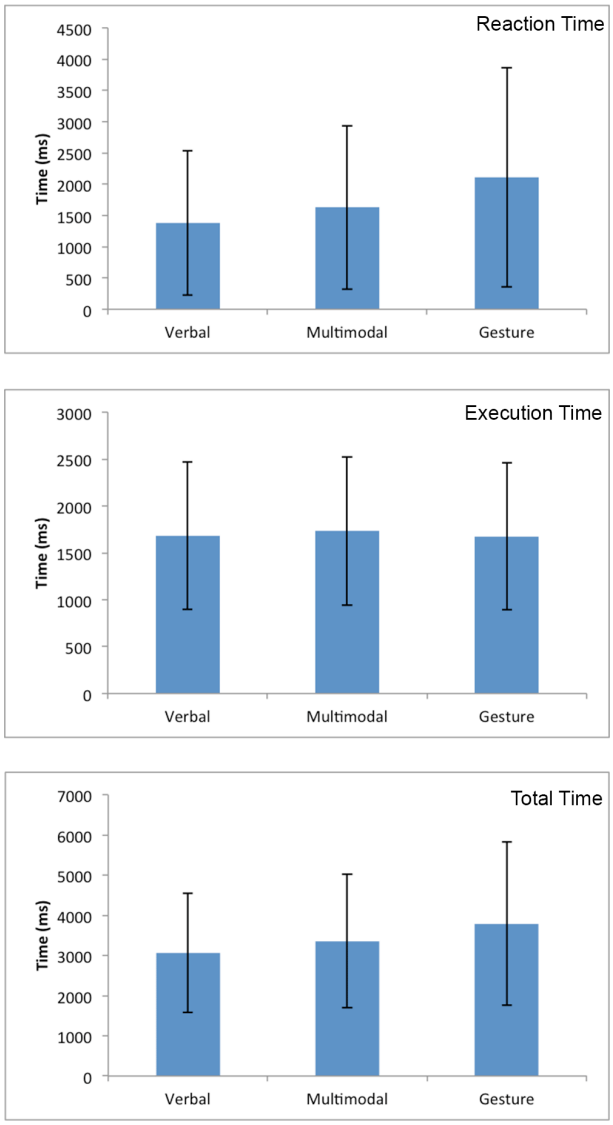


Figure 3.8 – Mean value and standard deviation of reaction time (RT), execution time (ET), and total time (TT) for each type of instruction.

feeling, etc) from the subjective questionnaire.

The simulator sickness was assessed before and after the experimental session using the SSQ. We found that the scores (by subtracting the pre-experimental measures to post-experimental measures) to the various sub-scales are relatively low. To study the impact of such immersive system (i.e., simulator effect), we applied the Kolmogorov-Smirnov test to verify that the variables failed to satisfy normality assumptions (Pre-score: K-S, $d = 0.25607$, $p < 0.05$; Post-score: K-S, $d = 0.29235$, $p < 0.01$). So we computed a Wilcoxon signed rank test on two paired samples on the pre (Mean = 164.21, SD = 221.34) and post (Mean = 155.02, SD = 212.07) scorings of SSQ. No effect of cybersickness was found after the experiment ($N = 17$, $T = 68$, $Z = 0.402373908$, $p = 0.687409138$).

From the subjective questionnaire, we found that the participants were moderately confident in their success of tasks and they felt only a little nervous about the dual-presence during the collaboration (Table 3.4).

Table 3.3 – Mean value and standard deviation of reaction time (RT), execution time (ET) and total time (TT) for each type of instruction.

Instruction Type	Reaction Time (millisecond)		Execution Time (millisecond)		Total Time (millisecond)	
	Mean	SD	Mean	SD	Mean	SD
Verbal	1384	1157	1682	785	3065	1481
Multimodal	1627	1306	1731	789	3358	1658
Gesture	2113	1755	1677	785	3790	2036

Table 3.4 – Global result of the subjective questionnaire for all participants.

Item	Mean	SD	Range
Confidence of task success	29.00	5.46	[0, 40]
Instruction understanding	15.76	6.39	[0, 28]
Ease of dual-presence	16.92	5.32	[0, 36]
Experience with virtual environment (VE)	1.76	1.36	[0, 4]
Experience with Virtual Reality (VR)	1.24	1.27	[0, 4]
Experience with video games	2.08	1.00	[0, 4]

3.5.3.2 Inter-group Analyses

We conducted inter-group comparison to check the relationship between the participants' collaborator choice and their individual characteristics. Regarding the simulator sickness, with Wilcoxon signed rank test on two paired samples, no matter in which group, participants did not experience cybersickness (Virtual group: N = 16, T = 20.5, Z = 0.236939551, p = 0.812703838; Real group: N = 9, T = 11.5, Z = 0.422577127, p = 0.672604101).

We used the Student's t-test with subjective questionnaire responses as inter-subject variables to compare the Real and Virtual group. We found a main effect on instruction understanding between participants in the Virtual group and those in the Real group (Table 3.5). Participants in the Real group declared weaker instruction understanding than participants in the Virtual group (Figure 3.9). We also found an effect on user's video game experience. Participants in the Virtual group played video games more often than those in the Real group (Table 3.5).

3.6 Discussion

In the absence of instructions to impose the participants to collaborate with the avatar in our experiment, we left the participants to choose between the real or virtual collaborator. In the experimental scenario that we created, a virtual object was always present in the pointing direction of the real experimenter so the real pointing gesture became "meaningful". According to the result, the physical presence of the experimenter both in the visual and audio modalities

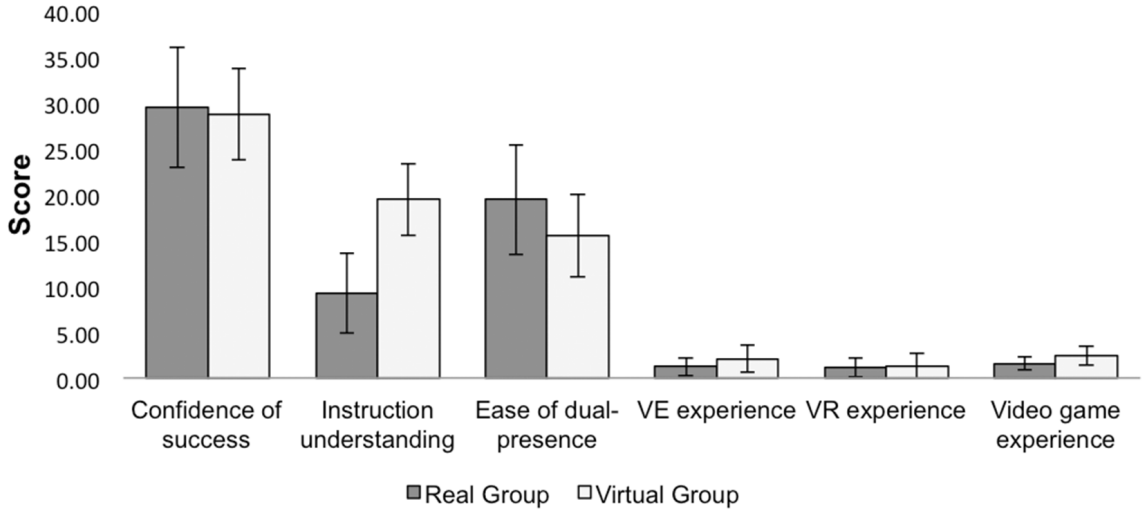


Figure 3.9 – Comparison of Real and Virtual groups according to the result of the subjective questionnaire.

Table 3.5 – Mean value and standard deviation of the result of subjective questionnaire and result of Student test between the Real group and Virtual group.

Item	Real group		Virtual group		Between groups	
	Mean	SD	Mean	SD	Student test	p-value
Confidence of success	29.44	6.54	28.75	4.97	t(23) = -0.30	0.77
Instruction understanding	9.22	4.32	19.44	3.90	t(23) = 6.05	**0.000004**
Ease of dual-presence	19.44	6.00	15.50	4.49	t(23) = -1.87	0.07
VE experience	1.22	0.97	2.06	1.48	t(23) = 1.52	0.14
VR experience	1.11	1.05	1.31	1.40	t(23) = 0.37	0.71
Video game experience	1.56	0.73	2.38	1.02	t(23) = 2.11	**0.046**

appeared to cause a significant proportion of participants to choose the real person instead of the avatar. In fact, 33% of all participants chose to work with the real experimenter and ignored the animated avatar. Meanwhile the rest (59%) interacted successfully with the avatar. So our hypothesis on participants' choice of collaborator (H1) was rejected. Three of those who finally chose the avatar have asked "Whom should I work with?" and we told them "Do as you like". However, none of the participants in the Real or Mixed group asked this question.

Moreover, almost all users made their choice in the early stage of the experiment, which either stayed stable during all trials or was gradually stabilized from trial to trial (especially participants V1 and V2 in Figure 3.5). Only the two users from the Mixed group (the participants M1 and M2 oscillated between the real experimenter and the avatar from the beginning to the end of the experiment (Figure 3.6). This result was corroborated by the fact that the collaborator choice was not significantly impacted by delocalization angle or by the instruction type. However, inter-group comparisons showed that participants' familiarization with video games had a significant impact on their collaborator choice. Participants who chose the avatar were those who played more video games. The results of the subjective questionnaire also indicated that

participants in the Real group had problems understanding the instructions compared to those in the Virtual group.

Regarding the impact of different independent variables on task performance, we first assumed that the spatial proximity of the collaborator and the avatar should provide stronger visual conflicts and thus would take the participants longer time to accomplish the task. Actually, this hypothesis (H2) was also rejected. There were no significant results confirming that small angle were more disturbing than larger ones. The distance between the real experimenter and her avatar did not seem to have an influence on participants' decision making. This may be the result of the co-effect of visual (Figure 2.13) and visual-auditive conflicts (Figure 2.14). Larger delocalization angle may cause weaker visual conflicts, but the visual-auditive inconsistency will become stronger. Further studies are needed to separate these two factors (e.g., using spatialized 3D audio to map the audio source to the avatar) and to investigate their individual influence on task performance.

Then for the comparison between different instruction modalities, overall participants had the best performance with verbal-only instruction (M = 1384, SD = 1157), and the weakest performance with gesture-only instruction (M = 2113, SD = 1755), while the multimodal instruction was medium (M = 1627, SD = 1306). This result confirmed our hypothesis on the influence of instruction modality on task performance (H3) that participants should be more efficient with verbal-only instructions than with instructions that involve gestural information when facing the dual-presence of the experimenter.

Unimodal commands like deictic gestures (pointing) or speech usually have an ambiguity problem (Bangerter, 2004). For example, pointing precision depends on the distance between hand and target, the density of objects in the pointing zone, etc. The efficiency of verbal command also varies according to the task and context. In our crown-picking task, these ambiguities were removed by asking participants to pick the nearest crown next to the experimenter, so the reaction time for different types of instructions depended on the time that the participants needed to receive and process the information. Studies on human reaction to visual-auditory stimuli have shown that humans have similar reaction time for auditory and visual stimuli (around 392 milliseconds) (Suied et al., 2009) or are slightly quicker with auditory (284 milliseconds) than visual (331 milliseconds) stimuli (Shelton and Kumar, 2010). To process information conveyed in the instructions, relational expressions in verbal instructions like "Go to the crown on my left" required the participants to rely on an allocentric reference frame with spatial updating abilities (Riecke et al., 2007), whereas pure pointing gestures can be handled within participants' egocentric reference frame. So normally participants should be quicker with gestural instructions than with verbal instructions because the former did not require participants to do mental rotations which may take several seconds depending on the rotation angle (Shepard and Metzler, 1971). Additionally, multimodal systems usually provide better task efficiency compared to unimodal systems in spatial domains (Oviatt, 1999). The opposite result (i.e., the user spent less time with verbal instructions than with multimodal and gestural ones) that we observed in our experiment can be explained by the fact that participants were disturbed by the visual conflict induced by the dual-presence of the experimenter and this visual conflict was enhanced by the pointing gesture.

3.7 Conclusion

As the results showed, first users had an a priori choice of collaborator (avatar or real person) and this choice did not change under different experimental conditions. Second, perceptual conflicts had an impact on users' performance in term of task completion time.

The goal of this experiment was to study user behavior and performance influenced by simultaneous real and virtual stimuli coming from an experimenter and her avatar (dual-presence) in a collaborative virtual context offered by multi-stereoscopic projection-based immersive device. In a two-user scenario, participants performed an object-picking task according to three types of instructions (verbal-only, gesture-only or multimodal instructions) given by an experimenter and her avatar which was delocalized from the real person according to a regular distribution of rotational angles (greater the angle, farther the experimenter and her avatar).

We found that most users had an a priori choice of collaborator, which stayed stable through all trials (neither the delocalization angle nor the instruction type had an influence on this choice). 33% of participants chose to interact with the real experimenter, 59% of participants worked with the avatar, while the rest remained undecided. This choice may be the result of combined effects of many factors, for example, user's experience with video games, the quality of the visual display, the reactiveness of the motion tracking system, etc.

When a user faced the dual-presence of another co-located user, it seemed that the cognitive process helped him/her to solve the perceptual conflicts and allowed him/her to follow a coherent strategy. However, it was at a cost of task performance for instructions that involved gesture information.

All these results provided us three possible research axes to design effective solutions to manage co-located collaboration in multi-user immersive systems.

First, in future experiments we can explicitly ask users to follow the virtual representation (the avatar) of the experimenter despite their perception of the dual-presence. It would be interesting to observe the reaction of users and to compare their performance with the results of this experiment, especially for users who were in the Real and Mixed group.

Second, tasks that require more complicated interactions and have longer duration between collaborators could be tested (with pairs of experimenter and participant, or pairs of participants) to further investigate the impact of certain factors (e.g., delocalization angle) on task performance.

At last, as we observed that the dual-presence brought conflicting information in the visual channel and thus reduced user performance for instructions including gestural interactions, an important issue could be to propose scene management approaches to limit the visual duality (i.e., visual perception of real user and his/her avatar). We could also add some automatic control (e.g., redirections) when users are in individual mode to prevent them from seeing each other's real body when navigating in the virtual world, as what we will discuss in the next chapter.

4

Study on User Cohabitation

Chapter Abstract

This chapter summarizes conceptual details about a navigation paradigm we designed called Altered Human Joystick. Then it presents cohabitation metrics that we proposed and several user experiments to verify the effectiveness of alterations for cohabitation control and their impacts on users' navigation performance and subjective feelings.

Contents

4.1 Introduction

To support safe and functional individual navigation in an multi-user IVE, traditional navigation metaphors should be adapted according to the number of co-located users and the configuration of the immersive system (size and shape of the screen walls, effective tracking space, etc.) to avoid cohabitation problems that we summarized in Chapter 2.

This chapter consists of two parts. The first one (section 4.2) summarizes conceptual details about a navigation paradigm we designed called Altered Human Joystick metaphor. The idea of this metaphor is to let each user navigate with normal human joystick control and the system will introduce alterations to a user's navigation velocity according to his/her spatial relationship with other users and with respect to the physical workspace, in order to keep user's safety and immersion level.

The second part (section 4.3 to 4.5) presents several user experiments that we conducted to verify the effectiveness of cohabitation control and their impacts on users' navigation performance and subjective feelings. First, we defined a list of cohabitation metrics and tested several alterations accumulatively in a collaborative case study, then three refined experiments followed to give an exhaustive analysis of different combinations of these alterations respectively in translation-only, rotation-only and translation-rotation combined navigation conditions.

4.2 Approach - Altered Human Joystick

In an immersive virtual environment, the human joystick paradigm allows virtual navigation by physical movements relative to a fixed reference frame. It can not assure safe and efficient navigation because it does not take into account the spatial limitations of the physical system. Users need to make sure themselves to stay inside a safe zone without touching any physical obstacles and to stay within usable projection areas. So we proposed some alterations of the original paradigm to free user's mind from these constraints.

4.2.1 Basic Model

The basic navigation method that we use is a reduced version (from 6DoF to 3DoF) of a head-controlled navigation technique (Figure 4.1) proposed by Bourdot and Touraine (2002). It is actually a human joystick metaphor (McMahan et al., 2012), but with an additional degree of freedom for yaw rotation. User's head position and orientation are captured in real time with an optical tracker attached to the 3D glasses. With Human Joystick metaphor users' hands can be left over for object manipulation or gesture interaction. This ratio control technique requires a user to move physically and thus involves his/her vestibular system during navigation, which could potentially reduce symptoms of cybersickness (Chen et al., 2013).

To navigate, user first needs to choose a neutral reference frame (composed of a position and an orientation (P_0, θ_0)) during calibration. Then user can move in the real workspace relatively to this reference frame to control the linear and angular velocity of the vehicle $(\overrightarrow{v_{veh}}, \Omega_{veh})$, denoting (P, θ) the current configuration. The translation and rotation transfer functions are

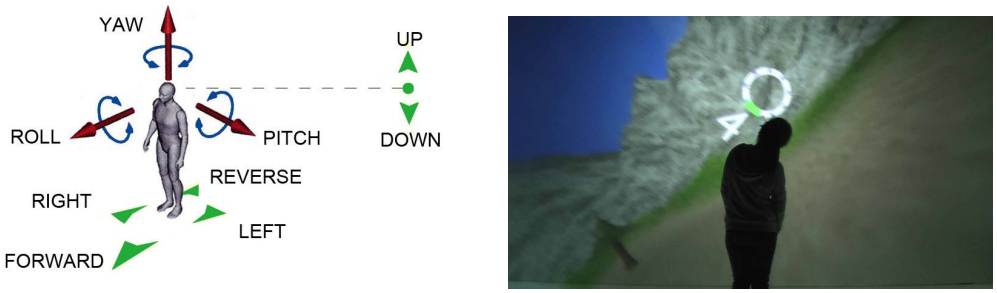


Figure 4.1 – Head-controlled navigation paradigm.

configured respectively by coefficients K_T and K_R.

$$\overrightarrow{v_{veh}} = K_T \cdot \overrightarrow{P_0 P} \tag{4.1}$$

$$\overrightarrow{\Omega_{veh}} = K_R \cdot \widehat{\theta_0 \theta} \tag{4.2}$$

4.2.2 Alterations

4.2.2.1 Altered Transfer Function

Rate control navigation metaphors are mainly characterized by their transfer functions. A transfer function transforms user's input into a change of the virtual vehicle's velocity. To restrain user's translational workspace, we can replace the linear transfer function with a divergent one integrating the distance limitation into the navigation model.

$$\overrightarrow{v_{veh}} = K_T \cdot \left(\frac{\Delta x}{X_m - \Delta x}, \frac{\Delta y}{X_m - \Delta y} \right) \tag{4.3}$$

where X_m is the minimum distance from the neutral position P_0 to the border of the physical system (Figure 4.2a). This divergent transfer function allows user to apply an infinite vehicle velocity when reaching X_m. So the workspace of a user is defined by the physical space that user is free to move inside before reaching any borders.

To avoid collision between users, we can consider each user as a moving border that restricts directly the workspace of other users. However, this method makes the velocity control too unstable when all users move the same time. A possible compromised solution is to assign each user a safe zone for proper use, and an overlapped zone to share with others. We adjust the "virtual border" of one user's workspace depending on the penetration of other users in the shared area. This way a user's velocity will only be influenced by others when he/she is inside the shared zone, one can make use of the entire shared zone as if he/she was alone in the system until the other user also needs to use it. Figure 4.3 shows an example of two users inside a 3-wall CAVE system. Usually CAVE-like systems are in a rectangular form, so the border formed by a user is chosen to be a vertical plane.

Considering rotations, physical constraints could also be integrated into the transfer function similarly as for translation. For example, in a 3-wall CAVE it is preferable to avoid user

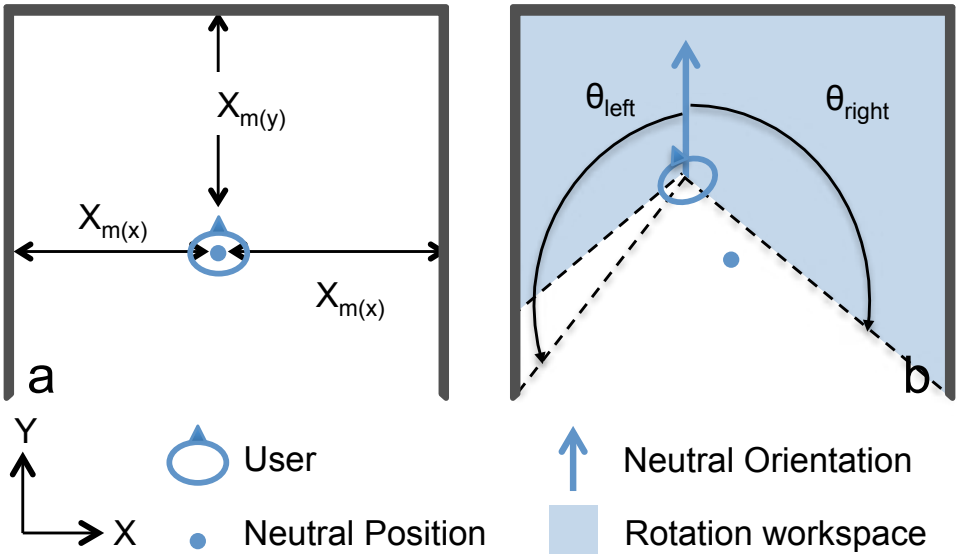


Figure 4.2 – a) The maximum available workspace X_m can have different values on x and y axis. b) User's rotational workspace is defined by θ_{max} on both sides of the neutral orientation.

occlusions and prevent users from seeing the empty screen behind them. However, due to the different nature between translation and rotation movements, we chose to use a saturated quadratic transfer function rather than a divergent form as follows:

$$\Omega_{veh} = \begin{cases} \frac{\Omega_{max}}{(\theta_{max})^2} \cdot (|\Delta\theta|)^2 & \text{if } |\Delta\theta| < \theta_{max} \\ \Omega_{max} & \text{if } |\Delta\theta| \geq \theta_{max} \end{cases} \tag{4.4}$$

where θ_{max} represents the maximum available rotational workspace (with which the user reaches the maximum rotation speed $\Omega_{max} = \pi\, rad.s^{-1}$) computed as the minimum of θ_{left}, θ_{right} and θ_{user} to provide a symmetric vehicle control around the neutral orientation θ_0 (Figure 4.2b).

4.2.2.2 Adaptive Neutral Reference Frame

The original human joystick metaphor has a fixed neutral reference frame (a neutral position and orientation) defined during calibration. This reference frame should give user the maximum workspace that the physical working environment can offer, for example, the center of a CAVE is often a good choice. However, when we have mobile obstacles in the physical working environment, e.g. multiple co-located users, it may be inappropriate to have a fixed reference frame. With the above methods of computing user's available workspace, a fixed neutral reference frame distribution will often lead to non-symmetric use of the total available workspace both for translation and rotation.

To optimize workspace usage, we can reconfigure each user's neutral reference frame as the center of his/her available workspace (Figure 4.4). In this way, each user's workspace is balanced between constraints introduced by the display system and by other users, and is always symmetric on the left and right with respect to the neutral reference frame.

