الجمهورية الجزائرية الديموقراطية الشعبية
وزارة التعليـم العالــي و البحــث العلمــي

Université Ferhat ABBAS Sétif 1
Faculté des Sciences

جامعـة فرحـات عباس سطيـف 1
كـلـيـة الـعـلـوم

Département : Informatique

MEMOIRE DE MASTER

DOMAINE : Mathématiques Informatique
FILIERE : Informatique
SPECIALITE : Génie Logiciel

Thème

Study of NUIs for virtual and augmented reality applications

Présenté par :
Abderahim Hamani

Dirigé par:
Pr. Mohamed Touahria

Promotion : 2019/2020

Study of NUIs for virtual and augmented reality applications

Master thesis

By

HAMANI ABDERAHIM

Department of Computer science
UNIVERSITY OF FERHAT ABBAS

A dissertation submitted to the University of Ferhat Abbas in accordance with the requirements of Master degree in COMPUTING SCIENCE in the Faculty of sciences.

SEPTEMBER 2020

Supervised by : Prof. Touahria Mohamed

Abstract

Currently we are living the transition from graphical user interfaces to natural user interfaces, which was made possible by the appearance of new technologies. This new paradigm offers new opportunities of interaction by allowing the user to interact directly with the digital content without the intermediate device. The interaction is carried out by the actions of the human body such as voice, brain activity, gaze, hand or head movement. This thesis explores the state of the art of natural user interfaces and the potential VR/AR technologies. Then, A conceptual model of NUIs for virtual and augmented reality applications is proposed to aid the designer to plan their design project for a natural 3D interaction technique. At the end, the model is applied to our Virtual reality game to define a natural and multimodel 3D interaction technique, based on hand gesture, voice and head movement modalities.

Keywords: 3D interaction, Virtual reality, Augmented reality, Natural user interface, Human-computer interaction.

RÉSUMÉ

Actuellement nous vivons la transition des interfaces graphiques vers les interfaces naturelles. cette transition est devenue possible grâce à l'apparition de nouvelles technologies. Ce paradigm offre de nouvelles opportunités d'interaction en permettant à l'utilisateur d'interagir directement avec le contenu digital sans l'utilisation d'un moyen intermédiaire. L'interaction maintenant est faite par les actions de l'utilisateur lui même (la voix, le cerveau, la main ou les mouvements de la tête). Ce mémoire explore l'état de l'art des interfaces naturelles et le potentiel réalité virtuelle/réalité augmentée. Un model conceptual des interfaces naturelles pour les applications réalité virtuelle et augmentée est proposé pour aider les concepteurs à planifier leurs projets de conception pour les techniques de l'interaction. À la fin, le modèle est appliqué à notre jeu de réalité virtuelle pour définir une technique d'interaction 3D naturelle et multimodel, basée sur les trois modalités: voix, mouvement de la tête et la main.

Mots clés: interaction 3D, réalité virtuelle, réalité augmentée, NUI, Interactions homme-machine.

ملخص

في الوقت الحاضر نشهد الانتقال من واجهات المستخدم الرسومية إلى واجهات المستخدم الطبيعية ، والتي أصبحت ممكنة بفضل ظهور التقنيات الجديدة. يوفر هذا النهج الجديد فرصاً جديدة للتفاعل ، من خلال السماح للمستخدم بالتفاعل مباشرة مع المحتوى الرقمي بدون الجهاز الوسيط. يتم التفاعل بواسطة الجسم البشري مثل الصوت ونشاط الدماغ والنظرة وحركة اليد اوالرأس. هذه الأطروحة تقترح شرحا لأحدث تقنيات واجهات المستخدم الطبيعية وتقنيات الواقع الافتراضي المعزز المحتملة. تم اقتراح نموذج مفاهيمي لواجهة المستخدم في تطبيقات الواقع الافتراضي و المعزز لمساعدة المصمم على تخطيط مشروع التصميم الخاص به لتقنية التفاعل ثلاثي الأبعاد الطبيعية. ثم ، يتم تطبيق النموذج على لعبة الواقع الافتراضي لدينا لتحديد تقنية تفاعل ثلاثية الأبعاد طبيعية ومتعددة النماذج ، استناداً إلى إيماءات اليد وصوت وحركة الرأس.

الكلمات المفتاحية : التفاعل ثلاثي الأبعاد ، الواقع الافتراضي ، الواقع المعزز ، واجهة المستخدم الطبيعية ، التفاعل بين الإنسان والحاسوب.

ACKNOWLEDGEMENTS

At the end of this thesis, in the first place, i will give thanks to Allah who made my work easier by putting in my path people who have helped and supported me enormously.