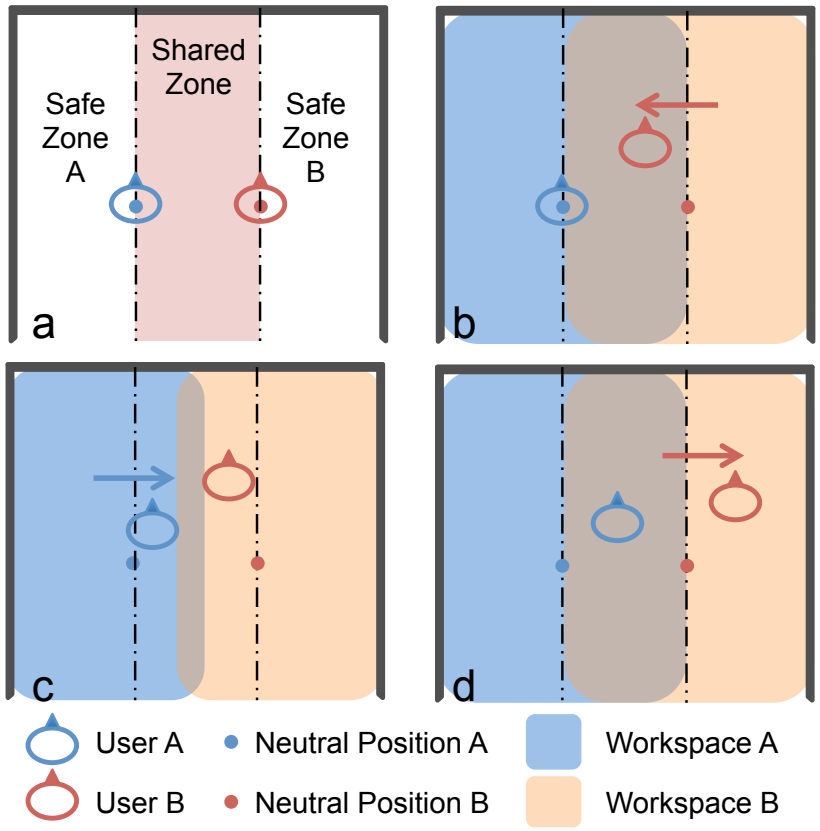


Figure 4.3 – Adaptive border of user's workspace for translation in a two-user case : a) Initial state. b) User B moves into the shared zone while user A is still in the safe zone, both users' workspaces are not influenced and user B can use the entire shared zone. c) User A also moves into the shared zone and pushes the border of user B's workspace to the right. d) User B moves back to the safe zone, the entire shared zone can now be used by user A.

Although this method could make better use of the total workspace, it also potentially increases the mutual influence between users, which may be disturbing for the navigation control. When users have relatively small workspace, the variation of neutral reference frame may remain imperceptible during navigation, which is to be tested.

4.2.3 Combinations

With the two categories of alterations presented above, we can create a set of similar navigation methods by combining these modifications. To see more clearly the impacts of each modification, we created three groups of navigation methods (see table 4.1). In the first two groups (Group T and Group R), we deactivated respectively rotation and translation, and we kept both for the last group (Group TR). For Group R, since users are not supposed to change their positions, the neutral orientations will always be fixed, though they will be different depending on how they are computed (be facing the front screen or be the bissectrice of rotational workspace).

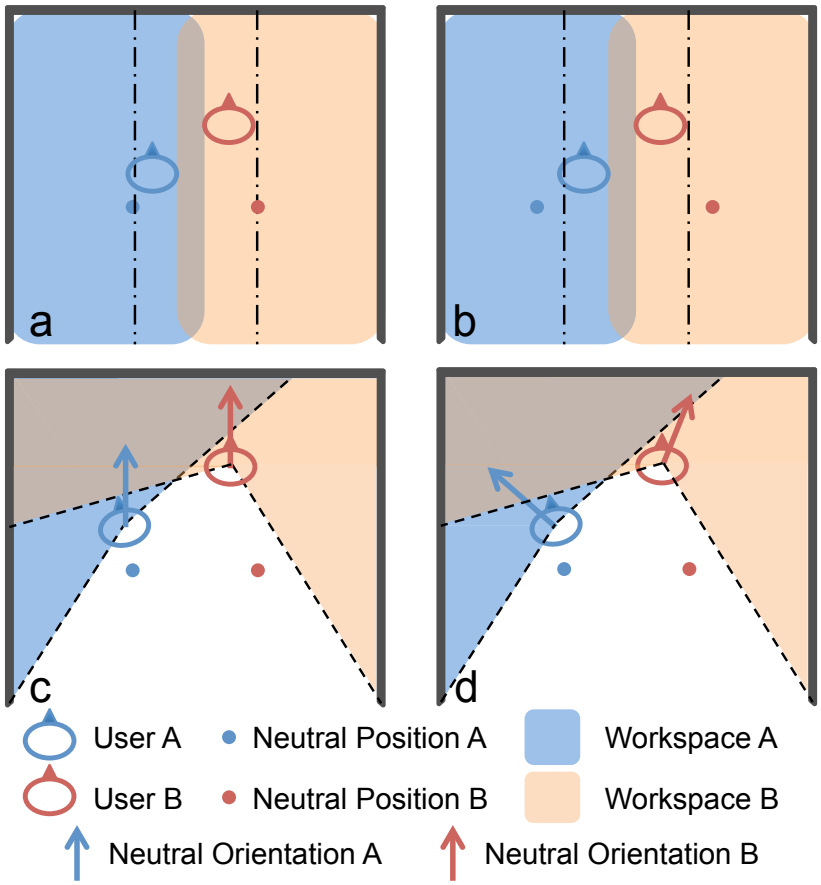


Figure 4.4 – (a, c) Users' neutral reference frames are fixed to the initial distribution. (b, d) Users' neutral reference frames are computed dynamically according to both users' positions.

4.3 Evaluation Framework

To evaluate the contributions of alterations described above, we defined several categories of metrics. Then we used them to compare these altered human joystick models by doing different navigation tasks within our CAVE platform.

4.3.1 Measures

We are interested in information coming from two different sources: how well users could achieve goals and manage navigation in the virtual world, and how they position themselves with respect to the screens and other users in the physical workspace. Here we summarized three categories of measurements that we used for evaluation.

4.3.1.1 Navigation Performance Metrics

For a pure navigation task, measurements like task finishing time (or speed) and trajectory length are often used to evaluate user's navigation efficiency.

Table 4.1 – A List of Altered Versions of Human Joystick divided in Three Groups ("/" means undefined)

Technique	Translation		Rotation	
	Form	Neutral Position	Form	Neutral Orientation
T1	Linear	Fixed	/	/
T2	Linear	Adaptive	/	/
T3	Divergent	Fixed	/	/
T4	Divergent	Adaptive	/	/
R1	/	/	Linear	Face front screen
R2	/	/	Linear	Bissectrice
R3	/	/	Saturated Quadratic	Face front screen
R4	/	/	Saturated Quadratic	Bissectrice
TR1	Linear	Fixed	Linear	Fixed
TR2	Linear	Fixed	Linear	Adaptive
TR3	Linear	Adaptive	Linear	Fixed
TR4	Linear	Adaptive	Linear	Adaptive
TR5	Divergent	Fixed	Saturated Quadratic	Fixed
TR6	Divergent	Fixed	Saturated Quadratic	Adaptive
TR7	Divergent	Adaptive	Saturated Quadratic	Fixed
TR8	Divergent	Adaptive	Saturated Quadratic	Adaptive

4.3.1.2 User Cohabitation Capacity Metrics

User cohabitation concerns both users' spatial relationship with the limitation of physical system and disturbance between users. We propose an evaluation model containing four variables covering all these aspects to measure user cohabitation capacity in a restricted physical workspace.

When several autonomous entities move within a restricted physical workspace, whether they will have cohabitation problem depends not only on where they are, but also on the moment when they have spatial coincidence. So all these variables combine spatial and temporal dimensions, and higher scores suggest higher disturbance level.

Hazardous Area Consumption In the real world, we choose to define an "Hazardous area" including all positions closer than $D_{Hazard} = 1$ meter to any physical obstacle as shown on Figure 4.5a.

In this area, user's presence and safety may be in balance depending both on the user real speed $\overrightarrow{v_{real}}$ which may affect collision probability, and on his/her relative distance to the screens $d_{screen[i]}$. So we compute a relative score V_{hac} based on $\overrightarrow{v_{real}}$ and $d_{screen[i]}$ when the user is inside

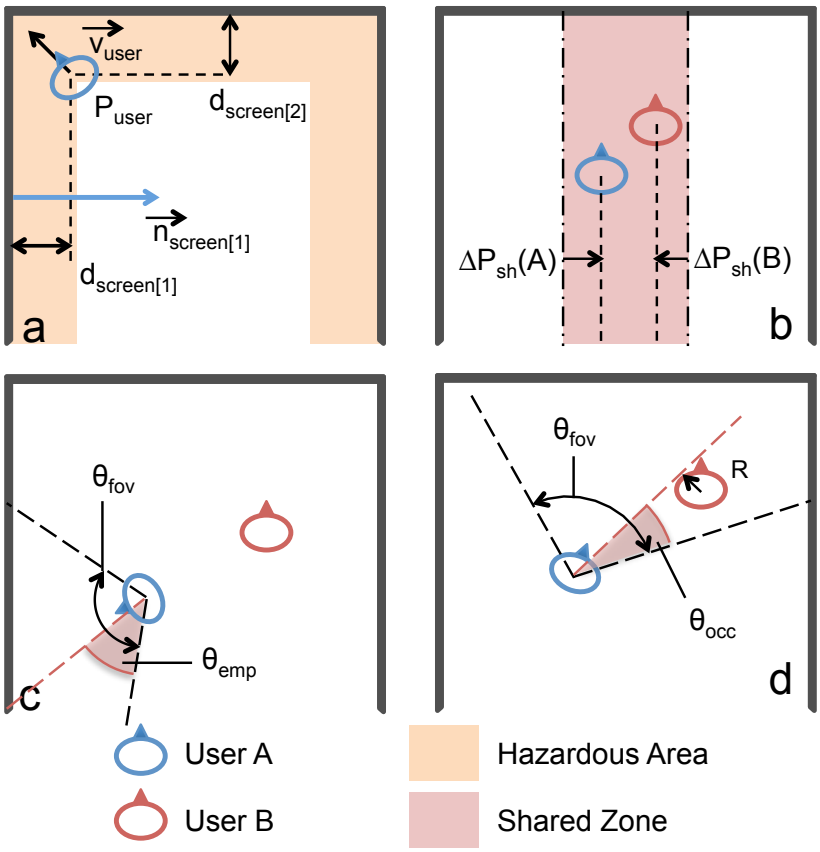


Figure 4.5 – (a) The hazardous area. (b) User's penetration $\Delta P_{sh}(t)$ in the shared zone. (c) Illustration of seeing empty screen angle $\theta_{emp}(t)$ and (d) the occlusion angle $\theta_{occ}(t)$.

the hazardous area.

$$S = \begin{cases} \overrightarrow{v_{real}} \cdot \overrightarrow{n_i} & \text{if } \overrightarrow{v_{real}} \cdot \overrightarrow{n_i} \leq 0 \\ 0 & \text{if } \overrightarrow{v_{real}} \cdot \overrightarrow{n_i} > 0 \end{cases} \tag{4.5}$$

$$V_{hac} = \sum_{t_{initial}}^{t_{final}} \left(\sum_{i=1}^{i=3} \left(\frac{S}{d_{screen[i]}} \right) \right) \tag{4.6}$$

where $\overrightarrow{n_i}$ is the screen's normal vector.

Shared Workspace Occupation In a two-user case, they share part of their translation workspace. To estimate the probability of user collision in the shared zone, we measure each user's penetration distance in the shared workspace $\Delta P_{sh}(t) = \left| (\overrightarrow{P} - \overrightarrow{P_0}).\overrightarrow{x} \right|$ while the other user is also inside the shared area (cf. Figure 4.5b), and $\Delta P_{sh}(t) = 0$ otherwise.

Using this penetration information, we can define a variable V_{Psh} allowing to evaluate users' shared area usage taking in account both amplitude and duration of each user's usage situation as:

$$V_{Psh} = \int_{t_{init}}^{t_{final}} \Delta P_{sh}(t) \cdot dt \tag{4.7}$$

Empty Screen Perception We define a variable V_{emp} to evaluate to what extent and how long a user sees the empty screen behind by using $\theta_{emp}(t)$ (Figure 4.5c)

$$V_{emp} = \int_{t_{init}}^{t_{final}} \theta_{emp}(t) \cdot dt \tag{4.8}$$

User Occlusion In order to evaluate the occlusion phenomenon, we define a penalty angle $\theta_{occ}(t)$ which quantifies the non-usable part of the field of view where the perception of other users disrupts the perception of the virtual world as shown on Figure 4.5d. This angle can be expressed, using the current user position P, the head orientation $\overrightarrow{H}$, the total available field of view θ_{fov} , the other user position P' and a bounding cylinder radius R which represented the size of the disturbing user, as:

$$\theta_{occ} = \frac{\theta_{fov}}{2} - (\widehat{\overrightarrow{H},\overrightarrow{PP'}}) + arcsin(\frac{R}{\left\|\overrightarrow{PP'}\right\|}) \text{ , with } \theta_{occ} \geq 0 \tag{4.9}$$

Using this occlusion angle, we define a variable V_{occ} which allows to evaluate the importance of the occlusion problem considering both occlusion amplitude and duration computed as:

$$V_{occ} = \int_{t_{init}}^{t_{final}} \theta_{occ}(t) \cdot dt \tag{4.10}$$

4.3.1.3 Subjective Questionnaires

Questionnaires are very useful tools to get user's subjective feelings (like presence and cybersickness) and explanations of certain choices and behaviors.

In our experiments, since we compared similar navigation conditions with minor differences, the level of presence was not our primary concern. However, these alterations could be source of cybersickness, so we used the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993) to assess user's level of cybersickness before and after each test. In addition, we developed a light-weight cohabitation questionnaire to further investigate user's feeling towards disturbances coming from physical surroundings as well as from the other user (see Appendix D).

4.3.2 Pre-test Condition

In all the experiments below we tested the cohabitation of two users. The initial distribution of both users' neutral reference frames was defined as follows: each user's neutral position had equal distance to the nearest obstacle on the left and right, and they were moved 0.6m away from the front screen to leave more space for user to move forward. The neutral orientations were chosen to be face the front screen (Figure 4.6). At the beginning of each task trial, users were always situated at the neutral position and facing the neutral orientation.

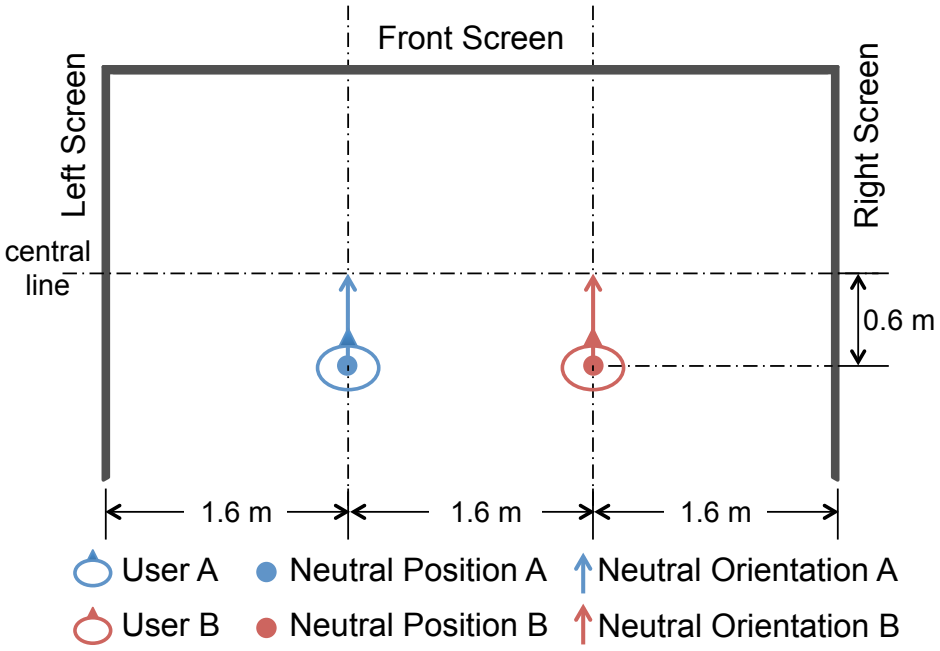


Figure 4.6 – The initial distribution of users' neutral reference frames in our CAVE platform.

Before passing real tests, participants joined a training session during which they spent about three to five minutes in a virtual playground to get familiar with the original human joystick metaphor.

4.4 Collaborative Case Study

4.4.1 Experimental Design

We conducted a first user study (Chen et al., 2015) with 36 participants divided into pairs to test modified transfer functions in a two-user working scenario. We defined three test conditions:

- **C1**: Original human joystick with linear transfer functions for translation and rotation (served as control condition).
- **C2**: Human joystick technique with modified transfer functions.
- **C3**: C2 plus adaptive neutral orientation.

Back then, the idea of adaptive neutral reference frame was not fully developed and we were not sure whether that would be beneficial, so we formed C3 in addition to C2 as a beta test of reorientation of the neutral reference frames (their positions remained fixed).

We assumed that modified transfer functions (C2) would improve users' cohabitation capacity, thus improve their navigation performance compared to linear transfer functions (C1). We also assumed that as adaptive neutral orientation provided better distribution of user's rotational workspace, users should have less occlusion with C3 than C2.

Participants were invited to accomplish an object-finding task inside a virtual factory (Figure 4.7). The factory had five similar sections each containing a common working zone and two storage areas connected by corridors. The task was for each participant to bring target objects from different storage areas to the common working zone. Since the task was relatively complicated (based on a realistic working scenario) and had a pre and post questionnaire session (for the evaluation of cybersickness), we used a between-subject design with each pair of participants testing one of the three conditions.

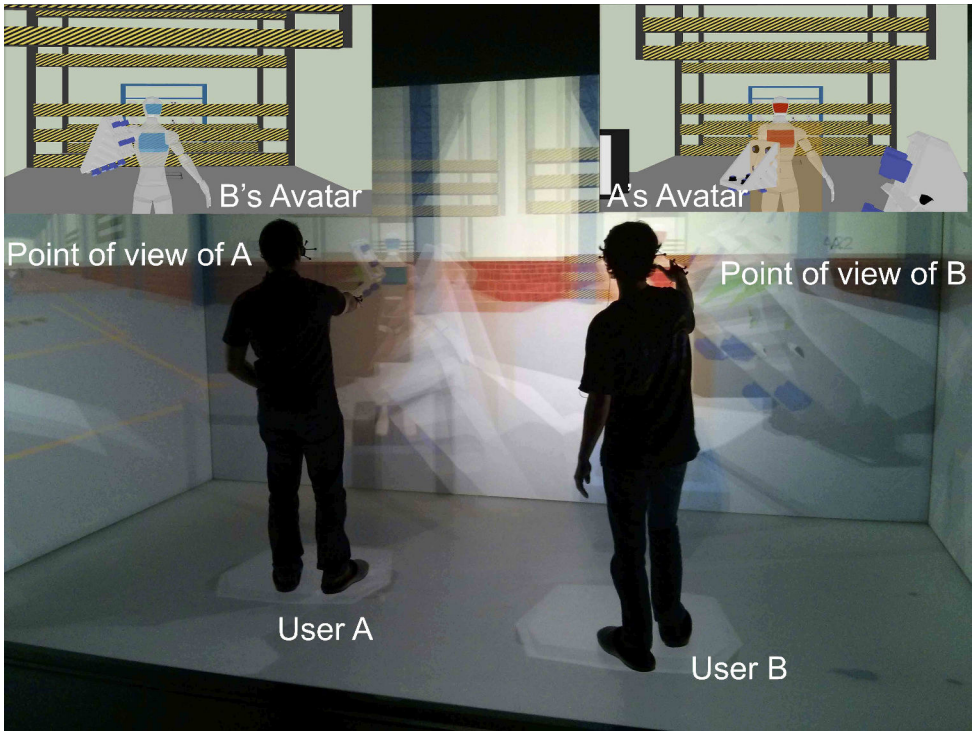


Figure 4.7 – Two users worked inside a multi-stereoscopic CAVE system to finish an object-finding task. They were virtually face-to-face to put down collected objects.

4.4.2 Results

To summarize, participants had similar navigation performance both in terms of speed and precision under different conditions. And for user cohabitation capacity, main effects were found on Shared Workspace Occupation (V_{Psh}) (H(2, N=36) = 6.109610, p=0.047) with C3 (mean=1.402, sd=0.921) less than C2 (mean=2.826, sd=1.431), and on Occlusion Angle (V_{occ}) (H(2, N=36) = 21.08269, p< 0.0001) with C3 (mean=14.7, sd=31.0) less than C1 (mean=153.4, sd=71.4) and C2 (mean=90.3, sd=84.9).

Participants under all three conditions experienced mild cybersickness symptoms (Total score: C1: mean=15.58, sd=20.41; C2: mean=18.39, sd=19.37; C3: mean=13.4, sd=10.51), however, no significant difference were found between these conditions.

4.4.3 Discussion

The lack of significant difference between C1 and C2 could be explained by several observations. First, during navigation tasks, users were not necessarily synchronized with each other

and were free to choose their navigation speed and trajectory. Regarding navigation control, they also had the choice between yaw rotation and lateral displacements, combined with forward movement to achieve the same virtual trajectory. Therefore it was difficult to predict users' real world movements and the recorded data had a large subject-dependent variation. Second, users were generally more cautious and conservative than we expected. They rarely went too close to the screen or to the other user to get high navigation speed by fear of running through virtual walls or barrels, which lessened cohabitation problems. The average navigation velocities measured under the three conditions were similar (around 2 m/s), which meant that the divergent part of C2 and C3 were seldom reached by users with the given gain values.

Finally, C3 was proved to have better support for reducing user occlusion. The way we defined the neutral orientations provided symmetric rotational workspace for users and tended to maintain an open angle between neutral orientations. In addition, we observed that C3 was also the best technique to minimize the occupation of the shared workspace (cf. V_{Psh}). Actually, as the neutral orientation defined user's main walking axis, the adaptive neutral orientations allowed more frequent use of safe zones rather than the shared zone in the middle.

Although we did not find significant results to prove the benefits of modified transfer functions as we assumed in this experiment, important lessons were learned for future study of user cohabitation in a multi-stereoscopic immersive system. First of all, for a given task inside an immersive device with fixed size, a properly chosen gain value can reduce the chance for having user cohabitation problems even with linear transfer function. In other words, to show the limitations of linear transfer functions, we should use smaller gain value. Second, it was difficult to distinguish the influence of each individual modification due to the divergence in users' navigation strategies. To tackle that, we could test translation and rotation under separate conditions (Group T and Group R) as presented in Table 4.1. At last, we found it beneficial to have adaptive neutral orientation, which led us to extend the alterations by setting the neutral position P_0 of each user dynamically to see if we can get similar benefits as performing neutral orientation change. So besides C1, C2 and C3, which corresponded respectively TR1, TR5 and TR6 (Table 4.1), we made a complete list of eight conditions to test all possible combinations.

Moreover, the alterations tested in this experiment did not seem to provoke more symptoms of cybersickness, which made it viable for us to extend the model and conduct more refined experiments.

4.5 Follow-up Experiments

We conducted three additional experiments to complete previous user study. As explained above, we globally lowered the base gain value for all transfer functions. In Experiment T (Translation only) and Experiment R (Rotation only), we created two special groups of navigation conditions by cutting off respectively rotation and translation, while in Experiment TR (Translation with Rotation) we kept both. All the three groups of conditions were listed in table 4.1 and to be tested with different types of navigation tasks.

We compared these conditions in terms of navigation performance and cohabitation capacity to see the influence of each alteration and the effect of their combinations. What's different from previous user study was that participants were equipped with closed headphones to avoid perceiving the other user by audio cues (foot steps, verbal communications, etc.) since the alterations of the navigation metaphor only provided visual feedback.

Based on the results of previous experiment, we did not measure the level of cybersickness this time because of the number and similarity of test conditions. Instead, we used a light-weight questionnaire in Experiment TR (Appendix D) to get additional information on user cohabitation.

For all the experiments below, we performed a one-way ANOVA test to compare different variables as within-subject factors. For post-hoc analyses, we used TukeyHSD multiple comparisons of means. All the analyses were performed with R[1]. The results presented in this section were considered statistically significant when $p < 0.05$. Results having a p value between 0.05 and 0.1 were also presented as a trend effect.

4.5.1 Participants

12 participants (11 male, 1 female), aged from 24 to 54 (mean=32) were involved in our user study. They were divided into six pairs and each pair passed successively Experiment T, R and TR.

4.5.2 Experiment T

4.5.2.1 Experimental Design

To compare the four navigation techniques in the first group (T1 - T4), we asked participants to accomplish a navigation task several times with each time a different technique. Thus, each navigation technique was tested once and the four conditions were randomized for different groups of participants.

The task was to travel to the other side of a straight corridor by following a given path marked by a white line on the floor and crossing all the gates along the path (Figure 4.8). Each pair of participants had parallel corridors with symmetric path to walk through so they have more chance to collide in the physical workspace. To begin, participants left their departure areas and traveled to destination zones on the other side of the corridors. Once both participants arrived, they were teleported to the departure areas to begin the next round with a different navigation technique till all the conditions were tested.

4.5.2.2 Results

We first tested the effects on two independent variables: the translation transfer function (linear or divergent) and neutral position type (fixed or adaptive). Then to see interaction effects between these two variables, we further compared the four navigation techniques.

Results showed that the transfer function type had a significant influence on multiple dependent variables while the neutral position type did not (Table 4.2). For navigation time, divergent transfer function (mean=89.5, sd=17.8) was faster than the linear one (mean=113.0, sd=21.1) ($F(1, 46)=17.25$, $p<0.001$). For Hazardous Area Consumption, divergent transfer function (mean=111.57, sd=15.17) had lower score than the linear one (mean=132.58, sd=22.22) ($F(1, 46)=14.63$, $p<0.001$). Same for Shared Workspace Occupation, divergent

[1] http://www.r-project.org/

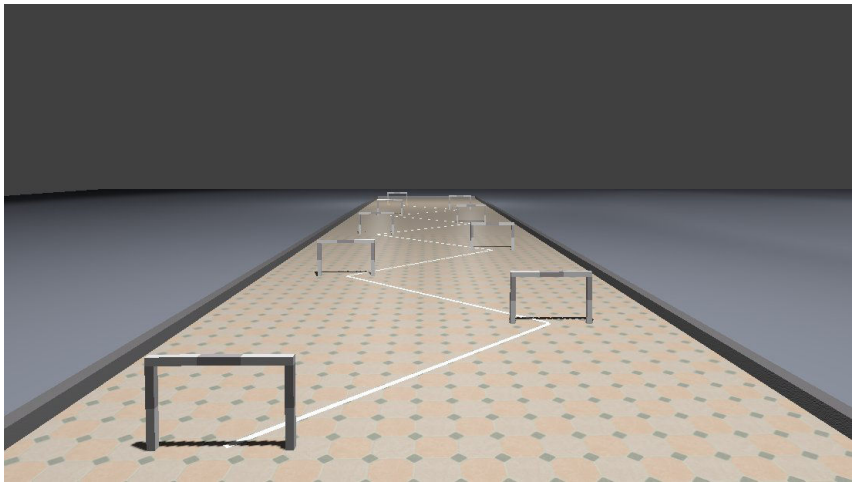


Figure 4.8 – A screen-shot of the virtual scene used in Experiment T.

Table 4.2 – Results of Experiment T (Translation only), ∅ means non significant result.

Variable	Transfer Function Type	Neutral Position Type	Technique T1 ∼ T4
Time	Divergent<Linear	∅	T1>{T3, T4}; T2>T4
Length	∅	∅	∅
Hazardous Area Consumption	Divergent<Linear	∅	T1>{T3, T4};
Shared Workspace Occupation	Divergent<Linear	∅	T1>T4; T1>T3 (0.054); T2>T4 (0.055)

transfer function (mean=5.18, sd=3.26) had lower score than the linear one (mean=12.58, sd=8.99) ($F(1, 46)=14.38$, $p<0.001$).

4.5.3 Experiment R

4.5.3.1 Experimental Design

We compared the four navigation techniques in the second group (R1 - R4) which only allowed rotations. We asked participants to turn each time virtually 180 degrees to perform an object-picking task. Each navigation technique was tested once and the four conditions were randomized for different groups of participants.

The task was to pick up an object behind a user and put it into the orange cube on the table in front (Figure 4.9). Participants were asked to stay on the neutral position and to keep their feet fixed on the floor since we did not want to involve translation. As a consequence, in this experiment we could only test a specific value of adaptive neutral orientation (the bissectrice of rotational workspace) compared to an empirically chosen direction (facing front screen).

We used green arrows to indicate the right rotation direction for the participants so overall they had equal left and right rotations.

4.5.3.2 Results

We first tested the effects on two independent variables: the rotation transfer function (linear or saturated quadratic) and neutral orientation type (face front screen or bissectrice). Then to see interaction effects between these two variables, we further compared the four navigation techniques.

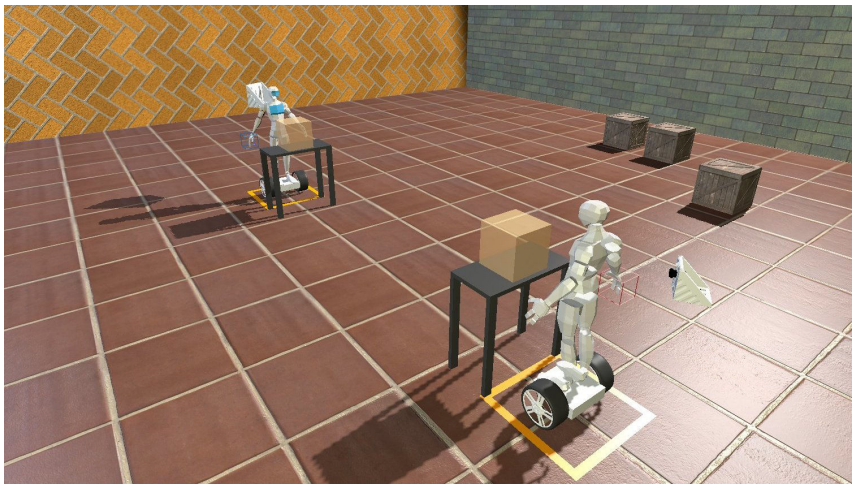


Figure 4.9 – A screen-shot of the virtual scene used in Experiment R.

Table 4.3 – Results of Experiment R (Rotation only), ∅ means non significant result.

Variable	Transfer Function Type	Neutral Orientation Type	Technique R1 ∼ R4
Time	Saturated quadratic<Linear	∅	R2>{R3, R4}
Empty Screen Perception	Saturated quadratic<Linear	∅	{R1, R2}>{R3, R4}
User Occlusion	Saturated quadratic<Linear	∅	R1>R3 (0.076), R1>R4 (0.053); R2>{R3, R4}

Results showed that the transfer function type had a significant influence on multiple dependent variables while the neutral orientation type did not (Table 4.3). For task finish time, saturated quadratic transfer function (mean=13.6, sd=5.0) was faster than the linear one (mean=27.6, sd=15.9) ($F(1, 46)=17.09$, $p<0.001$). For Empty Screen Perception, saturated quadratic transfer function (mean=59.57, sd=34.61) had lower score than the linear one (mean=196.21, sd=92.06) ($F(1, 46)=46.32$, $p<0.001$). For User Occlusion, saturated quadratic transfer function (mean=66.69, sd=51.69) also had lower score than the linear one (mean=264.82, sd=219.62) ($F(1, 46)=18.51$, $p<0.001$).

4.5.4 Experiment TR

4.5.4.1 Experimental Design

This last experiment aimed to evaluate the 3DoF version human joystick with full translation and rotation control. We compared the eight navigation techniques in the third group (TR1 -

TR8), we asked participants to perform a navigation task several times with each time a different technique. Each navigation technique was tested once and the eight conditions were randomized for different groups of participants.

Similar to Experiment T, the task was to travel to the other end of an "S" shaped corridor by following a given path marked by a white line on the floor and crossing all the gates along the path (Figure 4.10). Each pair of participants had symmetric path to walk through and they were free to use translation combined with rotation.

To begin, participants left their departure areas and traveled to destination zones on the other side of the corridors. Once both participants arrived, they were transported to a separated virtual space to finish the aforementioned cohabitation questionnaire by selecting a response with a tracker on the right hand (Figure 4.11). There were six answer levels from negative to positive answers represented by six floating cubes from left to right. For example, the leftmost one meant "I do not agree at all" and the rightmost "I totally agree", etc. When both participants finished all the seven questions, they were transported again to the departure areas to begin the next round with a different navigation technique till all the conditions were tested.

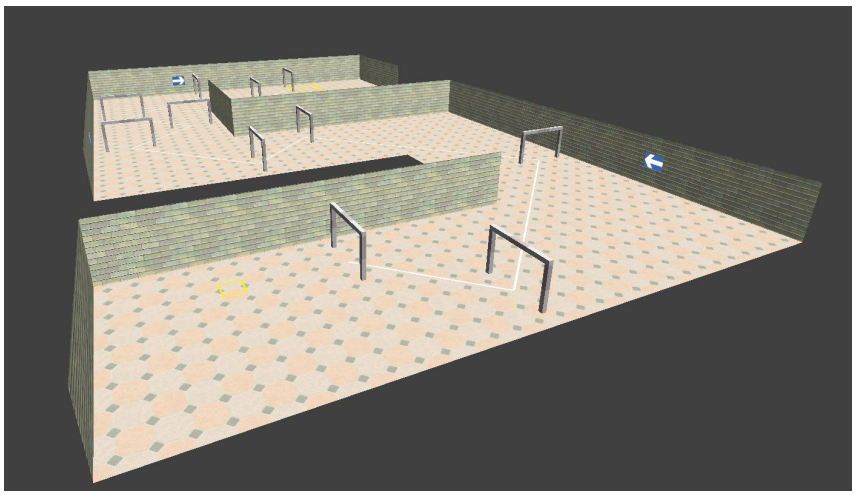


Figure 4.10 – A screen-shot of the virtual scene used in Experiment TR.

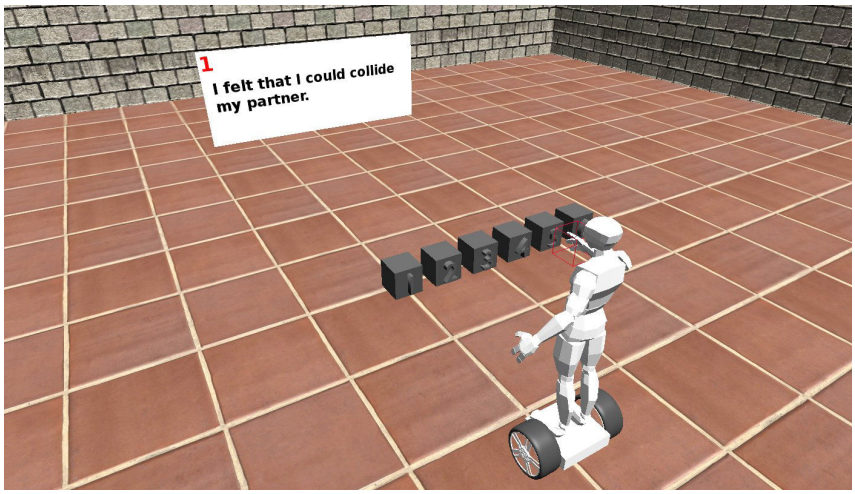


Figure 4.11 – Participants filled a questionnaire about user cohabitation presented on a virtual screen.

4.5.4.2 Results

We first tested the effects on three independent variables: the transfer function type (linear or modified), neutral position and neutral orientation types (fixed or adaptive). Then to see interaction effects between these variables, we further compared the eight navigation techniques.

Results showed that the transfer function type had a significant influence on multiple dependent variables. Modified transfer functions (divergent for translation, saturated quadratic for rotation) tended to provide better navigation efficiency and improved cohabitation capacity (Table 4.4). Figure 4.12 showed a detailed view of results on cohabitation measurements grouped by navigation techniques.

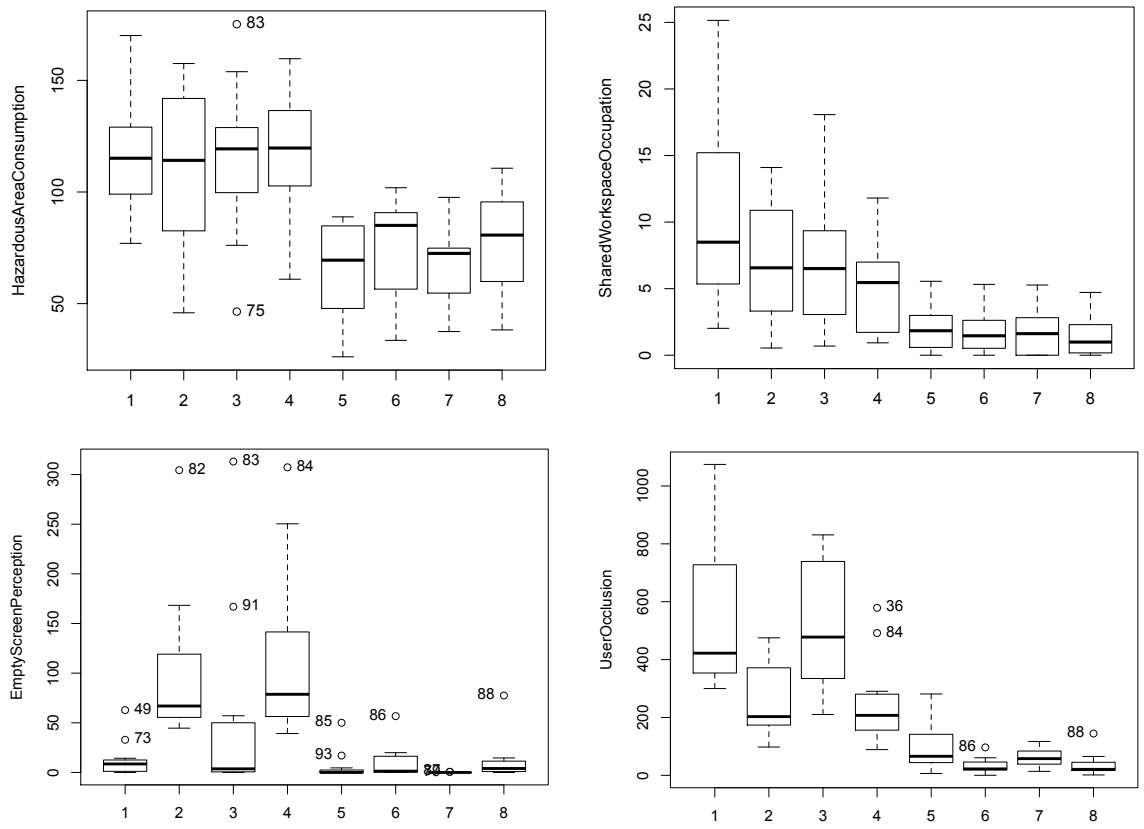


Figure 4.12 – Boxplot for conditions from TR1 to TR8 (marked as 1 to 8) on cohabitation measurements in Experiemnt TR. The numbers in the graph are outlier observations.

Modified transfer function (mean=67.1, sd=13.8) was faster than the linear one (mean=101.2, sd=14.8) ($F(1, 94)=136.6$, $p<0.001$), and had a shorter path (mean=100.51, sd=2.41) than the linear one (mean=104.21, sd=4.46) ($F(1, 94)=25.45$, $p<0.001$), knowing that the optimal path was 98 meters long (marked by white line on the floor in Figure 4.10). For Hazardous Area Consumption, modified transfer function (mean=71.59, sd=20.77) had lower score than the linear one (mean=114.86, sd=29.82) ($F(1, 94)=68.05$, $p<0.001$). For Shared Workspace Occupation, modified transfer function (mean=1.71, sd=1.57) also had lower score than the linear one (mean=7.50, sd=5.60) ($F(1, 94)=47.5$, $p<0.001$). Concerning rotational aspects, for Empty Screen Perception, modified transfer function (mean=6.83, sd=15.53) had lower score than the linear one (mean=68.78, sd=83.08) ($F(1, 94)=25.79$, $p<0.001$). For User Occlusion, modified transfer function (mean=55.69, sd=52.57) also had lower score than the

Table 4.4 – Results of Experiment TR, ∅ means non significant result.

	Variable	Transfer Function Type	Neutral Position Type	Neutral Orientation Type	Technique TR1 ∼ TR8
Navigation Performance	Time	Modified<Linear	∅	∅	{TR1, TR2, TR3, TR4}>{TR5, TR6, TR7, TR8}
	Length	Modified<Linear	∅	∅	TR3>{TR5, TR6, TR7}; TR4>TR7; TR2>TR7 (0.052); TR4>TR5 (0.06)
Cohabitation Capacity	Hazardous Area Consumption	Modified<Linear	∅	∅	{TR1, TR2, TR3, TR4}>{TR5, TR6, TR7, TR8}
	Shared Workspace Occupation	Modified<Linear	Adaptive<Fixed (p=0.1)	∅	**TR1**>{**TR4**, TR5, TR6, TR7, TR8}; {TR2, TR3}>{TR5, TR6, TR7, TR8}
	Empty Screen Perception	Modified<Linear	∅	Fixed<Adaptive	**TR1**<{**TR2, TR4**}; **TR3**<**TR4** {TR2, TR4}>{TR5, TR6, TR7, TR8};
	User Occlusion	Modified<Linear	∅	Adaptive<Fixed	**TR1**>{**TR2, TR4**, TR5, TR6, TR7, TR8}; TR2>{TR6, TR7, TR8}; **TR3**>{**TR2, TR4**, TR5, TR6, TR7, TR8}; TR4>{TR6, TR7, TR8};
Questionnaire	Question 1	Modified<Linear	∅	∅	TR1>{TR5, TR6, TR7, TR8}; TR4>TR6 (0.054)
	Question 2	Modified<Linear	∅	∅	{TR1, TR4}>{TR5, TR6, TR7, TR8}
	Question 3	Modified<Linear	∅	∅	{TR1, TR3, TR4}>{TR5, TR6, TR7, TR8}
	Question 4	Modified<Linear	∅	∅	{TR1, TR2, TR3, TR4}>{TR5, TR6, TR7, TR8}
	Question 5	Modified<Linear	∅	∅	TR1>{TR5, TR6, TR7, TR8}; TR2>TR5; TR3>TR5; TR3>{TR6, TR7, TR8} (0.065); TR4>TR5 (0.065)
	Question 6	Modified<Linear	∅	∅	{TR1, TR3, TR4}>{TR5, TR6, TR7, TR8}
	Question 7	Modified<Linear	∅	∅	{TR1, TR3}>{TR5, TR6, TR7, TR8}; TR4>TR6

linear one (mean=393.58, sd=233.79) ($F(1, 94)=95.44$, $p<0.001$).

The type of neutral reference frame did not seem to influence navigation efficiency, however, we observed some effects on cohabitation capacity. For Empty Screen Perception, fixed neutral orientation (mean=17.14, sd=51.65) had lower score than the adaptive one (mean=58.46, sd=74.59) ($F(1, 94)=9.957$, $p=0.00215$). However, fixed neutral orientation (mean=304.09, sd=284.59) had higher score than the adaptive one (mean=145.18, sd=147.77) for User Occlusion ($F(1, 94)=11.79$, $p<0.001$). In addition, neutral position type tended to have an effect on Shared Workspace Occupation (trend effect): adaptive neutral position (mean=3.76, sd=3.89) had lower score than the fixed one (mean=5.45, sd=5.86) ($F(1, 94)=2.76$, $p=0.1$).

Regarding the cohabitation questionnaire, we got similar results from question 1 to question 7: modified transfer functions globally had lower scores than linear transfer functions, whereas neutral reference frame type did have significant influence on users' subjective feeling towards user cohabitation (Figure 4.13).

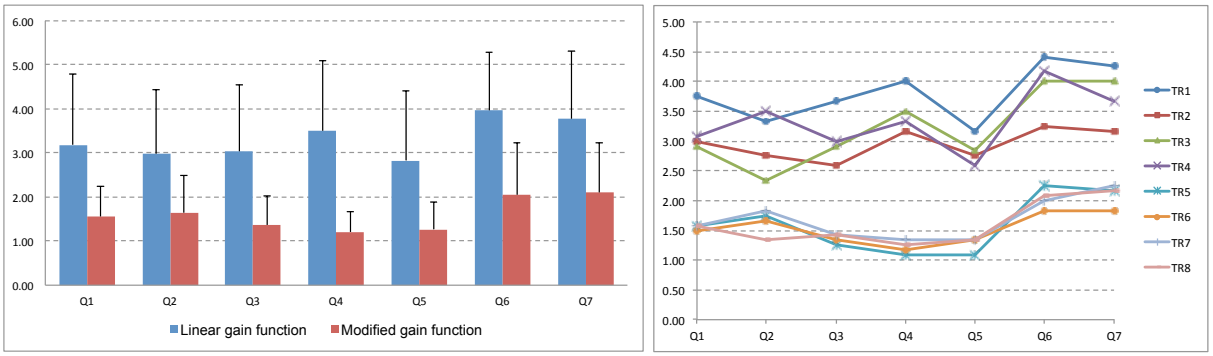


Figure 4.13 – Cohabitation questionnaire scores grouped by transfer function type (left) and by navigation condition (right).

4.5.5 Discussion

From the above three experiments, we could clearly see that modified transfer functions had a significant influence on users' navigation efficiency and cohabitation capacity, shown by both objective and subjective measurements. The divergent and quadratic transfer functions allowed users to get a much higher navigation speed before reaching the borders of the physical workspace. As discussed in previous experiment, linear transfer functions with higher gain value could also reduce cohabitation problems, but at a cost on low speed control precision, which can be observed from difference of trajectory lengths in Experiment TR. So in general, the modified transfer functions, which take the border of physical workspace into consideration for velocity computation, are more suitable than linear transfer functions to ensure safe navigation in restricted physical workspace with human joystick metaphor.

We could also see that in Experiment TR, adaptive neutral orientation reduced user occlusion, same as we found in previous experiment, though we were not able to confirm this effect in Experiment R (rotation only test) since users did not change their positions. However, on the other hand, adaptive neutral orientation led to more empty screen perception, which was not observed in previous experiment. In fact, as shown by Figure 4.4, adaptive neutral orientation provides a compromised solution between two kinds of constraints, the empty screen and the

other user, and it is normal that moving the neutral orientation away from the other user will result in more empty screen perception.