I would like to give my sincere thanks to my mother and father and every member of my family who encouraged me day and night. I would like to thank every one pushed me forward to successed, i'm really greatfull to all of you.

My heartfelt thanks to Professor Khababa Abdallah for his help and encouragement that he never stopped communicating to me..

HAMANI Abderahim

TABLE OF CONTENTS

List of Tables

List of Figures

Figure **Page**

INTRODUCTION

This chapter gives a highlight on the motivation, problem statement and the general organization of this work. At the beginning, the motivation and the statement of the work are lightly described. Subsequently, the aim of this work is well explained and then we conclude with a vague outline of the general organization of this work.

Problematic and Motivation

After the dominance of the WIMP paradigm during 30 years, here comes the era of a new transition of interface paradigms from graphical to natural user interfaces. These interfaces are expected to facilitate our interaction with systems without the use of extra devices, Since they use the learnt skills of human beings.

Natural User Interfaces intend to allow the user to interact in natural and intuitive way with computer systems. A lot of efforts have been included to enhance the user's experience by merging the new input and output devices such as head mounted displays, motion tracking cameras and controllers. This advances in the field of NUIs can provide new interaction techniques efficient and adapted to the context of use.

Nowadays, Virtual Reality (VR) and augmented Reality (AR) is mainly created by generating visual effects through different display device systems. Indeed, VR and AR have changed the interaction by providing a three dimensional (3D) immersive experience. which make the user's senses believe that they actually are inside the virtual environment.

As the technology moves forward, the designers are facing problems to create, study and evaluate natural user interfaces for VR and AR applications. Another problem is the lack of formalisms and decision-making tools to guide them in their planning design or to provide solutions for certain problems.

Aim of the work

The aim of this work is to study the concept of NUIs and its different natural interaction modalities (hand, voice, lip/body movement...etc), then a comparison between these interfaces were established. At the end we propose a conceptual model of NUIs for Virtual and Augmented Reality (VR / AR) applications, and present a new "serious" game that combines three modalities

for word learning.

In summary, the main objectives can be identified as follows:

- State of the art of natural user interfaces.
- Identification of common characteristics of NUIs.
- Comparisons between modalities according to these characteristics.
- Propose a conceptual model of NUIs for VR/AR applications.
- Implement multimodal virtual reality educational applications.

Thesis structure

This thesis is organized in four chapters as follows:

- **Chapter 1:** provides an overview on topics covered in this thesis, such as user interfaces, virtual and augmented reality. We specifically focus on the three dimensional interaction here.

- **Chapter 2:** this chapter discusses the state of the art of the natural user interfaces and modern technologies.

- **Chapter 3:** documents our proposed conceptual model of natural user interfaces for Virtual and augmented reality applications.

- **Chapter 4:** presents the system design and the implementation of multimodal Virtual reality educational application. In this chapter we include the description of interaction techniques and modalities used to interact with the system.

- **Conclusion:** this chapter summarizes the major findings of our study, and gives insights into future work.

CHAPTER 1

BACKGROUND AND PRELIMINARIES

This chapter introduced and define the key concepts and the major terms related to this work.This chapter is divided into three sections: first, an introduction to user interfaces is presented with brief history of their evolution, second, the concept of 3D interaction is further explained and detailed , and third, the last section is dedicated to Virtual and augmented reality.

1.1 User interface

As long as humans continue to interact with computer systems, user interfaces will be needed. The quality of an interface becomes more important in order to achieve high quality of communication between human and computer.

The next sections provide a definition for user interfaces, present a short review of user interfaces paradigms history and elaborate on the user experience concept.

1.1.1 Definitions

The interface is the part of the system that enables the user to see, feel and touch, which is the channel of communication between the user and the physical or abstract object [30]. The term "interface" can refer to either a hardware and software that mediates the interaction between humans and computers [15]. The goal of this interaction is to allow users to control and communicate with the virtual environment.

The design process of user interface includes all aspects of the computing system that are visible to the user, and must be a part of the design process of the computer system from the very beginning.

1.1.2 Evolution of interfaces paradigms

Humans and computer systems use user interfaces to communicate efficiently. As computer systems evolve, user interfaces evolve with them and offer new opportunities of interaction. The first type of interaction appears through the command line. User typing commands in command line interfaces (CLI) to operate in the computer. Afterwards came the graphical user interfaces (GUI). These types of interface were much simpler and easier to use. The GUI uses a mouse and keyboard to allow the user to control 2D space. Nowadays we are witnessing the transition from graphical user interfaces to natural user interfaces (NUI), which was made possible by the appearance of the new technologies. These technologies bring new opportunities to the user to interact with computers using his five sens. Some researchers are already discussing what will come next: organic user interfaces (OUI). These four interface paradigms are further explained in the next sections.