Overall, if we combine these two rotation-related variables into a new one representing user's perception of real world, techniques with adaptive neutral orientation (mean=62.51, sd=61.49) still had much lower total score than those having fixed neutral orientations (mean=462.36, sd=241.03) ($p < 0.001$). So using adaptive neutral orientation with modified transfer functions could globally lower the perception of physical environment.

Although we had expected similar benefits from adaptive neutral position as we got by changing neutral orientations, we did not get any effect on measurements used in Experiment T, and only got a trend effect ($p = 0.1$) on shared workspace occupation.

A closer look on the effects of adaptive neutral reference frames showed that, significant differences lay only between techniques with linear transfer functions (bold part in Table 4.4 and Figure 4.12). The modified transfer functions largely compensated cohabitation problem no matter the type of neutral orientation and position. Despite the fact that we did not find significant differences between techniques with modified transfer functions (from TR5 to TR8), we could still get some observations from Figure 4.12: first, the results of linear transfer function group (TR1 to TR4) and modified transfer function group (TR5 to TR8) seemed to have the same distribution, though the fluctuation was in a much smaller scale for modified transfer function group, which meant the adaptive neutral frames did have an impact on conditions in the modified transfer function group, but not enough to be observable; second, TR8 was not always the technique that provided best cohabitation capacity, so it might not be a good idea to change the neutral position and orientation simultaneously, or it should be done in a more intelligent way.

Regarding the questionnaire of cohabitation, we almost got the same result with objective measurements: two clusters of techniques were clearly identified - techniques with linear transfer functions nearly doubled the scores on all the questions (Figure 4.13). However, participants were not able to feel the influence of adaptive neutral reference frames, whose effects may be covered by the significant improvements brought by modified transfer functions. Another test with merely modified transfer functions should be added if we want to further compare different types of neutral reference frames with subjective measurements.

4.6 Conclusion

In this chapter we presented Altered Human Joystick metaphor, which is a combination of modifications with human joystick paradigm in order to enable individual navigation for co-located users in a multi-user IVE and to solve associated cohabitation problems. The alterations were designed to help users avoid physical obstacles and non-usable part of the display to assure users' safety and level of immersion by directly integrating limitations of the physical system and the influence of other users into the paradigm. Then we summarized an evaluation model to assess user cohabitation capability, using different variables to quantify the risk of collision and perceptual disturbance.

We conducted a series of user evaluations to investigate the influence of each alteration by testing their combinations with various navigation tasks. Results showed that modified transfer functions could provide a larger velocity range with relatively small base gain value compared

to linear transfer functions, thus could largely improve users' cohabitation situation both for translation and rotation. Adaptive neutral orientations can efficiently reduce user occlusion, more generally user's perception of physical environment. Whereas adaptive neutral positions only showed some potential benefits and needed to be further developed. For example, the optimal distribution of the neutral positions is not necessarily the median of each user's workspace, but can be more flexible depending on both users' activities in real time (velocity in physical workspace, navigation strategy, etc.).

Finally, in spite of our work to propose measurements that characterize efficiently navigation performance as well as user cohabitation, some limitations remain to be solved. For example, a user staying for a very long time in the peripheral field of view of another user may not be as disturbing as a user standing right in front of another for a few seconds even though the latter has a lower occlusion score.

Moreover, the way we define users' workspaces for translation (i.e. division of shared zone and safe zones) is system-dependent. It works well for rectangular workspaces, but not necessarily for circular or non-regular setups. The following chapter presents our efforts to propose a more generic cohabitation model with more advanced navigation control.

5

Dynamic Navigation Model

Chapter Abstract

This chapter presents a novel dynamic navigation model that we design for co-located collaboration. This model manages user cohabitation by integrating constraints from physical workspace and allows interactive navigation control by adding virtual constraints. First we explain our motivation and different components of the dynamic model at the conceptual level, then we present implementation details from data structure to code organization. At last, we discuss about different advantages brought by the model as well as possible future improvements.

Contents

5.1 Introduction

The Altered Human Joystick metaphor allows individual navigation for multiple users, however, each user is constrained in a predefined safe zone and the common workspace is shared by using a system dependent protocol (which works for CAVE-like immersive systems). Moreover, in many rate control navigation metaphors, the virtual vehicle's velocity is exclusively controlled by user's input and can not be influenced by constraints from the virtual world (e.g. collisions with objects).

To generalize the Altered Human Joystick metaphor and to allow interactive control of the virtual vehicle, we designed a dynamic navigation model named DYNAMIC (DYnamic NAvigation Model for Immersive Collaboration), which on the one hand, integrates real world information into the virtual navigation control to manage user cohabitation no matter the geometric properties of the immersive system, and on the other hand, takes into account influences coming from the virtual world and eventually from the vehicles of other users.

This chapter mainly contains four sections. The first section explains our motivation of developing this dynamic navigation model. The second part describes core concepts of this model: the acceleration-based transfer function and how we integrate constraints coming from the physical workspace and the virtual world. Then the third section presents implementation details and the last one some related discussion about contributions of this navigation model.

5.2 Motivation

The dynamic navigation model is the result of responses to two different research questions that we are working on: how can we generalize the Altered Human Joystick metaphor to extend the number of users regardless of the geometric configuration of immersive systems? And how can we make the virtual vehicle to be responsive to constraints coming from the virtual world to achieve interactions with virtual objects and other users during navigation?

This new model is dynamic because the transfer functions used in human joystick metaphors are replaced by their derivatives, so the vehicle is controlled by accelerations instead of setting directly its navigation velocity. As a consequence, the virtual vehicle is transformed from a pure reference frame to a virtual entity possessing its own kinematic state, providing a uniform interface to receive inputs from various sources. First, to integrate constraints from the physical workspace, we use a potential field method that generates accelerations for the virtual vehicle in order to influence user's navigation control. Second, virtual vehicles can now react to constraints in the virtual worlds in form of physical links. We can simulate physical interactions between a vehicle and virtual objects, or directly between different users' vehicles during navigation.

5.3 Concepts

5.3.1 Transfer Function

The original human joystick metaphor and the altered versions that we present in the previous chapter are based on transfer functions that map a user's relative position and orientation related to a neutral reference frame in the real world to the vehicle's translation and rotation velocities. In this previous model, a neutral position and orientation (p_0, q_0) need to be defined before starting navigation. The vehicle's translation and rotation velocity are expressed as:

$$\begin{cases} \overrightarrow{v_{veh}} = f(\Delta p) = f(p_u - p_0) \\ \Omega_{veh} = f(\Delta q) = f(q_0^{-1} \cdot q_u) \end{cases} \tag{5.1}$$

where $p_u(x, y, z)$ is user's current position and $q_u(q_x, q_y, q_z, q_w)$ is a quaternion representing user's orientation in the real world.

In the dynamic model, the virtual vehicle is not merely a spatial reference frame, but a virtual entity possessing its own kinematic state. The user input passes from $(\Delta p, \Delta q)$ to their derivatives:

$$\begin{cases} \overrightarrow{a_{veh}} = f(\overrightarrow{v_u}, \overrightarrow{a_u}) \\ \dot{\Omega}_{veh} = f(\omega_u, \dot{\omega}_u) \end{cases} \tag{5.2}$$

If we define a configuration c as the following:

$$c = \begin{bmatrix} p \\ q \end{bmatrix}, \ \dot{c} = \begin{bmatrix} \overrightarrow{v} \\ \omega \end{bmatrix}, \ \ddot{c} = \begin{bmatrix} \overrightarrow{a} \\ \dot{\omega} \end{bmatrix} \tag{5.3}$$

to have a uniform representation of linear and angular information, where $\dot{c}$ and $\ddot{c}$ are the first and second temporal derivatives of c. A linear version of the new transfer function can be then expressed as:

$$\ddot{c}_{veh} = K_1 \cdot \dot{c}_u + K_2 \cdot \ddot{c}_u \tag{5.4}$$

$$\begin{bmatrix} \overrightarrow{a_{veh}} \\ \dot{\omega}_{veh} \end{bmatrix} = \begin{bmatrix} K_{T1} & 0 \\ 0 & K_{R1} \end{bmatrix} \begin{bmatrix} \overrightarrow{v_u} \\ \omega_u \end{bmatrix} + \begin{bmatrix} K_{T2} & 0 \\ 0 & K_{R2} \end{bmatrix} \begin{bmatrix} a_u \\ \dot{\omega}_u \end{bmatrix} \tag{5.5}$$

Instead of setting directly the vehicle's position or velocity in the virtual world, user's physical movements "inject" accelerations to animate the virtual vehicle. This transfer function allows easy integration of physical constraints into the vehicle control. Moreover, the navigation velocity no longer depends on user's real world configuration, neither do we need a predefined neutral reference frame.

5.3.2 Physical Workspace constraints

In a multi-user immersive system, for a given user, the surrounding environment as well as other users are obstacles that should be avoided. Obstacle avoidance is a classical problem in mobile robot motion planning (Latombe, 2012). Among all existing methods to guide a robot to navigate over a field occupied by obstacles, potential field method is a commonly used reactive approach that can be easily parameterized and is suitable for real-time processes (Khatib, 1986; Hellström, 2011).

Generally, two kinds of potential fields lead to two different behaviors of the robot. An attractive field generates an action vector that leads the robot toward the goal, while a repulsive field pushes the robot away from an obstacle. Here we want to use the same principle to guide users by applying repulsive fields to all obstacles in the real world (e.g. screens, walls, other users, etc.). The goal is to keep users away from all the obstacles when they move around in the physical workspace. However, a major problem is that unlike motor-driven robots, users are autonomous actors that can not be directly influenced by the potential field. So, like the divergent transfer function described in the previous chapter (section 4.2.2.1), we modify the output of the transfer function (i.e. $\overrightarrow{a_{veh}}$) by integrating obstacles as real world constraints into the navigation control in order to influence user's input (i.e. $\overrightarrow{v_u}$ and $\overrightarrow{a_u}$).

5.3.2.1 Potential Field

To achieve the above objective, obstacles generate one potential field for a user's translational movements and another for rotations. The workspace borders are usually fixed while the users are mobile entities, so the global potential fields change dynamically according to users' positions in the physical workspace.

Translation To avoid collisions between users and to prevent them from going beyond workspace borders, each physical obstacle produces a repulsive field in which force vectors are in the direction of the surface normal, thus the distribution of potential field depends on obstacles' geometric form. For example, the potential field of a flat wall or screen generates parallel force vectors while those of a user's field (the user is simplified as a cylinder) spread outward from user's position (Figure 5.1).

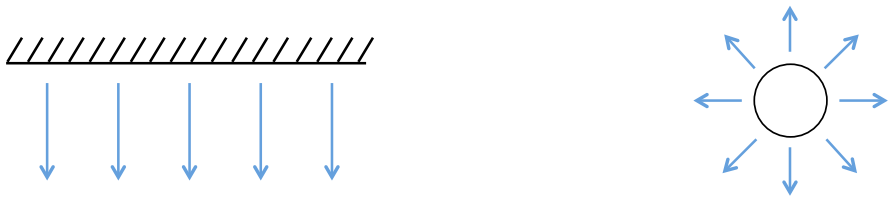


Figure 5.1 – The repulsive translational potential fields generated by obstacles.

If we name $\overrightarrow{n_i}$ as a unit surface normal vector, d_i the distance between a user and the obstacle

i, the force vector generated for the user $\overrightarrow{ft_i}$ can be expressed by a negative exponent function:

$$\overrightarrow{ft}_i = \overrightarrow{n}_i \cdot K_T \cdot \frac{1}{d_i^2} \tag{5.6}$$

where K_T is a coefficient to modulate the field power.

Rotation In projection-based systems, the potential field can help users avoid seeing empty screens and occlusions. Even with HMDs, angular potential fields are still useful if we want to influence a user's orientation in the physical workspace. Angular potential field works in a similar way as for translation, obstacles provide forces to make a user to turn towards one direction or another (clockwise or anti-clockwise). The potential fields associated with screen edges help user stay within projected area, while the potential field of a user makes other users to look away (Figure 5.2).

Figure 5.2 – The repulsive angular potential field generated by obstacles.

We can name $sign \in \{-1, 1\}$ to indicate the rotation direction (computed using user's orientation compared to the direction vector $\overrightarrow{P_u P_o}$, P_u and P_o are respectively the user's and obstacle's position). θ_i is the angle that the user needs to cover starting from the current orientation till seeing the obstacle i, then the force vector generated for the user $\overrightarrow{fr_i}$ can be expressed by a negative exponent function:

$$fr_i = sign \cdot K_R \cdot \frac{1}{\theta_i^2} \tag{5.7}$$

where K_R is a coefficient to modulate the field power.

5.3.2.2 Solutions

With the above potential fields associated with each obstacle, users will naturally keep a distance with all obstacles no matter the system configuration, which means we no longer need to specify particular working zones or shared zone for each user as we did in Chapter 4. So now the question is how users can "perceive" and be influenced by the potential fields.

Optimal Configuration In a physical workspace filled with several obstacles, the global potential field for a user is the combination of potential fields of all obstacles, and each user will

have a different potential field distribution since he/she is considered as an obstacle for the other users. For each user, the superimposed potential fields result in one or several (local minimum values) special positions and orientations where the forces are neutralized. We can use optimal configuration c_{opt} to denote the combination of optimal position and orientation:

$$c_{opt} = \begin{bmatrix} p_{opt} \\ q_{opt} \end{bmatrix} \tag{5.8}$$

where

$$\sum \overrightarrow{ft_i} = [0,0] \; and \; \sum fr_i = 0 \tag{5.9}$$

Figure 5.3 shows an example inside a three-wall CAVE, the final potential field for user A is the combination of the fields of other two users and those of all physical borders (a virtual back wall is added to assure users in the range of the screen floor). Optimal position and orientation are marked by green color.

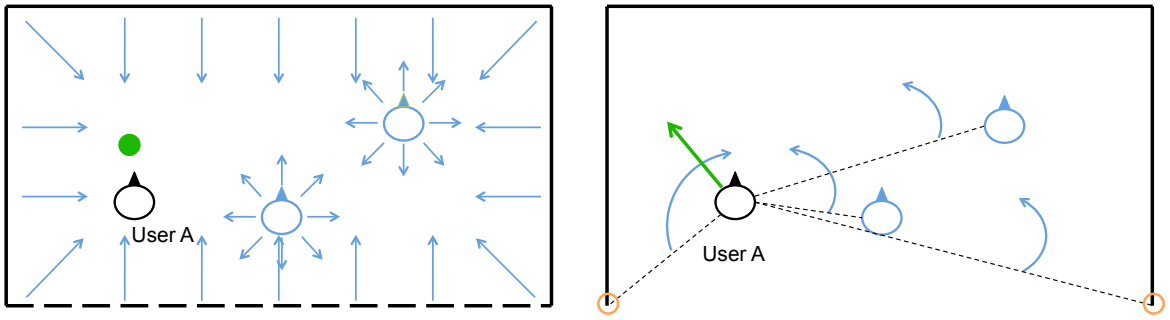


Figure 5.3 – The global translational and rotational potential field for user A in a three-wall CAVE.

To find the optimal configuration for a user in real time, we use an iterative method as presented in Algorithm 1. The idea is to create a virtual configuration $c(p,q)$ starting from user's current configuration $c_u(p_u,q_u)$ and is pushed by the sum of all potential fields step by step till we reach a certain precision ϵ_T (ϵ_R) or a given maximum round number n_T (n_R) - in case the algorithm does not converge in time. This allows us to find the optimal configuration for each user during a rendering frame.

The potential fields constantly update optimal configurations for each user, we can use the difference between user's current configuration and corresponding optimal configuration to modify user's navigation control. Here we got two options:

Option 1: Scaling We can use the difference to scale user's input. For translation, the difference can be represented by a vector $\overrightarrow{d}$ pointing from the optimal position to user's current position. By splitting user's velocity $\overrightarrow{v_u}$ (and acceleration $\overrightarrow{a_u}$) to different axis, these components can then be scaled by the corresponding components of $\overrightarrow{d}$:

$$\overrightarrow{v'_{ui}} = \overrightarrow{v_{ui}} \cdot (1 + \frac{\overrightarrow{v_{ui}} \cdot \overrightarrow{d_i}}{|\overrightarrow{v_{ui}} \cdot \overrightarrow{d_i}|} \cdot K_{ot} \cdot |\overrightarrow{d_i}|), \; i \in \{x,y\} \tag{5.10}$$

Algorithm 1 Iterative function for optimal configuration computing.

Require: $\epsilon_T \geq 0$ **and** $n_T > 0$ **and** $\epsilon_R \geq 0$ **and** $n_R > 0$

```
p ← p_u
p_opt ← p + Σ ft_i
round ← 0
while |p_opt − p| > ε_T and round < n_T do
    p ← p_opt
    p_opt ← p + Σ ft_i
    round ← round + 1
end while
q ← q_u
q_opt ← q · Σ fr_i
round ← 0
while (q_opt^{-1} · q).angle > ε_R and round < n_R do
    q ← q_opt
    q_opt ← q · Σ fr_i
    round ← round + 1
end while
return p_opt, q_opt
```

where K_{ot} is a coefficient to increase or reduce the influence of potential field. When $\overrightarrow{v_{ui}}$ is in the same direction with $\overrightarrow{d_i}$, the final input $\overrightarrow{v'_{ui}}$ will be amplified compared to $\overrightarrow{v_{ui}}$, otherwise it will be reduced (Figure 5.4a).

Similarly, for rotational control, we use θ to represent the angle between user's optimal orientation q_{opt} and current orientation q_u. The angular velocity will be amplified if the user turns away from q_{opt}, otherwise it is reduced to encourage user's turn faster towards q_{opt} (Figure 5.4b).

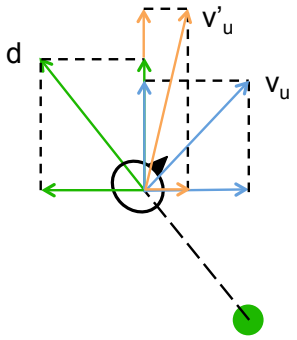
$$\omega'_u = \omega_u \cdot (1 + sign \cdot K_{or} \cdot \theta),\ sign \in \{-1, 1\} \tag{5.11}$$

where $sign$ is computed depending on the relationship between user's angular velocity and the rotation direction starting from q_{opt} to q_u. K_{or} also serves as a scale factor like K_{ot}. Both K_{ot} and K_{or} are set small enough so the potential field will not inverse the direction of user's input.

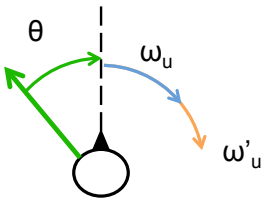
Option 2: Acceleration Insertion Instead of scaling user's input, we can also insert additional accelerations to "pull" users towards the optimal configuration. We can consider that the user and optimal configuration are connected by a mass spring damper system (a PD system in automatics which is useful to represent interconnection between two mobile entities). For translation, the total acceleration is the sum of acceleration coming from the transfer function $\overrightarrow{a_u}$ and $\overrightarrow{a_o}$ generated by the optimal position (Figure 5.5a). $\overrightarrow{a_o}$ is expressed as follows:

$$\overrightarrow{a_o} = K_{ot} \cdot (p_u - p_o) + B_{ot} \cdot (\overrightarrow{v_u} - \overrightarrow{v_o}) \tag{5.12}$$

where K_{ot} and B_{ot} are respectively spring and damper coefficients, p_o and $\overrightarrow{v_o}$ are the position and velocity of the optimal configuration.



(a) Velocity vectors are scaled by $\overrightarrow{d}$.



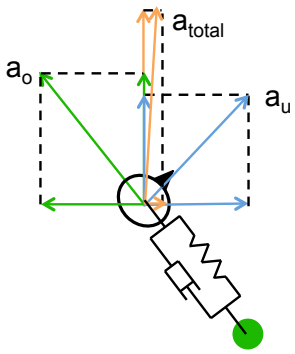
(b) Angular velocity is scaled by θ.

Figure 5.4 – Scaling of user's input by the difference between user's current configuration and the optimal configuration.

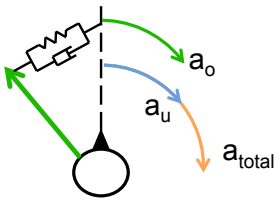
For rotation, the total angular acceleration $\dot{\omega}_{total}$ is the sum of $\dot{\omega}_u$ and $\dot{\omega}_o$ (Figure 5.5b). $\dot{\omega}_o$ is expressed as follows:

$$\dot{\omega}_o = K_{or} \cdot (q_u - q_o) + B_{or} \cdot (\omega_u - \omega_o) \tag{5.13}$$

where K_{or} and B_{or} are respectively spring and damper coefficients, q_o and ω_o are the optimal orientation and its rotational velocity.



(a) Additional linear acceleration $\overrightarrow{a_o}$.



(b) Additional angular acceleration $\dot{\omega}_o$.

Figure 5.5 – Inserting additional accelerations by the difference between user's current configuration and the optimal configuration.

Comparison These two options give users similar navigation experiences in terms of perceived acceleration as they are both based on the difference between user's current configuration and the optimal configuration decided by the potential field distribution. The major difference is that option 2 continues to accelerate the navigation as long as the user is not at the optimal configuration, while with option 1, the vehicle is accelerated only when the user intends to move physically.

The principal of option 1 is that a static user should not be "penalized" by the variation of surrounding potential fields caused by the movements of other users. This method assures the

stability of users' control of the navigation speed, however, it does not guarantee that users stay always near their optimal configurations. On the contrary, option 2 introduces accelerations to restrain users to their optimal configurations so they will hardly enter critical situations. Consequently, users are never in a stable state outside their optimal configurations.

For now we keep both options in the dynamic model because they both have advantages and drawbacks. User studies are needed to identify appropriate parameters for each option and to compare their usability in terms of navigation efficiency and comfort.

5.3.2.3 Command Configuration

Since users are mobile in the physical workspace, the potential field for each user is changing all the time depending on the positions of other users, the movements of one user change instantaneously the optimal configuration of another, which will induce instability to the navigation control, especially for option 2.

To reduce this instability, a possible solution is to define a command configuration $c_{command}$ as an intermediate layer between c_u and c_{opt}. The $c_{command}$ indicates where the user should be in the real world and tries to join c_{opt} with a certain delay. For example, we can create a mass spring damper link between $c_{command}$ and c_{opt} to reduce the influence of c_{opt} on the virtual vehicle.

Moreover, the optimal configuration defined by the potential field may not always be the preferred destination as we may have some other real world constraints to take into account. For example, a user may need to leave more workspace for another user based on their individual task requirements. So the command configuration is useful as a higher abstraction level for us to integrate other real world constraints besides obstacles' potential fields.

5.3.3 Virtual World constraints

In most rate control navigation methods, the virtual vehicle is just a mobile reference frame whose position and orientation is directly set by user commands, so it can not be influenced by physical constraints defined in the virtual world (e.g. the vehicle can cross virtual objects like walls or floors). However, the dynamic vehicle defined in DYNAMIC receives user commands in form of accelerations, so forces in the virtual world could be easily applied to the vehicle by giving the vehicle a certain virtual mass.

5.3.3.1 Interaction with Virtual World

Various physical behaviors can be simulated when using the dynamic model to navigate in the virtual world:

- Viscosity of the virtual space or friction of the floor (add an acceleration in the opposite direction of the vehicle's current velocity).
- Attraction or repulsion fields associated with a zone or a location (e.g. to stay within the boundary of the virtual scene or to show places of interest).