Command Line Interfaces

Represent the first type of interfaces that allow the users to interact with computers. With command line interfaces, the user is allowed to enter the commands into the computer via keyboard only. These interfaces support a big list of instructions which require the user to remember the commands in order to interact with the computer efficiently.

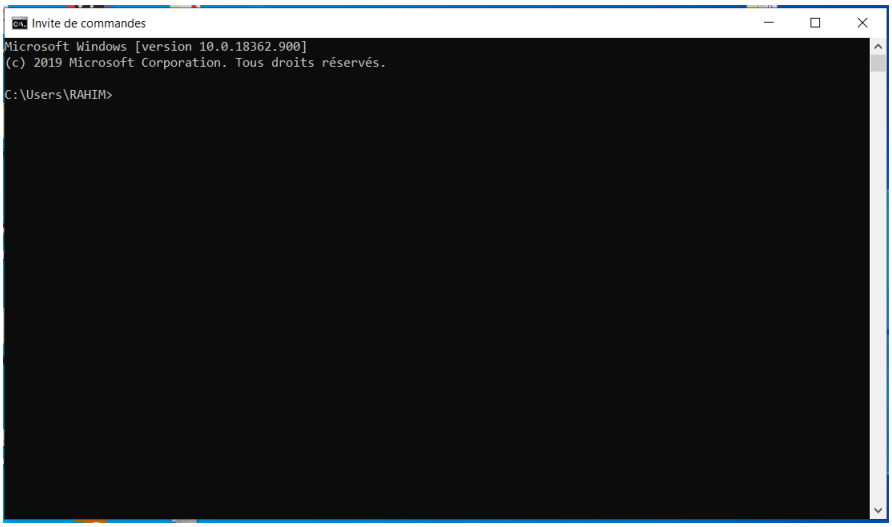

Figure 1.1: Windows 10 command line.

The command line interface requires skills from the user and uses a lot of cognitive load to enter a lot of commands to interact with the computer.Therefore, the user's experience feels abstract. However, these interfaces are still available and still used in a smaller niche by experienced users (Figure 1.1).

Graphical User Interfaces

Graphical user interfaces are a turning point in computing history. The interaction is much simpler and easier to use by the general public according to the command line interface. GUIs represent the content in graphical icons using metaphor. The metaphor helps the user to recognize the operation from the graphical objects. As an example the user can drag the folder to the recycle bin to delete it.

These interfaces start to become popular, and support another indirect interaction. The user is required to use an immediate device to communicate with the computer. This input device is a keyboard and mouse which is called a "X-Y Position Indicator" in the beginning.

The main design principle behind graphical user interfaces is What You See Is What You Get (WYSIWYG). This principle implies that what the user sees on the computer is what he will get as an end result.