- Collision with certain objects (e.g. a bounding box of virtual obstacles).

These physical interactions can be managed directly by the dynamic model that includes a simple physics engine, so the virtual world does not need to be physicalized. Otherwise, we can also associate the virtual vehicle to an entity in a physicalized virtual world to leave physical constraints to a separate physics engine.

5.3.3.2 Interaction between Vehicles

When multiple users navigate in the same virtual space with the dynamic model, physical links could be established between vehicles to allow different levels of navigation control:

- Navigation with rigid link: there is a rigid physical link between pairs of vehicles, one user's move will have an instant influence on the navigation of another. This is typical for object co-manipulation - when several users move together a heavy or large object to another place.

- Navigation with soft link: there is a soft physical link between pairs of vehicles. For example, a user's vehicle is attached to the vehicle of another user by a spring so the former could be guided from one place to another while keeping a certain degree of autonomy.

- Individual navigation: no physical constraints exist between different vehicles.

Users can switch from one navigation state to another by changing the physical link connecting the vehicles (Figure 5.6). Consequently, the dynamic navigation model broadens supported collaborative scenarios in multi-user immersive virtual environment by enabling these different navigation states.

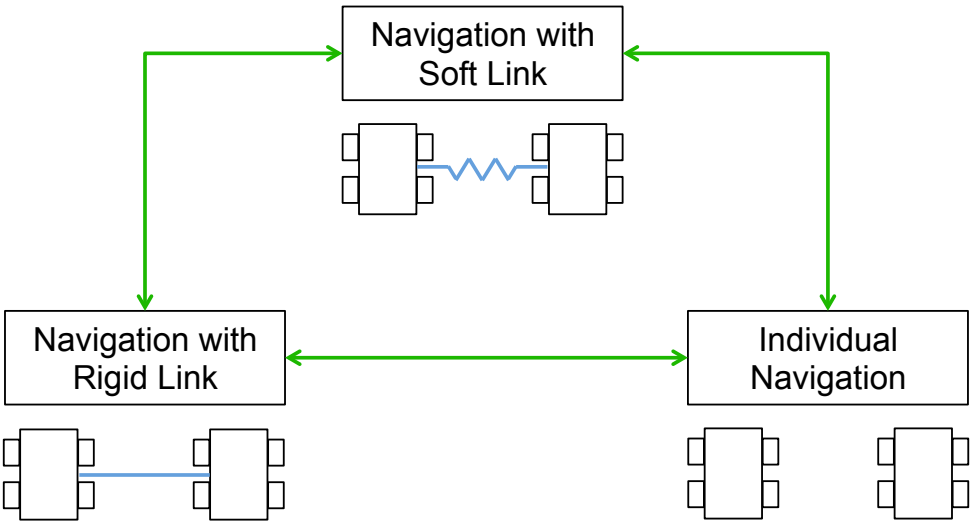


Figure 5.6 – Three levels of interactive navigation control (example of two vehicles).

5.4 Implementation

In this section, we explained how we put concepts described above together into a functional model, and how we created specific data structure to store all kinematic information of users and their vehicles. Finally, we presented an object-oriented design of the model with detailed descriptions of main components.

5.4.1 Control Process

In virtual reality applications, the general program loop contains three steps: user input collection, logic processing and rendering update. As shown in Figure 5.7, depending on user's input, the dynamic model computes with its three main components the accelerations to apply to the virtual vehicle.

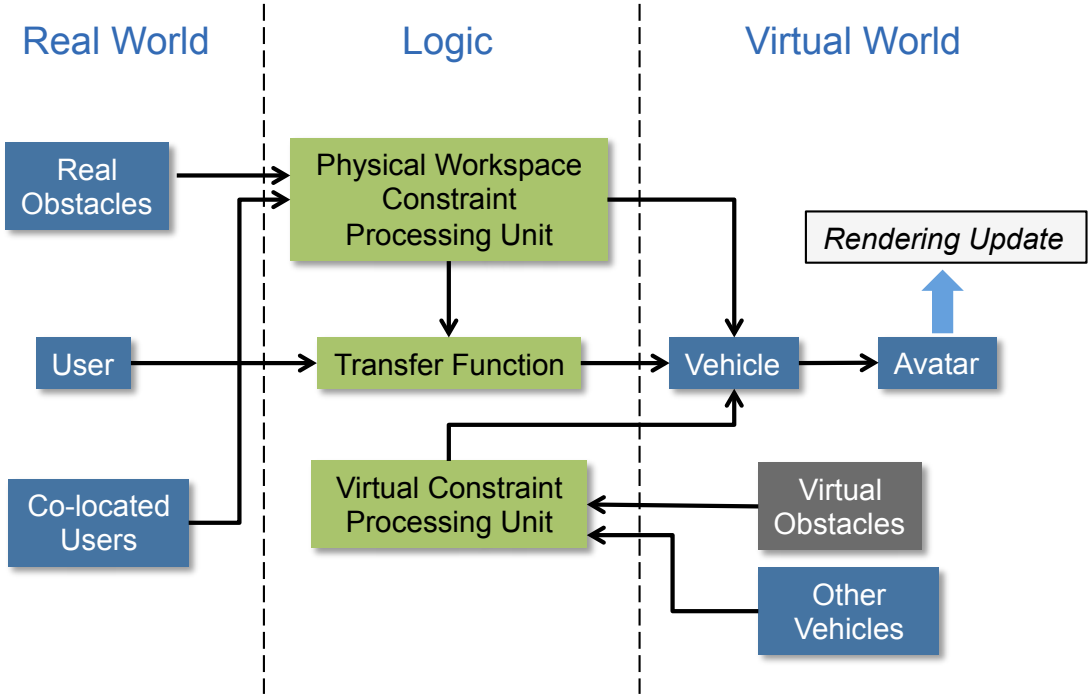


Figure 5.7 – The control process of DYNAMIC for each user.

Since we defined two options to integrate the influence of potential fields (section 5.3.2.2), there are two corresponding ways to compute the total acceleration for the vehicle. If we name acc_{user} the acceleration (both linear and angular components) generated from the transfer function, acc_{obs} from the potential field, and acc_{vir} the one from virtual constraints, the total acceleration acc_{total} for option 2 can be expressed as:

$$acc_{total} = acc_{user} + acc_{obs} + acc_{vir} \tag{5.14}$$

While for option 1, the potential field scales acc_{user} instead of inserting an additional acceleration, so we get:

$$acc_{total} = scale(acc_{user},\ (\overrightarrow{d}, \theta)) + acc_{vir} \tag{5.15}$$

where $\overrightarrow{d}$ and θ represent the difference between user's current configuration and the optimal configuration decided by the potential field distribution.

With the total acceleration acc_{total} and the vehicle's configuration c_{veh} in the previous frame, we can get a new c_{veh} to correctly position user's corresponding stage and avatar in the virtual world to get proper rendering results.

5.4.2 Data Structure

In this dynamic model, we often need different kinds of kinematic information about an object such as its velocity and acceleration, along with its static configuration to control the virtual navigation. In 3D space, an object's kinematic state contains two parts of information respectively for translation and rotation, except for points (considered as particles) that have only positional information.

So we created a variable named KS (kinematic state) to assemble all motion-related data of a given object and implement corresponding rules to update it in each frame. Table 5.1 lists different components of an object's KS.

Table 5.1 – Components of an object's KS

Components	Translation Part	Rotation Part
Configuration $c(p, q)$	$p(x, y, z)$	$q(w, x, y, z)$
Velocity $v(\overrightarrow{v}, \omega)$	$\overrightarrow{v}(v_x, v_y, v_z)$	$\omega(\omega_x, \omega_y, \omega_z)$
Acceleration $a(\overrightarrow{a}, \dot{\omega})$	$\overrightarrow{a}(a_x, a_y, a_z)$	$\dot{\omega}(\dot{\omega}_x, \dot{\omega}_y, \dot{\omega}_z)$

The translation part $(p, \overrightarrow{v}, \overrightarrow{a})$ is expressed in vectors, while we represent the rotation part by quaternion and Euler angles $(q, \omega, \dot{\omega})$. Quaternion is a convenient math tool to express spatial rotations (or orientation relative to a reference frame), but is a little complicated to get its derivatives, so instead, angular velocity and acceleration are stored using Euler angles.

An object's KS changes constantly while moving in 3D space. Typically, in the virtual world, we can animate an object by injecting a new acceleration on it to modify its current KS, which will result in a corresponding velocity and configuration change. However, the KS of a real world entity located by optical tracking (e.g. a user's body) will be directly updated by getting a new static configuration. Implementation details about KS update can be found in Appendix E.

5.4.3 Components

We implemented a first version of DYNAMIC with Python for BlenderVR (Appendix B), and it can be easily ported to other object-oriented languages. Figure 5.8 is a UML class diagram of DYNAMIC (omitting interfaces that communicate with codes managing user input and rendering).

Here we give explanations to a list of main components (in form of classes) that constitute the model:

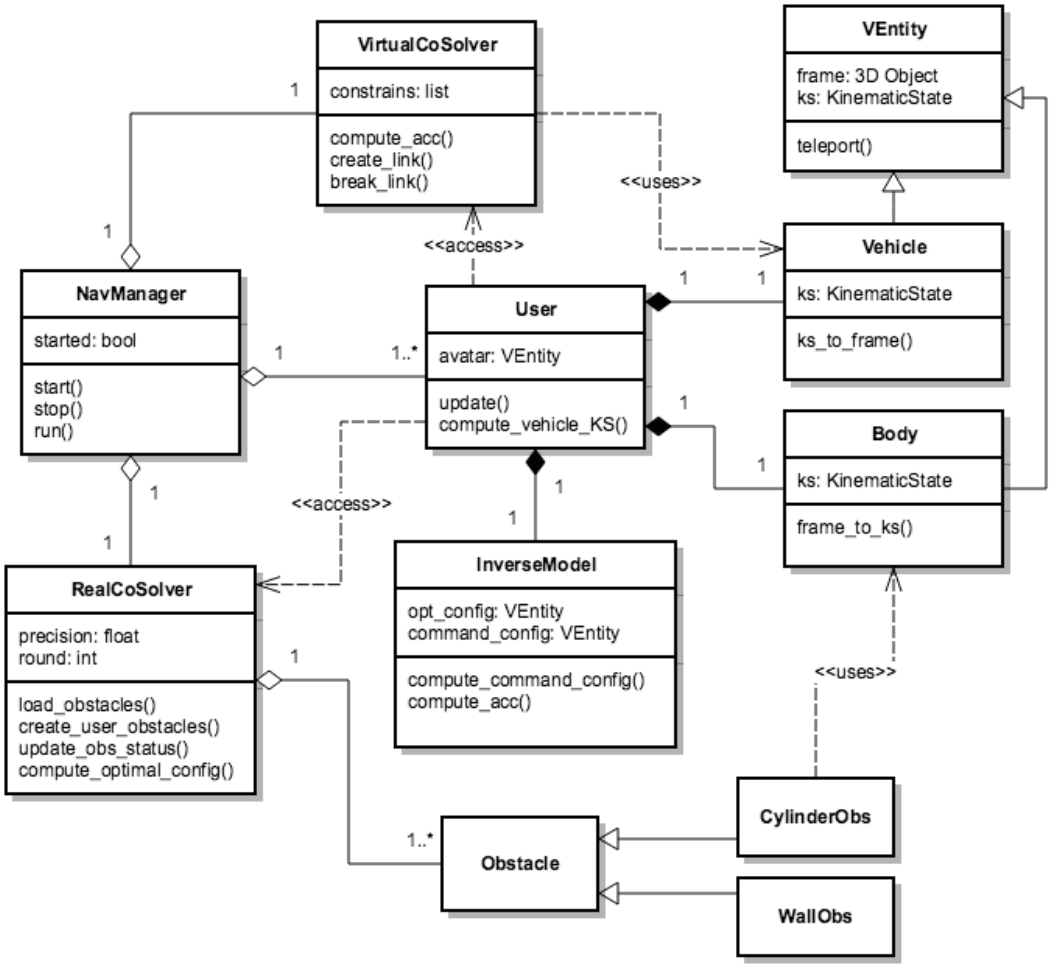


Figure 5.8 – The UML class diagram of DYNAMIC.

- *NavManager*: The main class of the navigation model that contains a list of users and constraint solvers, it initiates calls to transfer function of each user and to functions defined in these solvers. The navigation can be activated or deactivated through a control boolean.

- *VEntity*: A basic class used to represent a virtual entity involved in the simulation. Each virtual entity has its own KS, and eventually a mesh for visualization in the virtual world. For example, the *Vehicle* and *Body* (user's physical body in the real world) are all inherited classes of *VEntity*. A virtual entity's initial configuration can be set by "teleportation" without affecting other KS properties.

- *User*: A class groups all user-related information and combines accelerations from different sources to compute a new KS for the corresponding vehicle.

- *RealCoSolver*: A *Real Constraint Solver* that manages constraints from the physical workspace by computing an optimal configuration for each user.

- *InverseModel*: A unit contained in each user that manages the command configuration and implements option 1 or option 2 mentioned in section 5.3.2.2.

- *VirtualCoSolver*: A *Virtual constraint Solver* that processes a list of constrains defined in the virtual world. It creates or breaks the physical link between a pair of vehicles and computes accelerations accordingly.

5.5 Discussion

The dynamic model constitutes a framework for co-located immersive collaboration by providing a generic and extensible way to manage constraints from different sources. Here is a detailed discussion of its various advantages:

5.5.1 Generic Control Process

The dynamic model has a generic control process thanks to the potential field method, which is compatible with different types of immersive systems and various number of users.

The dynamic model can be applied to most popular immersive systems regardless of their shape, size and behavioral interface instantiation. For example, whether the visual immersion is provided by screens with retro-projection as in CAVE-like systems, or directly by HMDs, whether we have large or relatively limited workspace, whether the tracking is assured by optical cameras or accelerometers, the potential field method can always propose an optimal configuration for each user according to the boundary of physical workspace. Another advantage of potential field method is that mobile obstacles can also be included, and the global potential field is updated in real time according to the positions of mobile obstacles within the workspace boundary. This allows several (from 1 to N, N depends on the total available workspace) tracked users to share the same immersive device. And unlike with Altered Human Joystick, since users do not have specific private zones, the reallocation of workspace takes effect automatically as users join or leave the common workspace.

When there is only one user in the immersive system, the dynamic model reduces to a single user version that is similar to other navigation techniques, while the management of physical workspace and virtual world constraints remains active (Figure 5.9). Moreover, as we can manage users situated in the same physical workspace, we can also extend the dynamic model to cover multiple networked workspaces, so both co-located and remote users could collaborate in the same virtual world via a uniform navigation model. Figure 5.10 shows an example of the control process where four users from two distinct physical workspaces connected via different immersive systems collaborate in the same virtual world. Two of them (user A and B) share the same CAVE while the other two (user C and D) wear HMDs. Each user has an independent vehicle controlled by corresponding transfer function. In this situation, the *Navigation Manager* (section 5.4.3) has more than one processing units for physical workspace constraints (one for each connected physical workspace) with a single processing unit for virtual world constraints.

Besides serving multiple users, this model could also be extended to manage the coexistence of users and robotic devices during immersive collaboration. For example, the Scale 1™ (Figure 1.5c) is a robotic arm that can move with the user in control to provide haptic feedback over a large workspace. Our dynamic model can not only help other co-located users to avoid colliding with the arm (the arm is considered a mobile obstacle), but also be extended to make the arm to “be aware of” the positions of users when moving inside the shared physical

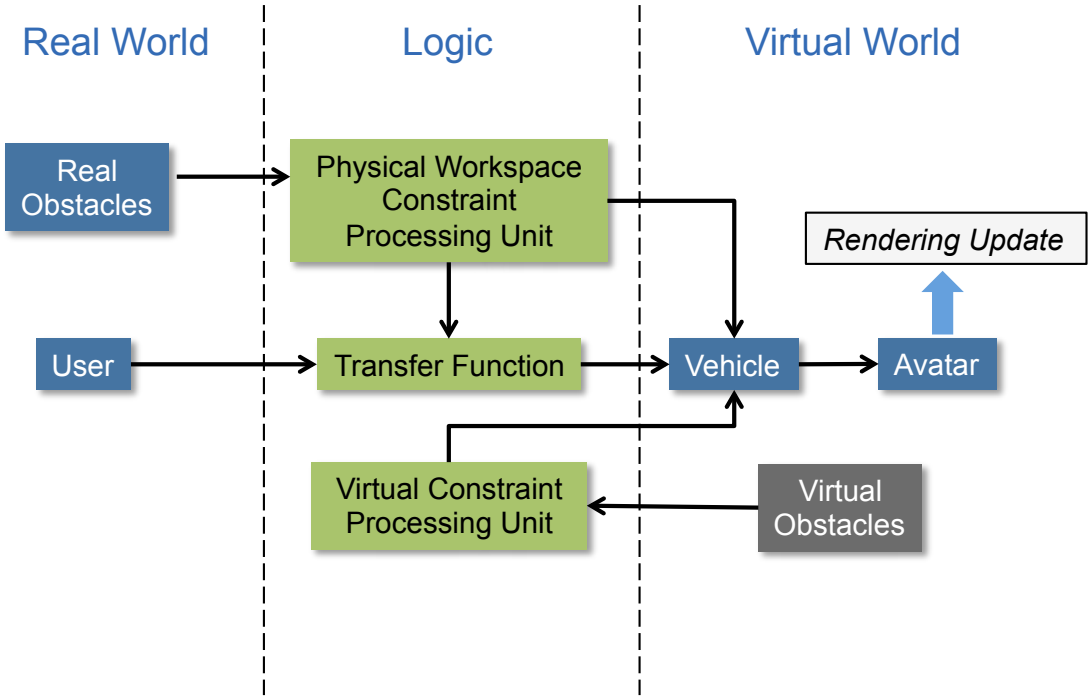


Figure 5.9 – The control process of DYNAMIC for a single user.

workspace.

5.5.2 Flexible Metaphor Integration

The transfer function used in the dynamic model is a modified version of human joystick that is based on user's movements in the physical workspace. Since the management of constraints from physical workspace and virtual world is independent with the navigation transfer function, different navigation metaphors can replace the initial human joystick metaphor. In other words, the model can be easily adapted to work with other navigation techniques like walking or steering metaphors.

For example, when several users equipped with HMDs walk inside the same room, the dynamic model can achieve the "Redirected Walking" effect by using a transfer function that maps user's position to a certain vehicle configuration $c(p, q)$ (not velocity or acceleration). With appropriate redirections, it is not only possible to restrict users in a limited workspace, but also to avoid user collisions. It is also interesting to combing the work from Kohli et al. (2005) that allows to redirect users to provide proper passive haptic feedback. We can imagine that users can be guided to meet each other in the physical workspace and in the virtual world the same time (users pass from individual mode to consistent mode).

With flexible metaphor integration, a user can switch from one navigation metaphor to another depending on the size of the virtual scene and task requirements (e.g. users walk in a virtual room and then "fly" to another place). Also, several metaphors can be used at the same time by users in collaboration to achieve each user's different task role.

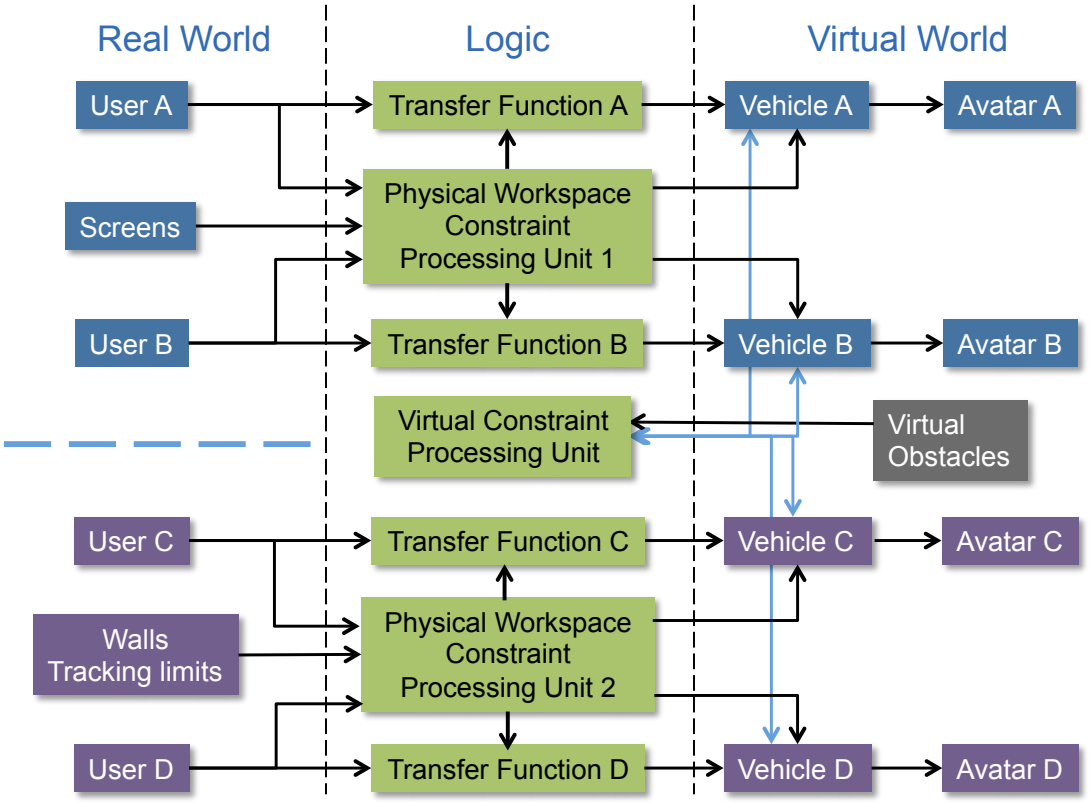


Figure 5.10 – An example of control process of DYNAMIC for co-located and remote users.

5.5.3 Mode Switching

The dynamic model provides better support for collaborative work as it enables virtual constraints on vehicles, which could be helpful for collaborative mode switching. To pass from individual mode to consistent mode, users need to reach specific positions in the physical workspace and their avatars should be seen as superimposed with the real person from other users' perspective. With the dynamic model, users can be guided to special locations in the physical workspace by imposing command configurations (section 5.3.2.3), and we can add attractive potential field to destinations in the virtual world so vehicles (under the avatars) could be well positioned in order to merge different *stages*. For example, in Figure 5.11, two users collaborate in the same physical workspace. They first concentrate on individual tasks, and then navigate to a common place of interest for *closely-coupled* collaboration (section 1.3.1). The dynamic model can make sure that they enter predefined special positions and soft links between vehicles can be added once they arrive to help them stay in consistent collaborative mode.

5.5.4 Adaptive Parameter Settings

The dynamic model contains a list of parameters as shown in equation 5.5 and from equation 5.10 to 5.13. These parameters can be configured in advance or adapted in real time according to user-related or task-related factors. Here is a list of factors that we identified:

- User-related factors:

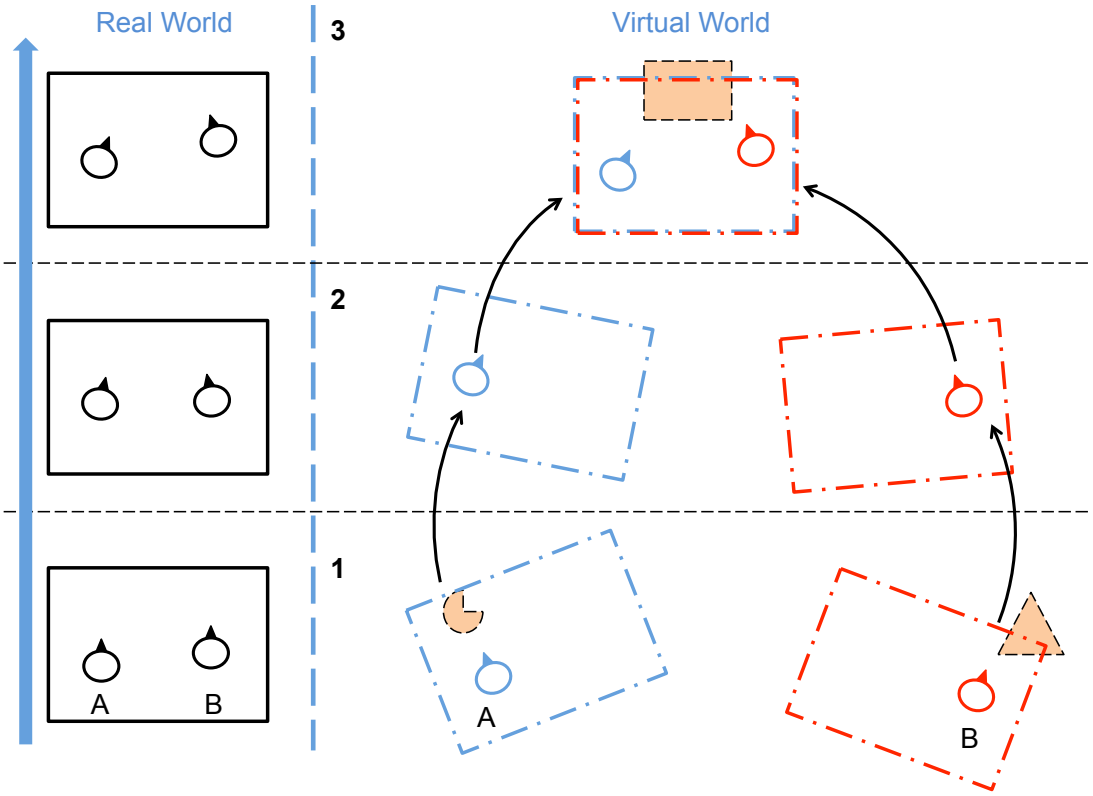


Figure 5.11 – An example of collaborative mode switching: users pass from individual mode to consistent mode (from phase 1 to 3). Users' *stages* are represented by dotted rectangles.

  - User's profile: age, sex, height & weight, game experience, familiarity to VR interfaces, etc.
  - User's state: motivation, fatigue, cybersickness, etc.

- Task-related factors:
  - Requirements of interaction devices (e.g. whether props or haptic devices are involved).
  - Requirements of the size of virtual space (i.e. whether users always stay at the same place, or they need to travel across a medium or large distance).

All these factors influence users' dynamics inside the physical workspace. Having adaptive parameters allows the dynamic model to support various use cases under heterogeneous system settings.

5.5.5 Summary

These advantages make the dynamic model a powerful conceptual framework for supporting collaborative work in immersive virtual environment, and especially, it can be adapted to work with other existing navigation metaphors with minimal efforts.

With the concepts and code structure described above, we implemented a first version of DYNAMIC with parameters that are set empirically according to the configuration (form and

size) of our immersive system. Further studies are required to find rules to model collaborative mode transitions and to define proper parameters to improve navigation control.

5.6 Conclusion

In this last chapter, we presented a novel dynamic navigation model named DYNAMIC - a generalized version of the Altered Human Joystick metaphor that we designed for co-located collaboration in multi-user immersive systems. On the one hand, this model integrates constraints from physical workspace by a potential field method to solve user cohabitation problems. On the other hand, it takes into account constraints coming from the virtual world to enable interactive control of a user's vehicle and interactions between users.

This dynamic model offers a generic solution for user cohabitation in terms of supported number of user and type of immersive systems, and it allows flexible instantiation of transfer functions as well as adaptive parameter settings. It enriches collaborative scenarios by enabling collaborative modes switching and it could also be extended to support collaboration between remote sites.

Based on the current implementation, further studies can be conducted to refine this model regarding user comfort during navigation control and to include more advanced viewpoint control to better support immersive collaboration.

Conclusion

General Conclusion

Multi-user display technology makes it possible for several users to experience visual immersion simultaneously inside the same immersive system. Under the circumstance, co-located users share the same virtual environment on top of a common physical workspace, which requires them to handle physical constraints from the real world while immersed in the virtual world.

This dissertation describes our efforts in aim to support collaborative work among co-located users in multi-user immersive virtual environments.

Depending on the relationship between users' spatial reference frames, we defined two complementary modes of collaboration - consistent and individual mode, which are both useful by supporting collaborative work in different ways. Our goal is to study issues related to individual mode in order to design a valid and efficient framework covering both collaborative modes for co-located collaboration in immersive virtual environment. We identified two categories of spatial issues related to the use of individual mode: perceptual conflicts during user communication and interaction, and cohabitation in limited physical workspace during individual navigation. Our research work is then organized around the study of these two aspects.

First, we conducted an experiment to test users' reactions to perceptual conflicts due to the introduced avatar representation for co-located collaboration. We wanted to examine how perceptual conflicts generated by the dual-presence of a user's physical body and the associated avatar would alter users' communication and their task performance. We artificially created conflicting situations and accentuated them with instructions involving spatial information. The result showed that, the dual-presence induced 33% of participants to interact with the real collaborator than her avatar, and this choice was made at the beginning and stayed stable till the end. Moreover, users' task performance was not influenced by the distance between the two forms of collaborator, and users took longer time to react to instructions containing deictic gestures than verbal commands.

So when users faced the dual-presence of collaborator, it seemed that the cognitive process helped them to solve the perceptual conflicts and allowed them to follow a coherent strategy. However, it was at a cost of task performance for instructions that involved gesture information. Although further studies are needed to get a more general conclusion, we can already say that in individual mode, it is preferable to avoid perceiving the dual-presence during collaboration, especially when user communication involves spatial information.

Second, we studied issues related to physical workspace management for co-located users. We summarized cohabitation problems along with corresponding measurement metrics in terms

of spatial relationship between users and workspace boundaries, and also workspace arrangement among co-located users. Then we concentrated on the design and evaluation of appropriate navigation paradigms to allow individual navigation while solving user cohabitation problems. The paradigm that we proposed was an altered version of Human Joystick metaphor that included additional controls to manage users' relationship with system borders and to minimize inter-user interferences.

We conducted a series of user evaluations to investigate the influence of each alteration by testing their combinations with various navigation tasks. Results showed that modified transfer functions largely improved users' cohabitation situation both for translation and rotation by providing a larger velocity range with relatively small base gain value compared to linear ones. Adaptive neutral orientations can efficiently reduce user's perception of physical environment, whereas adaptive neutral positions only showed some potential benefits and needed to be further developed.

At last, based on observations and feedback from user cohabitation study, we designed a dynamic navigation model named DYNAMIC. It is a generic version of altered human joystick, which integrates obstacles in the physical workspace by a potential field method to solve user cohabitation problems, and constraints from virtual world to enable interactive control of a user's vehicle and different levels of interaction between users. This model has several advantages: it offers a generic solution for user cohabitation in terms of supported number of user and type of immersive systems, and it allows flexible instantiation of transfer functions with adaptive parameter settings. It enriches collaborative scenarios by enabling collaborative modes switching and could also be extended to support remote collaboration.

Although our research is initially motivated by supporting co-located collaboration in large-scale multi-user immersive virtual environment, and experiments described in this thesis are conducted in a multi-user CAVE system, most of our observations and developed paradigms are applicable to other types of immersive devices (e.g. multiple HMDs), so our work contributes to co-located collaboration from a more generic perspective. Also, as the dynamic model can be extended to manage collaboration among remote sites, we make one step further towards a collaborative framework involving both remote and co-located users connected by immersive virtual environments.

Future Work

This thesis presented some basic concepts and several paradigms that we designed to support co-located immersive collaboration, while there are still many questions to be answered and new directions to be explored in the future. Moreover, by searching solutions to manage user cohabitation, we arrive at a dynamic model supporting different immersive systems with a generic control process, which also leads us to new research challenges on the way to fully support immersive collaboration.

Perceptual Conflicts

Regarding future studies on perceptual conflicts, a first complementary study that we can carry out is to explicitly ask users to follow the collaborator's avatar despite the perception of the

dual-presence to observe their reaction and to compare their performance with the results of previous experiment, especially for whose were in the Real and Mixed group.

Second, tasks that require more complicated interactions and have longer duration between collaborators could be tested to further investigate the impact of certain factors on task performance.

Additionally, the effects of visual-auditive conflicts were not yet assessed. Users may have different spatial relationship (have a large or medium distance) in the virtual world while staying side-by-side in the physical workspace, which may influence users' comprehension of the situation. So we need to test whether we should use headphones to map users' speech to avatars' positions by 3D audio, or more generally in term of design choice, whether do we need to completely separate co-located users in individual mode by putting them in a pseudo-remote situation, or the displaced audio has little effect and we can let users talk directly without computer mediation.

Collaborative Mode Switching

The dynamic model allows automatic transition between consistent mode and individual mode with predefined virtual locations for collaboration. Consequently, we need a standard procedure to define these collaboration spots according to the properties of immersive systems. For example, if users need to collaborate within a space (e.g. at two different sides of a plane) larger than the immersive system (CAVE-like system or room for HMDs), in this situation consistent mode is impossible. However, if users just need to work around a table, then we can put spots with attractive potential field around it and activate mode switching. More generally, a task-oriented collaboration model is needed so that we can divide a complex collaborative task into different phases and associate appropriate collaborative modes accordingly.

Moreover, in CAVE-like systems, mode switching is a critical moment in term of perception as users are "merging" or "splitting" with their avatars from other users' perspectives. So further study on perceptual conflicts will also help us to better understand the effects of dual-presence in this particular situation.

Navigation

The dynamic model allows adaptive parameter settings and we identified a list of potential influencing factors (section 5.5.4). The next question is how we can take into account these factors to improve the dynamic model in terms of user comfort and stability during navigation.

In term of users' comfort, an important question is that in general, how we can find appropriate parameters (e.g. gain value for the transfer function and the inverse model) for the dynamic model with all these user-related factors. So whether do we need a calibration process to identify parameters for each user, or can we find parameters that correspond to a generic model describing human reaction to accelerations? Furthermore, we need a more complete model to describe human activity in immersive virtual environment. For example, instead of using information relative to a user's head, torso or part of the limb, it would be interesting to model the motion of the whole body to better express a user's moving intention and real world dynamics.

Another important factor that we should consider for virtual navigation is user's cybersickness. In the experiments that we conducted, we used subjective questionnaire to assess the severity of cybersickness symptoms. The problem is that with the questionnaire, we can only get a description of user's feeling afterwards, it is impossible to observe the trends of cybersickness variation during the simulation to see its interaction effect with modified navigation controls. If we can get more information about the cause and influencing factors of cybersickness, it will then be possible for us to improve the dynamic model by including another constraint - targeting cybersickness minimization.

Finally, advanced viewpoint controls in immersive systems for perspective change (e.g. from first person to third person, zoom in or out, etc.) could be integrated to further facilitate the execution of collaborative tasks.

Priority Management

The dynamic model unifies constraints from the physical workspace and virtual world, which facilitates system design. However, it could be troublesome if we simply sum up all the constraints to get the final acceleration. For example, when a user is near the physical workspace border, it may happen that accelerations generated by virtual constraints could neutralize the accelerations for obstacle avoidance, which would invalidate safety measures and lead the user to a dangerous situation. Also, a user could be cornered by the sum of different constraints with inadequate personal workspace when three or more users are present in the same workspace. So it is important to assign different priorities to diverse constraints: safety constraints should come before all the others, and we need to amplify constraints that encourage the redistribution of workspace among different actors to get them out of blocking situations. In certain cases, it may be helpful to add additional feedback (visual or auditive) if the constraints fail to assure a proper distribution of users with respect to the physical workspace.

Furthermore, as discussed in the last chapter, the dynamic model could be extended to manage the coexistence of users with robotic devices (e.g. haptic arms) during immersive collaboration. We need to design proper control laws to intelligently position the arm with respect to its master and other co-located users in real time, and similarly, we should be able to modulate constraints to leave enough workspace for the arm as for the users.

Appendices

A

Acronyms

- CAVE : Cave Automatic Virtual Environment
- CoP : Center of Projection
- CSCW : Computer Supported Collaborative Work
- CVE : Collaborative Virtual Environment
- DLP : Digital Light Processing
- DoF : Degree of Freedom
- FoV : Field of View
- HCI : Human Computer Interaction
- HMD : Head Mounted Display
- IVE : Immersive Virtual Environment
- KS : Kinematic State
- LC : Liquid Cristal
- LCD : Liquid Crystal Display
- LoD : Level of Detail
- IPD : Interpupillary Distance
- IPT : Immersive Projected Technology
- SHC : Social Human Communication
- SVN : Apache Subversion
- TSI : Transformed Social Interaction
- UML : Unified Modeling Language
- VR : Virtual Reality

B

Test Platform - EVE

All experiments were carried out within EVE system (Evolutive Virtual Environment[1]) using Virtools[2] and BlenderVR (Katz et al., 2015). It is a large CAVE-like system designed to provide multi-user and multi-sensorimotor perceptions in a large immersive virtual environment ($4.8m \times 2.7m \times 4.7m$) (Figure B.1). The visual display of EVE system is composed of three screens surrounding a floor screen. Stereoscopic images are retro-projected on all these screens. Double-stereoscopy is provided on a combination of active and passive separation technologies, generating independent stereoscopic views for two users.

This CAVE is equipped with a tracking system composed of infrared cameras to track user's motion. Markers installed on the shutter stereo glasses give us the information about the (head) position and orientation of the user so that adaptive stereoscopic images are computed for each user. Additional trackers - ART-Human Body Tracking Suit[3] (Figure B.2) are used for motion capture to enable full body interaction.

[1] http://www.limsi.fr/venise/EVEsystem
[2] http://www.3dvia.com/products/3dvia-virtools/
[3] http://www.ar-tracking.com

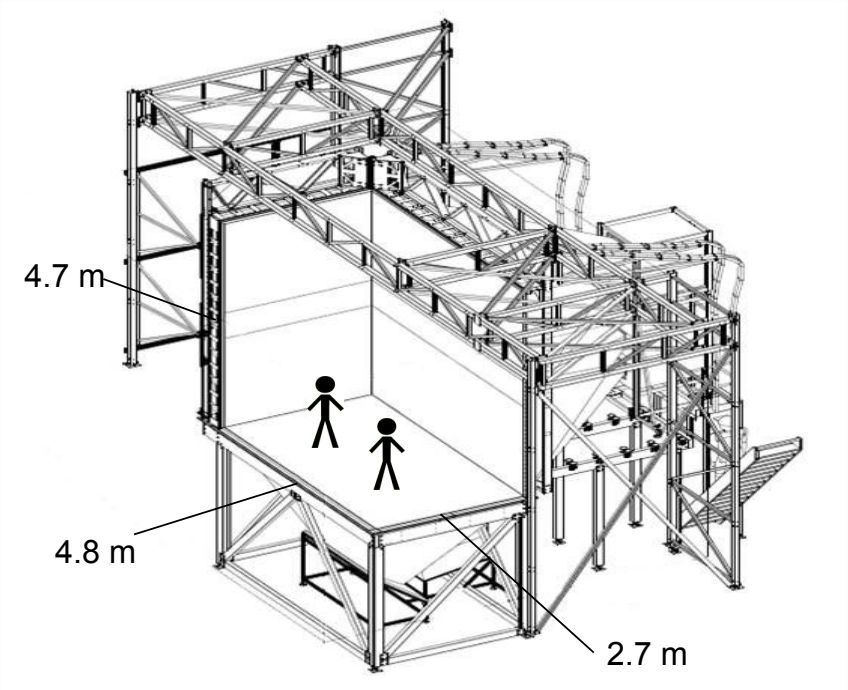


Figure B.1 – The Evolutive Virtual Environment (screen on the right-hand side is not shown in the figure).

Figure B.2 – The ART-Human Body Tracking Suit by Advanced Realtime Tracking GmbH.

C

Dual-Presence Questionnaire

The subjective questionnaire used in the experiment had a 4-point rating scale, ranging from 0 (do not agree at all), 1 (do not agree), 2 (neutral), 3 (agree), to 4 (fully agree). We regrouped the items into three categories according to three measured dimensions. Nine items measured the participant's feeling about his/her confidence on the success of collaborative task, 11 items estimated the participant's understanding of instructions, and 15 items measured participant's ease of dual-presence of the two forms of the collaborator.

1. I succeeded in the task.
2. I quickly identified the targets.
3. I was rarely mistaken.
4. I understood all the instructions.
5. It was hard to follow the instructions.
6. I easily identified the targets.
7. The instruction didn't allow me to identify the target to be reached.
8. I chose targets randomly.
9. I wasn't always sure of myself in the choice of target.
10. I had the impression that it was always the same target that was indicated by the instruction.
11. I felt confused during the experiment.
12. I preferred the instructions that were given only in verbal form.
13. Some instructions were ambiguous.

14. Some instructions didn't have a solution.
15. I forgot the physical presence of the assistant.
16. I only interacted with the virtual assistant.
17. The physical presence of the assistant disturbed me.
18. The physical presence of the assistant led me into mistakes.
19. When the instruction indicated "go to the crown on my left", I didn't understand which left was referred by the instruction.
20. When the instruction indicated "go to the crown on my right", I considered it as the right of the virtual assistant.
21. When the instruction indicated "go to the crown on my left", Sometimes I chose the left side of the virtual assistant and sometimes the left side of the real assistant.
22. The instruction "go to the crown on my left" has never been a problem to me.
23. I tried to put myself in the position of the virtual assistant to understand the instruction "go to the crown on my left".
24. I can understand the instruction better when the assistant pointed to a target with a gesture without specifying whether it was on her left or right.
25. During the experiment, I wasn't bothered by the physical presence and the virtual presence of the assistant.
26. I would prefer that the assistant who helped me in the task wasn't physically present in the display system.
27. I wasn't bothered by the physical presence of the assistant.
28. I found that the task was facilitated by the simultaneous presence of my real assistant and her avatar.
29. The proximity of the real assistant and the virtual assistant (the avatar) bothered me.
30. I prefer the situation in which the virtual assistant was far away from the real assistant.
31. When the real assistant and virtual assistant were close to each other, I had more difficulties to identify the target.
32. I always paid my attention to the instructions given by the virtual assistant regardless of her position in the virtual scene.
33. I had more trouble to rely on the virtual assistant when he/she was close to the real assistant.
34. When the real assistant and the virtual one were far from each other, it was difficult for me to decide which target to choose.
35. When the virtual and real assistants were combined (superimposed, both at the same location), I understood the instruction better.

D

Cohabitation Questionnaire

A light-weight questionnaire for evaluating user cohabitation:

1. I felt that I could collide my partner.
2. I felt that I could collide the screens.
3. I was afraid that my partner would collide me.
4. I was disturbed by my partner when I moved around.
5. My vision of the virtual scene was disrupted by my partner.
6. I was aware of the movements of my partner next to me.
7. I paid attention to my partner while performing my task.

E

Kinematic State Update

KS Update by New Acceleration

The Verlet integration is used to update the KS when receiving a new acceleration.

For translation, when we have a new $\overrightarrow{a_1}$, the corresponding position p_1 and velocity $\overrightarrow{v_1}$ are:

$$p_1 = p_0 + \overrightarrow{v_0}\Delta t + \frac{1}{2}\overrightarrow{a_0}\Delta t^2 \tag{E.1}$$

$$\overrightarrow{v_1} = \overrightarrow{v_0} + \frac{\overrightarrow{a_0} + \overrightarrow{a_1}}{2}\Delta t \tag{E.2}$$

And for rotation, giving $\dot{\omega}_1$, ω_1 and q_1 are computed as:

$$\omega_1 = \omega_0 + \frac{\dot{\omega}_0 + \dot{\omega}_1}{2}\Delta t \tag{E.3}$$

$$q_1 = q_a \cdot q_v \cdot q_0 \tag{E.4}$$

$$q_v = \left(\cos(\frac{\|\omega_0\|\Delta t}{2}),\ \sin(\frac{\|\omega_0\|\Delta t}{2})\frac{\omega_0(x)}{\|\omega_0\|\Delta t},\ \sin(\frac{\|\omega_0\|\Delta t}{2})\frac{\omega_0(y)}{\|\omega_0\|\Delta t},\ \sin(\frac{\|\omega_0\|\Delta t}{2})\frac{\omega_0(z)}{\|\omega_0\|\Delta t}\right) \tag{E.5}$$

$$q_a = \left(\cos(\frac{\|\dot{\omega}_0\|\Delta t}{2}),\ \sin(\frac{\|\dot{\omega}_0\|\Delta t}{2})\frac{\dot{\omega}_0(x)}{\|\dot{\omega}_0\|\Delta t},\ \sin(\frac{\|\dot{\omega}_0\|\Delta t}{2})\frac{\dot{\omega}_0(y)}{\|\dot{\omega}_0\|\Delta t},\ \sin(\frac{\|\dot{\omega}_0\|\Delta t}{2})\frac{\dot{\omega}_0(z)}{\|\dot{\omega}_0\|\Delta t}\right) \tag{E.6}$$

where Δt is the time lapse between two consecutive frames.

Update by New Configuration

When receiving a new configuration $c_1(p_1, q_1)$, new KS is computed as the follows:

$$\overrightarrow{v_1} = \frac{p_1 - p_0}{\Delta t} \tag{E.7}$$

$$\overrightarrow{a_1} = \frac{\overrightarrow{v_1} - \overrightarrow{v_0}}{\Delta t} \tag{E.8}$$

$$\omega_1 = \frac{Euler(q_0^{-1} \cdot q_1)}{\Delta t} \tag{E.9}$$

$$\dot{\omega}_1 = \frac{\omega_1 - \omega_0}{\Delta t} \tag{E.10}$$

F

Publications By the Author

Journal Articles

- **Weiya Chen**, Céline Clavel, Nicolas Férey, and Patrick Bourdot. Perceptual conflicts in a multi-stereoscopic immersive virtual environment: Case study on face-to-face interaction through an avatar. *Presence: Teleoperators and Virtual Environments*, 23 (4): 410–429, 2014.

Conference Proceedings

- **Weiya Chen**, Nicolas Ladeveze, Céline Clavel, Daniel Mestre, Patrick Bourdot. User co-habitation in multi-stereoscopic immersive virtual environment for individual navigation tasks. In *Proceedings of IEEE Virtual Reality*, pages 47–54, 2015.
- **Weiya Chen**, Anthony Plancoulaine, Nicolas Férey, Damien Touraine, Julien Nelson, and Patrick Bourdot. 6DoF navigation in virtual worlds: Comparison of joystick-based and head-controlled paradigms. In *Proceedings of the 19th ACM VRST*, pages 111–114, New York, NY, USA, 2013.