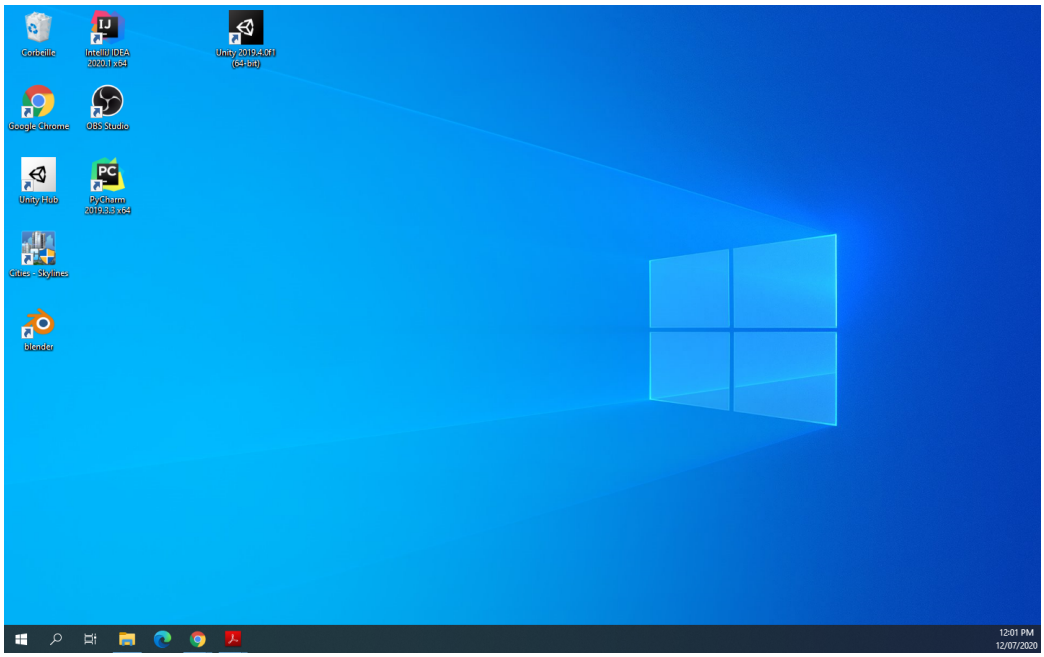

Figure 1.2: Windows 10 graphical user interface.

Another concept appeared with the graphical user interfaces is the concept of WIMP (Figure 1.2). WIMP stands for Windows, Icons, Menus and Pointer which are the four main elements that constitute a GUI.

Natural User Interfaces

Natural user interfaces are becoming more and more common every day, especially touch-screens and game consoles. The new devices facilitate spreading of this interaction paradigm. While the design of GUIs focus on What You See Is What You Get, NUIs focus on the human perception to achieve a natural interaction with the system. The main design principle behind natural user interfaces is What You Do Is What You Get (Figure 1.3).

Figure 1.3: Touch surface [77].

In NUIs, the interaction is direct, the user manipulates the object without the intermediaries devices such as mouse or keyboard. The virtual content is treated as a real object inside the virtual space.

Organic User Interfaces

User interface experts predict that organic user interface will replace natural user interface. In organic user interfaces everything is connected and fluid (Figure 1.4), like in an organic system[70]. This way of interaction should make the user forget that he is using or communicating with the machine. This interface will bring a new level of immersion and experience to the

user while using the objects, since The interactions between them is fully fluid.

Organic interfaces are user interfaces with fluid displays which make it easy to use since the form of the objects will clearly hint the user on how they should be used.

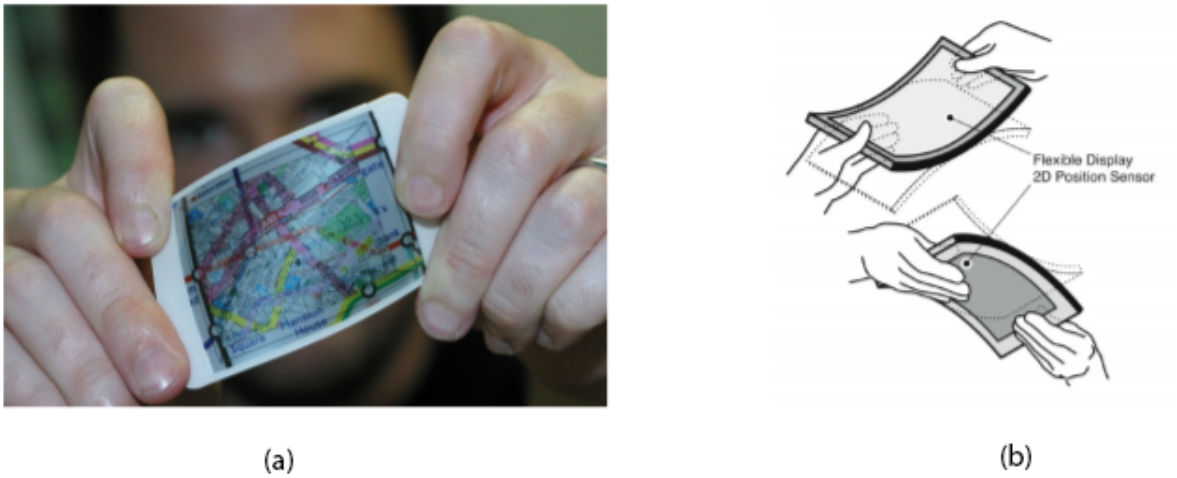

Figure 1.4: Organic user interface: (a) Gummi interface prototype. (b) Gummi interaction[70].

1.1.3 Paradigm transition

The transition between new interface paradigms and the present paradigm, does not necessarily mean that the old paradigm will disappear. As an example the command line interfaces are still used and useful in small precise operations with the current graphical user interfaces.

The current paradigm that dominates is the Graphical user interfaces. The majority of users are familiar with these digital experiences and they think it's enough to satisfy their needs. But in some domains, the graphical user interfaces paradigm has reached his limit and is not enough to achieve the goal. A new paradigm is emerging: that is natural user interfaces . This new paradigm allows the interaction to expand into another level that can serve users needs.