G

Synthèse en Français

Contexte et Motivation

L'Environnement Virtuel Collaborative (EVC) immersive se développe comme une convergence d'intérêts de recherche des communautés de la Réalité Virtuelle (VR) et du Travail Collaboratif Assisté par Ordinateur (TCAO) avec sa capacité à offrir d'immersion multi-sensorielle pour les utilisateurs en réseau. EVCs immersifs offrent un meilleur soutien pour les communications sociales et humaines, et permettent aux utilisateurs de se sentir «être ensemble» dans le même monde virtuel.

Les utilisateurs de EVC immersifs peuvent être répartis géographiquement ou situés dans le même espace de travail. Si on compare avec des situations à distance, les utilisateurs co-localisés collaborent dans le même monde virtuel au-dessus de l'espace de travail physique partagé. Cette co-localisation physique constitue un contexte mixte où les utilisateurs peuvent avoir des interactions et communications directes et celles qui sont médiatisées par ordinateur. Alors que beaucoup de travaux de recherche se concentrent sur la collaboration à distance en connectant plusieurs systèmes immersifs, la collaboration co-localisée est pour l'instant rarement étudiée en raison de la rareté des systèmes immersifs multi-utilisateurs.

Avec le développement de l'affichage numérique multi-utilisateur et la technologie de tracking, les configurations classiques basées sur la projection immersive (ex. CAVE) peuvent maintenant offrir l'immersion pour un groupe d'utilisateurs co-localisés en offrant différentes vues stéréoscopiques sans distorsion visuelle. Dans ce contexte, la coexistence de l'information du monde virtuel et réel, en particulier lorsque les utilisateurs ne partagent pas un référentiel spatial commun, offre aux utilisateurs une nouvelle expérience perceptive et cognitive.

Cette thèse de doctorat traite principalement des questions de perception et de cohabitation que nous avons identifiés dans le but d'assurer la sureté et l'efficacité pour les collaborations co-localisées dans des environnements virtuels immersifs multi-utilisateurs. Nous sommes intéressés à la façon dont les utilisateurs perçoivent et communiquent entre eux pour atteindre

un contexte de collaboration commun, et comment nous pouvons élargir des scénarios qui peuvent être pris en charge en intégrant des contrôles de point de vue plus flexibles. En outre, comme les utilisateurs partagent le même espace de travail physique limitée, il est nécessaire de trouver des solutions pour allouer assez espace de travail pour chaque utilisateur pour assurer leur niveau de sécurité et d'immersion tout en leur permettant de naviguer dans des mondes virtuels illimités.

Organisation de Thèse

Cette thèse contient deux parties principales qui sont divisées en cinq chapitres: Chapitre 1 et chapitre 2 présente le contexte général et de nos contributions à la collaboration co-localisée en termes de concepts, problèmes identifiés et les méthodes de recherche pertinents, puis Chapitre 3 à 5 décrivent en détail notre travail de recherche constitué de conception des paradigmes et des expériences de l'utilisateur.

Le chapitre 1 décrit le contexte général de cette thèse en présentant des notions et l'état de l'art de la Réalité Virtuelle (VR) et du Travail Collaboratif Assisté par Ordinateur (TCAO), et différents sujets liés à l'utilisation des environnements virtuels immersifs pour la collaboration.

Chapitre 2 se concentre sur la façon dont les utilisateurs co-localisés collaborent à l'intérieur des environnements virtuels immersifs multi-utilisateurs. Il commence par présenter la technologie d'affichage multi-utilisateur et les notions de base sur l'organisation spatiale des utilisateurs et la navigation virtuelle, puis parle de deux problèmes principaux identifiés au cours du processus de collaboration et de méthodologie de la recherche que nous avons adopté.

Le chapitre 3 décrit ensuite une étude de cas sur les conflits perceptifs causés par double-présence des utilisateurs lors de la collaboration co-localisée dans un environnement virtuel immersif multi-stéréoscopique. Nous avons examiné comment les conflits perceptifs modifient la communication de l'utilisateur et l'exécution des tâches.

Le chapitre 4 présente la mise en œuvre du Joystick Humain Altéré avec des métriques que nous avons définie pour évaluer la capacité de cohabitation en réponse à des problèmes de cohabitation identifiés. Une série d'expériences ont été menées pour tester différentes combinaisons de modifications afin de voir leurs impacts sur la distribution spatiale des utilisateurs dans l'espace de travail physique et la navigation dans le monde virtuel.

Enfin le chapitre 5 présente les concepts et les implémentations d'un nouveau modèle de navigation nommé DYNAMIC. Ce nouveau modèle gère la cohabitation des utilisateurs par l'intégration des contraintes de l'espace de travail physique et permet le contrôle de la navigation interactive en intégrant les contraintes du monde virtuel, le tout en vue de mieux soutenir la collaboration dans les environnements virtuels immersifs multi-utilisateurs.

Conclusion Générale

Technologie d'affichage multi-utilisateur permet à plusieurs utilisateurs d'avoir l'immersion visuelle simultanément à l'intérieur du même système immersif. Dans les circonstances, les utilisateurs co-localisés partagent le même environnement virtuel au-dessus de l'espace de tra-

vail physique commun, qui les oblige à traiter les contraintes physiques du monde réel tout en étant immergé dans le monde virtuel.

Cette thèse décrit nos efforts en vue de faciliter la collaboration entre les utilisateurs co-localisés en environnements virtuels immersifs multi-utilisateurs.

Selon la relation entre les référentiels spatiaux des utilisateurs, nous avons défini deux modes de collaboration complémentaires - Mode cohérent et individuel, qui sont tous utiles pour le travail collaboratif de différentes manières. Notre objectif est d'étudier les questions liées au mode individuel afin de concevoir un cadre valide et efficace couvrant les deux modes de collaboration pour la collaboration co-localisée dans un environnement virtuel immersif. Nous avons identifié deux catégories de problèmes spatiaux liées à l'utilisation du mode individuel: conflits perceptifs lors de la communication et l'interaction entre utilisateurs, et la cohabitation dans l'espace de travail physique limitée pendant la navigation individuelle. Notre travail de recherche est alors organisé autour des études de ces deux aspects.

Tout d'abord, nous avons mené une expérience pour tester des réactions des utilisateurs aux conflits perceptifs due à la représentation par des avatars introduits pour la collaboration co-localisée. Nous voulions examiner comment la perception des conflits générés par la double-présence de corps physique de l'utilisateur et l'avatar associé modifierait la communication des utilisateurs et de leur exécution de la tâche. Nous avons créé artificiellement des situations conflictuelles et les accentuées avec des instructions qui contient l'information spatiale. Le résultat a montré que, la double-présence induit 33% des participants à interagir avec la collaboratrice réelle que son avatar, et ce choix a été fait au début et est resté stable jusqu'à la fin. En outre, l'exécution de la tâche des utilisateurs n'a pas été influencée par la distance entre les deux formes de collaborateur, et les utilisateurs ont pris plus de temps à réagir aux instructions contenant les gestes déictiques que les commandes verbales.

Donc, lorsque les utilisateurs sont confrontés à la double-présence de collaborateur, il semble que le processus cognitif les a aidés à résoudre les conflits perceptifs et leur a permis de suivre une stratégie cohérente. Cependant, il était à un coût de l'exécution des tâches pour les instructions qui impliquaient des informations de geste. Bien que d'autres études sont nécessaires pour obtenir une conclusion plus générale, nous pouvons déjà dire que dans le mode individuel, il est préférable d'éviter de percevoir la double-présence lors de la collaboration, en particulier lorsque la communication de l'utilisateur implique l'information spatiale.

Deuxièmement, nous avons étudié les problèmes liés à la gestion de l'espace de travail physique pour les utilisateurs co-localisés. Nous avons fait un résumé des problèmes de cohabitation avec les mesures correspondant en termes de relation spatiale entre les utilisateurs et les limites de l'espace de travail et le partage d'espace de travail parmi les utilisateurs co-localisés. Ensuite, nous nous sommes concentrés sur la conception et l'évaluation des paradigmes de navigation appropriés pour permettre la navigation individuelle tout en résolvant les problèmes de cohabitation. Le paradigme que nous avons proposé est une version altérée de la métaphore de Joystick Humain qui inclut des contrôles supplémentaires pour gérer la relation des utilisateurs avec des frontières du système et minimiser les interférences inter-utilisateur.

Nous avons effectué une série d'évaluations des utilisateurs pour étudier l'influence de chaque modification en testant leurs combinaisons avec diverses tâches de navigation. Les résultats ont montré que les fonctions de transfert modifiées ont largement amélioré la situation de cohabitation des utilisateurs à la fois en translation et en rotation par rapport à ceux linéaires. L'orientation neutre adaptive peut réduire la perception de l'environnement physique, alors que

les positions neutres adaptatifs ont seulement montré certains avantages potentiels et avaient besoin d'être davantage développés.

Enfin, basé sur des observations et des commentaires de l'étude de cohabitation, nous avons conçu un modèle de navigation dynamique nommée DYNAMIC. C'est une version générique du Joystick Humaine Altéré, qui intègre les obstacles dans l'espace de travail physique par une méthode de champ de potentiel pour résoudre les problèmes de cohabitation, et les contraintes de monde virtuel pour permettre un contrôle interactif du véhicule et différents niveaux d'interaction entre les utilisateurs. Ce modèle présente plusieurs avantages: elle offre une solution générique pour la cohabitation des utilisateurs en termes de nombre d'utilisateurs pris en charge et le type de systèmes immersifs, et il permet l'instanciation flexible des fonctions de transfert avec les réglages de paramètres adaptatifs. Il enrichit les scénarios de collaboration en permettant les modes de collaboration commutation et pourrait également être étendu pour supporter la collaboration à distance.

Bien que notre recherche est d'abord motivée par soutenir la collaboration co-localisée en environnement virtuel immersif multi-utilisateurs à grande échelle, et les expériences décrites dans cette thèse sont menées dans un système CAVE multi-utilisateurs, la plupart de nos observations et les paradigmes développés sont applicables aux autres types de dispositifs immersifs (ex. plusieurs HMDS), de sorte que notre travail contribue à la collaboration co-localisée dans une perspective plus générale. En outre, comme le modèle dynamique peut être étendu pour gérer la collaboration entre les sites distants, nous faisons un pas de plus vers un cadre de collaboration entre utilisateurs distants et co-localisés reliés par des environnements virtuels immersifs.

List of Figures

List of Tables

Bibliography

Agrawala, M., Beers, A. C., McDowall, I., Fröhlich, B., Bolas, M., and Hanrahan, P. The two-user responsive workbench: support for collaboration through individual views of a shared space. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, pages 327–332, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. ISBN 0-89791-896-7. doi: 10.1145/258734.258875.

Aguerreche, L., Duval, T., and Lécuyer, A. 3-hand manipulation of virtual objects. In *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pages 153–156. Eurographics Association, 2009.

Aguerreche, L., Duval, T., and Lécuyer, A. Comparison of three interactive techniques for collaborative manipulation of objects in virtual reality. In *CGI 2010 (Computer Graphics International)*, 2010.

Argelaguet, F., Kunert, A., Kulik, A., and Froehlich, B. Improving co-located collaboration with show-through techniques. In *3D User Interfaces (3DUI), 2010 IEEE Symposium on*, pages 55–62, 2010. doi: 10.1109/3DUI.2010.5444719.

Arthur, K., Preston, T., Taylor, R., Brooks, F., Whitton, M., and Wright, W. Designing and building the pit: a head-tracked stereo workspace for two users. In *2nd International Immersive Projection Technology Workshop*, pages 11–12, 1998.

Bailenson, J. N. and Yee, N. A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 15(6):699–716, 2006.

Bailenson, J. N., Beall, A. C., and Blascovich, J. Gaze and task performance in shared virtual environments. *The journal of visualization and computer animation*, 13(5):313–320, 2002.

Bailenson, J. N., Swinth, K. R., Hoyt, C. L., Persky, S., Dimov, A., and Blascovich, J. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence*, 14(4):379–393, 2005.

Bangerter, A. Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, 15(6):415–419, 2004.

Banks, M. S., Held, R. T., and Girshick, A. R. Perception of 3-d layout in stereo displays. *Information display*, 25(1):12, 2009.

Bayon, V., Griffiths, G., and Wilson, J. R. Multiple decoupled interaction: An interaction design approach for groupware interaction in co-located virtual environments. *International journal of human-computer studies*, 64(3):192–206, 2006.

Beaudouin-Lafon, M. *Computer Supported Co-operative Work*, volume 7 of *Trends in Software*. John Wiley & Sons, 1999.

Beck, S., Kunert, A., Kulik, A., and Froehlich, B. Immersive group-to-group telepresence. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):616–625, 2013.

Becker, B. and Mark, G. Social conventions in collaborative virtual environments. In *Proceedings of Collaborative Virtual Environments*, pages 17–19. Citeseer, 1998.

Begault, D. R. et al. *3-D sound for virtual reality and multimedia*, volume 955. Citeseer, 1994.

Benford, S., Bowers, J., Fahlén, L. E., and Greenhalgh, C. Managing mutual awareness in collaborative virtual environments. In *Proceedings of the conference on Virtual reality software and technology*, pages 223–236. World Scientific, 1994.

Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., and Snowdon, D. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 242–249, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. ISBN 0-201-84705-1. doi: 10.1145/223904.223935.

Benford, S., Greenhalgh, C., Rodden, T., and Pycock, J. Collaborative virtual environments. *Commun. ACM*, 44(7):79–85, july 2001. ISSN 0001-0782. doi: 10.1145/379300.379322.

Bimber, O., Fröhlich, B., Schmalstieg, D., and Encarnação, L. M. The virtual showcase. In *ACM SIGGRAPH 2006 Courses*, page 9. ACM, 2006.

Biocca, F. Will simulation sickness slow down the diffusion of virtual environment technology? *Presence: Teleoperators and Virtual Environments*, 1(3):334–343, 1992.

Blake, R. and Sekuler, R. *Perception*. McGraw-Hill higher education. McGraw-Hill Companies, 2006. ISBN 9780072887600.

Blom, K., Lindahl, G., and Cruz-Neira, C. Multiple active viewers in projection-based immersive environments. *Immersive Projection TechnologyWorkshop*, 2002.

Boker, S. M., Cohn, J. F., Theobald, B.-J., Matthews, I., Brick, T. R., and Spies, J. R. Effects of damping head movement and facial expression in dyadic conversation using real–time facial expression tracking and synthesized avatars. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3485–3495, 2009.

Bolas, M., McDowall, I., and Corr, D. New research and explorations into multiuser immersive display systems. *Computer Graphics and Applications, IEEE*, 24(1):18–21, 2004.

Bolt, R. A. *"Put-that-there": Voice and gesture at the graphics interface*, volume 14. ACM, 1980.

Bourdot, P. and Touraine, D. Polyvalent display framework to control virtual navigations by 6DoF tracking. In *Virtual Reality, 2002. Proceedings. IEEE*, pages 277–278, 2002. doi: 10.1109/VR.2002.996537.

Bowman, D. A., Gabbard, J. L., and Hix, D. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, 11(4):404–424, 2002.

Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004. ISBN 0201758679.

Brown, B. and Bell, M. Cscw at play:'there'as a collaborative virtual environment. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 350–359. ACM, 2004.

Bryson, S. Virtual reality in scientific visualization. *Communications of the ACM*, 39(5):62–71, 1996.

Bullinger, H., Riedel, O., and Breining, R. Immersive projection technology - benefits for the industry. In *International Immersive Projection Technology Workshop*, pages 13–25, 1997.

Burgoon, M., Hunsaker, F. G., and Dawson, E. J. *Human communication*. Sage Beverly Hills, CA, 1994.

Burton, M., Pollock, B., Kelly, J. W., Gilbert, S., Winer, E., and de la Cruz, J. Diagnosing perceptual distortion present in group stereoscopic viewing. In *IS&T/SPIE Electronic Imaging*, pages 82910P–82910P. International Society for Optics and Photonics, 2012.

Bystrom, K.-E., Barfield, W., and Hendrix, C. A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 8(2):241–244, 1999.

Capin, T., Jovovic, M., Esmerado, J., Aubel, A., and Thalmann, D. Efficient network transmission of virtual human bodies. In *Computer Animation 98. Proceedings*, pages 41–48. IEEE, 1998.

Carlsson, C. and Hagsand, O. Dive—a platform for multi-user virtual environments. *Computers & graphics*, 17(6):663–669, 1993.

Carstensen, P. H. and Schmidt, K. Computer supported cooperative work: New challenges to systems design. In *In K. Itoh (Ed.), Handbook of Human Factors*. Citeseer, 1999.

Chen, W., Plancoulaine, A., Férey, N., Touraine, D., Nelson, J., and Bourdot, P. 6DoF navigation in virtual worlds: Comparison of joystick-based and head-controlled paradigms. In *Proceedings of the 19th ACM VRST*, pages 111–114, New York, NY, USA, 2013. ISBN 978-1-4503-2379-6. doi: 10.1145/2503713.2503754.

Chen, W., Clavel, C., Férey, N., and Bourdot, P. Perceptual conflicts in a multi-stereoscopic immersive virtual environment: Case study on face-to-face interaction through an avatar. *Presence: Teleoperators and Virtual Environments*, 23(4):410–429, 2014. ISSN 1054-7460. doi: 10.1162/PRES_a_00209.

Chen, W., Ladeveze, N., Clavel, C., Mestre, D., and Bourdot, P. User cohabitation in multi-stereoscopic immersive virtual environment for individual navigation tasks. In *Proceedings of IEEE Virtual Reality*, pages 47–54, 2015.

Churchill, E. F. and Snowdon, D. Collaborative virtual environments: an introductory review of issues and systems. *Virtual Reality*, 3(1):3–15, 1998.

Cirio, G., Marchal, M., Regia-Corte, T., and Lécuyer, A. The magic barrier tape: A novel metaphor for infinite navigation in virtual worlds with a restricted walking workspace. In *Proceedings of the 16th ACM VRST*, pages 155–162, New York, NY, USA, 2009. ISBN 978-1-60558-869-8. doi: 10.1145/1643928.1643965.

Cirio, G., Vangorp, P., Chapoulie, E., Marchal, M., Lécuyer, A., and Drettakis, G. 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)*, 18(4), 2012.

Clark, H. H. and Brennan, S. E. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.

Cobb, S. V., Nichols, S., Ramsey, A., and Wilson, J. R. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators and Virtual Environments*, 8(2):169–186, 1999.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., and Hart, J. C. The CAVE: Audio visual experience automatic virtual environment. *Commun. ACM*, 35(6):64–72, June 1992. ISSN 0001-0782. doi: 10.1145/129888.129892.

Dillman, D. A. *Mail and internet surveys: The tailored design method*, volume 2. Wiley New York, 2000.

Doat, T., Parizot, E., and Vézien, J.-M. Novative rendering and physics engines to apprehend special relativity. In *Proceedings of the 17th Eurographics Conference on Virtual Environments & Third Joint Virtual Reality*, EGVE - JVRC'11, pages 9–18, Aire-la-Ville, Switzerland, Switzerland, 2011. Eurographics Association. ISBN 978-3-905674-33-0. doi: 10.2312/EGVE/JVRC11/009-018.

Dodds, T. J., Mohler, B. J., and Bülthoff, H. H. Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PloS one*, 6(10): e25759, 2011.

Dodgson, N. A. Variation and extrema of human interpupillary distance. In *Electronic Imaging 2004*, pages 36–46. International Society for Optics and Photonics, 2004.

Dodgson, N. A. Autostereoscopic 3d displays. *Computer*, (8):31–36, 2005.

Dominjon, L., Lecuyer, A., Burkhardt, J.-M., Andrade-Barroso, G., and Richir, S. The "bubble" technique: interacting with large virtual environments using haptic devices with limited workspace. In *Eurohaptics Conference, World Haptics 2005. First Joint*, pages 639–640, March 2005. doi: 10.1109/WHC.2005.126.

Dourish, P. and Bellotti, V. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 107–114. ACM, 1992.

Duchowski, A. *Eye tracking methodology: Theory and practice*, volume 373. Springer Science & Business Media, 2007.

Duval, T., Lécuyer, A., and Thoma, S. Skewer: a 3d interaction technique for 2-user collaborative manipulation of objects in virtual environments. In *3D User Interfaces, 2006. 3DUI 2006. IEEE Symposium on*, pages 69–72. IEEE, 2006.

d'Angelo, D., Wesche, G., Foursa, M., and Bogen, M. The benefits of co-located collaboration and immersion on assembly modeling in virtual environments. In *Advances in Visual Computing*, pages 478–487. Springer, 2008.

Ellis, C. A., Gibbs, S. J., and Rein, G. Groupware: some issues and experiences. *Communications of the ACM*, 34(1):39–58, 1991.

Ellis, S. *Pictorial communication in real and virtual environments*. CRC Press, 1991a.

Ellis, S. R. Nature and origins of virtual environments: A bibliographical essay. *Computing Systems in Engineering*, 2(4):321–347, 1991b.

Ennis, C., McDonnell, R., and O'Sullivan, C. Seeing is believing: body motion dominates in multisensory conversations. In *ACM Transactions on Graphics (TOG)*, volume 29, page 91. ACM, 2010.

Férey, N., Nelson, J., Martin, C., Picinali, L., Bouyer, G., Tek, A., Bourdot, P., Burkhardt, J.-M., Katz, B. F., Ammi, M., et al. Multisensory vr interaction for protein-docking in the corsaire project. *Virtual Reality*, 13(4):273–293, 2009.

Fleury, C., Chauffaut, A., Duval, T., Gouranton, V., Arnaldi, B., et al. A generic model for embedding users' physical workspaces into multi-scale collaborative virtual environments. In *ICAT 2010 (20th International Conference on Artificial Reality and Telexistence)*, 2010.

Fox, J., Arena, D., and Bailenson, J. N. Virtual reality: a survival guide for the social scientist. *Journal of Media Psychology*, 21(3):95–113, 2009.

Froehlich, B., Hoffmann, J., Klueger, K., and Hochstrate, J. Implementing multi-viewer time-sequential stereo displays based on shuttered lcd projectors. In *4th Immersive Projection Technology Workshop, Ames, Iowa (May 2004)*. Citeseer, 2004.

Fröhlich, B., Blach, R., Stefani, O., Hochstrate, J., Hoffmann, J., Klüger, K., and Bues, M. Implementing multi-viewer stereo displays. In *WSCG*, pages 139–146, 2005.

Fuchs, P., Moreau, G., and Guitton, P. *Virtual reality: concepts and technologies*. CRC Press, 2011.

Garau, M. *The impact of avatar fidelity on social interaction in virtual environments*. PhD thesis, University College London, 2003.

Garau, M. Selective fidelity: Investigating priorities for the creation of expressive avatars. In *Avatars at Work and Play*, pages 17–38. Springer, 2006.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536. ACM, 2003.

Greenhalgh, C. and Benford, S. Massive: A virtual reality system for tele-conferencing. *ACM Transactions on Computer Human Interfaces (TOCHI)*, 2(3):239–261, 1995.

Greenhalgh, C., Purbrick, J., and Snowdon, D. Inside massive-3: flexible support for data consistency and world structuring. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 119–127. ACM, 2000.

Grossman, T. and Balakrishnan, R. Collaborative interaction with volumetric displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 383–392. ACM, 2008.

Grudin, J. Computer-supported cooperative work: History and focus. *Computer*, (5):19–26, 1994.

Grudin, J. and Poltrock, S. Taxonomy and theory in computer supported cooperative work. *The oxford handbook of organizational psychology*, 2:1323–1348, 2012.

Gutwin, C. and Greenberg, S. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 207–216. ACM, 1998.

Guye-Vuillème, A., Capin, T. K., Pandzic, S., Thalmann, N. M., and Thalmann, D. Nonverbal communication interface for collaborative virtual environments. *Virtual Reality*, 4(1):49–59, 1999.

Hale, K. S. and Stanney, K. M. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press, 2014.

Hand, C. A survey of 3d interaction techniques. In *Computer graphics forum*, volume 16, pages 269–281. Wiley Online Library, 1997.

Harwood, K. and Foley, P. Temporal resolution: An insight into the video display terminal (vdt)"problem". *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(4):447–452, 1987.

Hayashi, Y., Kojiri, T., and Watanabe, T. Immersive round-table interface in collaborative learning. In *Knowledge-Based Intelligent Information and Engineering Systems*, pages 769–776. Springer, 2007.

Heeter, C. Being there: The subjective experience of presence. *Presence: Teleoperators and virtual environments*, 1(2):262–271, 1992.

Held, R. T. and Banks, M. S. Misperceptions in stereoscopic displays: a vision science perspective. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 23–32. ACM, 2008.

Hellström, T. Robot navigation with potential fields. *Uminf-11.18, Department of Computing Science, Umeå University*, 2011.

Inkpen, K., Hawkey, K., Kellar, M., Mandryk, R., Parker, K., Reilly, D., Scott, S., and Whalen, T. Exploring display factors that influence co-located collaboration: angle, size, number, and user arrangement. In *Proc. HCI international*, 2005.

Interrante, V., Ries, B., and Anderson, L. Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments. In *3D User Interfaces, 2007. 3DUI '07. IEEE Symposium on*, pages –, March 2007. doi: 10.1109/3DUI.2007.340791.

Isenberg, P., Fisher, D., Paul, S., Morris, M. R., Inkpen, K., Czerwinski, M., et al. Co-located collaborative visual analytics around a tabletop display. *Visualization and Computer Graphics, IEEE Transactions on*, 18(5):689–702, 2012.

Iwata, H. and Yoshida, Y. Path reproduction tests using a torus treadmill. *Presence: Teleoperators and Virtual Environments*, 8(6):587–597, dec 1999. ISSN 1054-7460. doi: 10.1162/105474699566503.

Johansen, R. *Groupware: Computer support for business teams*. The Free Press, 1988.

Joslin, C., Giacomo, T. D., and Magnenat-Thalmann, N. Collaborative virtual environments: from birth to standardization. *Communications Magazine, IEEE*, 42(4):28–33, 2004.

Katz, B. F., Felinto, D., Touraine, D., Poirier-Quinot, D., and Bourdot, P. BlenderVR: Open-source framework for interactive and immersive VR. In *IEEE Virtual Reality (IEEE VR)*, pages 1–2, Arles, France, 2015.

Kazman, R. Making waves: On the design of architectures for low-end distributed virtual environments. In *Virtual Reality Annual International Symposium, IEEE*, pages 443–449, 1993.

Kennedy, R. S. and Fowlkes, J. E. Simulator sickness is polygenic and polysymptomatic: Implications for research. *The International Journal of Aviation Psychology*, 2(1):23–38, 1992.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3.

Khatib, O. Real-time obstacle avoidance for manipulators and mobile robots. *The international journal of robotics research*, 5(1):90–98, 1986.

Kim, T. and Biocca, F. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion.[1]. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.

Kim, Y. Y., Kim, H. J., Kim, E. N., Ko, H. D., and Kim, H. T. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42(5):616–625, 2005.

Kistler, D. J. and Wightman, F. L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3):1637–1647, 1992.

Kitamura, Y., Konishi, T., Yamamoto, S., and Kishino, F. Interactive stereoscopic display for three or more users. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 231–240. ACM, 2001.

Kohli, L., Burns, E., Miller, D., and Fuchs, H. Combining passive haptics with redirected walking. In *Proceedings of the 2005 international conference on Augmented tele-existence*, pages 253–254. ACM, 2005.

Konrad, T., Demirdjian, D., and Darrell, T. Gesture + play: full-body interaction for virtual environments. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pages 620–621. ACM, 2003.

Kooima, R., Prudhomme, A., Schulze, J., Sandin, D., and DeFanti, T. A multi-viewer tiled autostereoscopic virtual reality display. In *proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, pages 171–174. ACM, 2010.

Kulik, A., Kunert, A., Beck, S., Reichel, R., Blach, R., Zink, A., and Froehlich, B. C1x6: A stereoscopic six-user display for co-located collaboration in shared virtual environments. In *Proceedings of the 2011 SIGGRAPH Asia Conference*, SA '11, pages 188:1–188:12, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0807-6. doi: 10.1145/2024156.2024222.

Lanier, J. Virtual reality: The promise of the future. *Interactive Learning International*, 8(4): 275–79, 1992.

Latombe, J.-C. *Robot motion planning*, volume 124. Springer Science & Business Media, 2012.

LaViola, J. J., Jr. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1): 47–56, January 2000. ISSN 0736-6906. doi: 10.1145/333329.333344.

Lawson, B. D., Graeber, D. A., Mead, A. M., and Muth, E. Signs and symptoms of human syndromes associated with synthetic experiences. *Handbook of virtual environments: Design, implementation, and applications*, pages 589–618, 2002.

Lissermann, R., Huber, J., Schmitz, M., Steimle, J., and Mühlhäuser, M. Permulin: Mixed-focus collaboration on multi-view tabletops. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3191–3200, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557405.

Lok, B., Naik, S., Whitton, M. C., and Brooks, F. P. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(6):615–628, 2003.

Lombard, M. and Ditton, T. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.

López, P. G. and Skarmeta, A. F. G. Ants framework for cooperative work environments. *Computer*, (3):56–62, 2003.

Lopez, T., Bouville, R., Loup-Escande, E., Nouviale, F., Gouranton, V., and Arnaldi, B. Exchange of avatars: Toward a better perception and understanding. *Visualization and Computer Graphics, IEEE Transactions on*, 20(4):644–653, 2014.

Lucente, M. Interactive three-dimensional holographic displays: seeing the future in depth. *ACM SIGGRAPH Computer Graphics*, 31(2):63–67, 1997.

Macedonia, M. R., Zyda, M. J., Pratt, D. R., Barham, P. T., and Zeswitz, S. Npsnet: A network software architecture for large scale virtual environments, 1994.

Magnenat-Thalmann, N. and Egges, A. Interactive virtual humans in real-time virtual environments. *IJVR*, 5(2):15–24, 2006.

Maksakov, E., Booth, K. S., and Hawkey, K. Whale tank virtual reality. In *Proceedings of Graphics Interface 2010*, pages 185–192. Canadian Information Processing Society, 2010.

Marchal, M., Pettre, J., and Lecuyer, A. Joyman: A human-scale joystick for navigating in virtual worlds. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces*, 3DUI '11, pages 19–26, Washington, DC, USA, 2011. IEEE Computer Society. ISBN 978-1-4577-0063-7.

Margery, D., Arnaldi, B., and Plouzeau, N. A general framework for cooperative manipulation in virtual environments. In Gervautz, M., Schmalstieg, D., and Hildebrand, A., editors, *Virtual Environments '99*, Eurographics, pages 169–178. Springer Vienna, 1999. ISBN 978-3-211-83347-6. doi: 10.1007/978-3-7091-6805-9_17.

Martin, P., Bourdot, P., and Touraine, D. A reconfigurable architecture for multimodal and collaborative interactions in virtual environments. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*, pages 11–14. IEEE, 2011.

Martin, P., Férey, N., Clavel, C., Darses, F., and Bourdot, P. Sensorimotor feedback for interactive realism: Evaluation of a haptic driving paradigm for a forklift simulator. In *Proceedings of the 2012 International Conference on Haptics: Perception, Devices, Mobility, and Communication - Volume Part I*, EuroHaptics'12, pages 314–325, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-31400-1. doi: 10.1007/978-3-642-31401-8_29.

Matsukura, H., Nihei, T., and Ishida, H. Multi-sensorial field display: Presenting spatial distribution of airflow and odor. In *Virtual Reality Conference (VR), IEEE*, pages 119–122, 2011.

McGinity, M., Shaw, J., Kuchelmeister, V., Hardjono, A., and Favero, D. D. AVIE: a versatile multi-user stereo 360 interactive vr theatre. In *Proceedings of the 2007 workshop on Emerging displays technologies: images and beyond: the future of displays and interacton*, page 2. ACM, 2007.

McMahan, R., Bowman, D., Zielinski, D., and Brady, R. Evaluating display fidelity and interaction fidelity in a virtual reality game. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):626–633, April 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2012.43.

Meehan, M., Insko, B., Whitton, M., and Brooks Jr, F. P. Physiological measures of presence in stressful virtual environments. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 645–652. ACM, 2002.

Meyer, K., Applewhite, H. L., and Biocca, F. A. A survey of position trackers. In *Presence: Teleoperators and Virtual Environments*, volume 1, pages 173–200. MIT Press, 1992.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. Augmented reality: A class of displays on the reality-virtuality continuum. In *Photonics for Industrial Applications*, pages 282–292. International Society for Optics and Photonics, 1995.

Min, B.-C., Chung, S.-C., Min, Y.-K., and Sakamoto, K. Psychophysiological evaluation of simulator sickness evoked by a graphic simulator. *Applied ergonomics*, 35(6):549–556, 2004.

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., and Bülthoff, H. H. The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoperators and Virtual Environments*, 19(3):230–242, 2010.

Mori, M. The uncanny valley. *Energy*, 7(4):33–35, 1970.

Mori, M., MacDorman, K. F., and Kageki, N. The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, 19(2):98–100, 2012.

Mulder, J. D. and Boscker, B. A modular system for collaborative desktop vr/ar with a shared workspace. In *Virtual Reality, 2004. Proceedings. IEEE*, pages 75–280. IEEE, 2004.

Nam, C. S., Shu, J., and Chung, D. The roles of sensory modalities in collaborative virtual environments (cves). *Computers in Human Behavior*, 24(4):1404–1417, 2008.

Narayan, M., Waugh, L., Zhang, X., Bafna, P., and Bowman, D. Quantifying the benefits of immersion for collaboration in virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 78–81. ACM, 2005.

Narumi, T., Nishizaka, S., Kajinami, T., Tanikawa, T., and Hirose, M. Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 93–102. ACM, 2011.

Normand, J.-M., Spanlang, B., Tecchia, F., Carrozzino, M., Swapp, D., and Slater, M. Full body acting rehearsal in a networked virtual environment-a case study. *Presence: Teleoperators and Virtual Environments*, 21(2):229–243, apr 2012. ISSN 1054-7460. doi: 10.1162/PRES_a_00089.

Ogi, T., Yamada, T., Tamagawa, K., Kano, M., and Hirose, M. Immersive telecommunication using stereo video avatar. In *Virtual Reality, 2001. Proceedings. IEEE*, pages 45–51, 2001. doi: 10.1109/VR.2001.913769.

Oguz, S. O., Kucukyilmaz, A., Sezgin, T. M., and Basdogan, C. Haptic negotiation and role exchange for collaboration in virtual environments. In *Haptics Symposium, 2010 IEEE*, pages 371–378. IEEE, 2010.

Okada, K. Collaboration support in the information sharing space. *IPSJ Magazine*, 48(2): 123–125, 2007.

Ortega, M. and Coquillart, S. Prop-based haptic interaction with co-location and immersion: an automotive application. haptic audio visual environments and their applications. In *2005 IEEE International Workshop on*, page 6, 2005.

Otto, O., Roberts, D., and Wolff, R. A review on effective closely-coupled collaboration using immersive cve's. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, pages 145–154. ACM, 2006.

Oviatt, S. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81, 1999.

Paggio, P. and Jongejan, B. Multimodal communication in virtual environments. In *Multimodal Intelligent Information Presentation*, pages 27–45. Springer, 2005.

Pausch, R. and Crea, T. A literature survey for virtual environments: Military flightsimulator visual systems and simulator sickness. 1992.

Peck, T. C., Whitton, M. C., and Fuchs, H. Evaluation of reorientation techniques for walking in large virtual environments. In *IEEE VIrtual Reality*, pages 121–127, 2008.

Perlin, K., Paxia, S., and Kollin, J. S. An autostereoscopic display. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 319–326. ACM Press/Addison-Wesley Publishing Co., 2000.

Pinho, M. S., Bowman, D. A., and Freitas, C. M. Cooperative object manipulation in immersive virtual environments: framework and techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 171–178. ACM, 2002.

Pollock, B., Burton, M., Kelly, J. W., Gilbert, S., and Winer, E. The right view from the wrong location: Depth perception in stereoscopic multi-user virtual environments. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):581–588, 2012.

Ponto, K., Gleicher, M., Radwin, R. G., and Shin, H. J. Perceptual calibration for immersive display environments. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4): 691–700, 2013.

Pouliquen-Lardy, L., Milleville-Pennel, I., Guillaume, F., and Mars, F. How role distribution influences choice of spatial reference frames in a virtual collaborative task. In *Spatial Cognition 2014*, pages 106–108. Universität Bremen/Universität Freiburg, 2014.

Razzaque, S., Swapp, D., Slater, M., Whitton, M. C., and Steed, A. Redirected walking in place. In *Proceedings of the Workshop on Virtual Environments*, EGVE '02, pages 123–130, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association. ISBN 1-58113-535-1.

Reason, J. T. and Brand, J. J. *Motion sickness.* Academic press, 1975.

Rheingold, H. *Virtual reality*. Summit Books, New York, USA, 1991.

Rich, C. J. and Braun, C. C. Assessing the impact of control and sensory compatibility on sickness in virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 40, pages 1122–1125. SAGE Publications, 1996.

Riecke, B. E. and Feuereissen, D. To move or not to move: can active control and user-driven motion cueing enhance self-motion perception (vection) in virtual reality? In *Proceedings of the ACM Symposium on Applied Perception*, pages 17–24. ACM, 2012.

Riecke, B. E., Cunningham, D. W., and Bülthoff, H. H. Spatial updating in virtual reality: the sufficiency of visual information. *Psychological research*, 71(3):298–313, 2007.

Riege, K., Holtkamper, T., Wesche, G., and Frohlich, B. The bent pick ray: An extended pointing technique for multi-user interaction. In *3D User Interfaces, 3DUI 2006. IEEE Symposium on*, pages 62–65. IEEE, 2006.

Roberts, D., Wolff, R., Otto, O., and Steed, A. Constructing a gazebo: supporting teamwork in a tightly coupled, distributed task in virtual reality. *Presence: Teleoperators and Virtual Environments*, 12(6):644–657, 2003.

Roberts, D., Wolff, R., Otto, O., Kranzlmueller, D., Anthes, C., and Steed, A. Supporting social human communication between distributed walk-in displays. In *Proceedings of the ACM symposium on Virtual reality software and technology*, VRST '04, pages 81–88, New York, NY, USA, 2004. ACM. ISBN 1-58113-907-1. doi: 10.1145/1077534.1077550.

Ruddle, R. A. and Lessels, S. The benefits of using a walking interface to navigate virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 16(1):5:1–5:18, apr 2009. ISSN 1073-0516. doi: 10.1145/1502800.1502805.

Ruddle, R. A., Savage, J. C., and Jones, D. M. Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(4):285–308, 2002.

Salas, E., Prince, C., Baker, D. P., and Shrestha, L. Situation awareness in team performance: Implications for measurement and training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):123–136, 1995.

Salzmann, H. and Froehlich, B. The two-user seating buck: Enabling face-to-face discussions of novel car interface concepts. In *Virtual Reality Conference, VR '08. IEEE*, pages 75–82, 2008. doi: 10.1109/VR.2008.4480754.

Salzmann, H., Jacobs, J., and Froehlich, B. Collaborative interaction in co-located two-user scenarios. In *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, JVRC'09, pages 85–92, Aire-la-Ville, Switzerland, Switzerland, 2009a. Eurographics Association. ISBN 978-3-905674-20-0. doi: 10.2312/EGVE/JVRC09/085-092.

Salzmann, H., Moehring, M., and Froehlich, B. Virtual vs. real-world pointing in two-user scenarios. *2014 IEEE Virtual Reality (VR)*, 0:127–130, 2009b. doi: http://doi.ieeecomputersociety.org/10.1109/VR.2009.4811011.

Sato, M. Development of string-based force display: Spidar. In *8th International Conference on Virtual Systems and Multimedia*, pages 1034–1039. Citeseer, 2002.

Schmalstieg, D., Fuhrmann, A., Hesina, G., Szalavári, Z., Encarnaçao, L. M., Gervautz, M., and Purgathofer, W. The studierstube augmented reality project. *Presence: Teleoperators and Virtual Environments*, 11(1):33–54, 2002.

Schroeder, R. Copresence and interaction in virtual environments: An overview of the range of issues. In *Presence 2002: Fifth International Workshop*, pages 274–295, 2002.

Schroeder, R. Being there together and the future of connected presence. *Presence: Teleoperators and Virtual Environments*, 15(4):438–454, 2006.

Schroeder, R., Steed, A., Axelsson, A.-S., Heldal, I., Abelin, Å., Wideström, J., Nilsson, A., and Slater, M. Collaborating in networked immersive spaces: as good as being there together? *Computers & Graphics*, 25(5):781–788, 2001.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., and Van Der Mast, C. A. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001.

Schulze, J. P., Acevedo, D., Mangan, J., Prudhomme, A., Nguyen, P., and Weber, P. Democratizing rendering for multiple viewers in surround vr systems. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*, pages 77–80, 2012.

Scott, S. D., Grant, K. D., and Mandryk, R. L. System guidelines for co-located, collaborative work on a tabletop display. In *ECSCW 2003*, pages 159–178. Springer, 2003.

Shelton, J. and Kumar, G. P. Comparison between auditory and visual simple reaction times. *Neuroscience & Medicine*, 1(01):30, 2010.

Shepard, R. N. and Metzler, J. Mental rotation of three-dimensional objects. *Science*, 171 (3972):701–703, 1971.

Sheridan, T. B. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environment*, 1(1):120–126, jan 1992. ISSN 1054-7460.

Simon, A., Smith, R., and Pawlicki, R. Omnistereo for panoramic virtual environment display systems. In *Virtual Reality, 2004. Proceedings. IEEE*, pages 67–279, March 2004. doi: 10.1109/VR.2004.1310057.

Simon, A. First-person experience and usability of co-located interaction in a projection-based virtual environment. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 23–30. ACM, 2005.

Simon, A. Usability of multiviewpoint images for spatial interaction in projection-based display systems. *Visualization and Computer Graphics, IEEE Transactions on*, 13(1):26–33, 2007.

Simon, A. and Göbel, M. The i-cone™ - a panoramic display system for virtual environments. In *Computer Graphics and Applications, 2002. Proceedings. 10th Pacific Conference on*, pages 3–7. IEEE, 2002. doi: 10.1109/PCCGA.2002.1167834.

Slater, M. and Steed, A. Meeting people virtually: Experiments in shared virtual environments. In *The social life of avatars*, pages 146–171. Springer, 2002.

Slater, M. and Usoh, M. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality*, 1, 1994.

Slater, M., Usoh, M., and Steed, A. Depth of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 3(2):130–144, 1994.

Slater, M., Sadagic, A., Usoh, M., and Schroeder, R. Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators and virtual environments*, 9(1): 37–51, 2000.

Smith, G. Cooperative virtual environments: lessons from 2d multi user interfaces. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 390–398. ACM, 1996.

So, R. H., Lo, W., and Ho, A. T. Scene movement: An important cause of cybersickness. In *Seminars on Virtual reality: selected tools and applications*, 2002.

Spence, C. Just how important is spatial coincidence to multisensory integration? evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296(1):31–49, 2013.

Stanney, K. M., Kennedy, R. S., and Drexler, J. M. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 41, pages 1138–1142. SAGE Publications, 1997.

Stanney, K. M., Kingdon, K. S., Graeber, D., and Kennedy, R. S. Human performance in immersive virtual environments: Effects of exposure duration, user control, and scene complexity. *Human Performance*, 15(4):339–366, 2002. doi: 10.1207/S15327043HUP1504_03.

Stefik, M., Bobrow, D. G., Foster, G., Lanning, S., and Tatar, D. Wysiwis revised: Early experiences with multiuser interfaces. *ACM Trans. Inf. Syst.*, 5(2):147–167, apr 1987. ISSN 1046-8188. doi: 10.1145/27636.28056.

Steuer, J., Biocca, F., Levy, M. R., et al. Defining virtual reality: Dimensions determining telepresence. *Communication in the age of virtual reality*, pages 33–56, 1995.

Stoakley, R., Conway, M. J., and Pausch, R. Virtual reality on a wim: Interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. ISBN 0-201-84705-1. doi: 10.1145/223904.223938.

Sugita, N., Yoshizawa, M., Tanaka, A., Abe, K., Chiba, S., Yambe, T., and Nitta, S.-i. Quantitative evaluation of effects of visually-induced motion sickness based on causal coherence functions between blood pressure and heart rate. *Displays*, 29(2):167–175, 2008.

Suied, C., Bonneel, N., and Viaud-Delmon, I. Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*, 194(1):91–102, 2009.

Sutherland, I. E. 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.

Takaki, Y. and Nago, N. Multi-projection of lenticular displays to construct a 256-view super multi-view display. *Optics express*, 18(9):8824–8835, 2010.

Thalmann, D. The role of virtual humans in virtual environment technology and interfaces. In *Frontiers of Human-Centered Computing, Online Communities and Virtual Environments*, pages 27–38. Springer, 2001.

Usoh, M., Catena, E., Arman, S., and Slater, M. Using presence questionnaires in reality. *Presence*, 9(5):497–503, 2000.

Verheijen, E. N. G. *Sound reproduction by wave field synthesis*. PhD thesis, TU Delft, Delft University of Technology, 1998.

Vishwanath, D., Girshick, A. R., and Banks, M. S. Why pictures look right when viewed from the wrong place. *Nature neuroscience*, 8(10):1401–1410, 2005.

von Kapri, A., Rick, T., and Feiner, S. Comparing steering-based travel techniques for search tasks in a CAVE. In *IEEE Virtual Reality Conference (VR'11)*, pages 91–94, march 2011. doi: 10.1109/VR.2011.5759443.

Ware, C. and Osborne, S. Exploration and virtual camera control in virtual three dimensional environments. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, I3D '90, pages 175–183, New York, NY, USA, 1990. ACM. ISBN 0-89791-351-5. doi: 10.1145/91385.91442.

Welch, G. and Foxlin, E. Motion tracking survey. *IEEE Computer graphics and Applications*, pages 24–38, 2002.

Williams, B., Narasimham, G., Rump, B., McNamara, T. P., Carr, T. H., Rieser, J., and Bodenheimer, B. Exploring large virtual environments with an HMD when physical space is limited. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*,

APGV '07, pages 41–48, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-670-7. doi: 10.1145/1272582.1272590.

Witmer, B. G. and Singer, M. J. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, June 1998. ISSN 1054-7460. doi: 10.1162/105474698565686.

Wolff, R., Roberts, D. J., Steed, A., and Otto, O. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology*, 29(1):11–26, 2007.

Woods, A. J., Docherty, T., and Koch, R. Image distortions in stereoscopic video systems. In *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, pages 36–48. International Society for Optics and Photonics, 1993.

Zhao, S. Toward a taxonomy of copresence. *Presence: Teleoperators and Virtual Environments*, 12(5):445–455, 2003.