1.1.4 User Experience

The term of UX is related to the concept of user interface. The user experience is the user perception and feeling in the moment of interaction or the use of the system. This concept includes both of user interface and usability such as learnability, Efficiency, Effectiveness and Satisfaction.

The User experience is the internal that makes the system function well, and the User interface is who makes the application look nice and appealing.

1.2 3D interaction

The interaction consists of the direct manipulation of objects, it can be defined as a language of communication between human and machine. This language is the set of actions/reactions loop between human and computer through sensory and motor interfaces and interaction techniques [61].

1.2.1 Interaction paradigms

The interaction paradigm is a set of rules and techniques for a method that allows a user to interact with its virtual environment [76].

1.2.2 Interaction Metaphor

The metaphor of interaction means that a real object or concept is used as a virtual tool to interact with the virtual environment [72].

1.2.3 Interaction techniques

An interaction technique is the method that allows users to interact with an object in Virtual space [72]. It is how to use a device to select and manipulate virtual objects or accomplish a task on a computer. Interaction techniques may be simple or complex as a series of gestures [15].

The interaction technique describes the relationship between users and the machines they use. It determines how users can make inputs and how machines can return outputs back to users [80].

1.2.4 3D interaction tasks

Bowman [15] classified 3D interaction tasks in fourth categories:

- **Navigation:** is to explore the virtual world. Bowman [15] has divided the navigation in two types: the physical movement of the user from one place to another (travel) and the search of a route (way finding).
- **Selection:** is the designation of an object in the virtual environment.
- **Manipulation:** is the process that allows the user to change the properties of an object in the virtual world.
- **Application control:** allows users to execute a command to change the state of the system or the mode of interaction to achieve a specific objective.

1.2.5 Classification of 3D interaction Techniques

The existing 3D interaction techniques related to virtual environments can be classified by the universal interaction task (navigation, selection, manipulation and application control).

1.2.5.1 Selection and Manipulation Techniques

The user in a virtual and augmented environment has the ability to interact with one or more virtual objects. Thus, there are a variety of techniques that allow the user to select and manipulate virtual objects.

Poupyrev et al. [63] 1999 classify the selection and manipulation techniques into two main categories according to the position of the user and the distance between the user and the virtual object: exocentric techniques and egocentric techniques. The third category contains the hybrid techniques which combine both of them.

Exocentric interaction techniques

The user in this technique category is not inside the virtual environment but he/she has the ability to manipulate objects of the virtual world. The feeling of immersion will not be high since the user is considered as an actor who is not part of the scene. The metaphor of World In Miniature WIM is used to allow the user to interact indirectly with the virtual objects [9]. Therefore, the selection and the manipulation will be applied using the metaphor of the simple virtual hand.

Egocentric interaction techniques

Egocentric techniques consider the user within the virtual environment. Unlike exocentric category, these techniques provide more feeling of presence inside the virtual space and give the user a much better immersive experience. Egocentric techniques are divided into two types:[9]

- **Virtual hand:** or the simple virtual hand is a technique that allows a direct mapping of the user to use hand to select virtual objects in a virtual environment. Mappings is to make the virtual hand motion near to motion of the real hand. This technique is considered as the most natural and intuitive in Egocentric techniques. The user can select the virtual object by staying in contact with it or closing his wrist. The fundamental problem of this technique is that the only objects within the area of the user's reach can be selected and manipulated.

- **Virtual pointer:** this technique based on the metaphor of the virtual pointer, also uses a virtual representation of the hand. The selection was made by using a laser pointer, without touching the object by hand. An example of this technique is Ray-Casting, which is based

on the metaphor of the virtual ray. Unlike the virtual hand, an infinite laser can reach any point in the virtual environment. But Ray-Casting is problematic when it comes to selecting small objects that are covered by big one's since it selects the first object intersecting with the laser.

Hybrid interaction techniques

This Technique combines the characteristics of egocentric and/or exocentric techniques: [9]

- **HOMER:** is a hybrid technique that combines the Ray-casting technique for selection and the simple virtual hand for manipulation. HOMER is an abbreviation of Hand-centered Object Manipulation Extending Ray-casting. This technique switches between the mode of selection and manipulation to positioning and rotating the virtual object. HOMER allows a user to easily reposition an object no matter how far away is at the moment of selection. This technique also has the problem of Ray-casting.

- **Voodoo Dolls:** it is a hybrid technique that combines the Word In Miniature technique of the Exocentric category and one of the plan-image techniques of the Egocentric category. The "Voodoo Dolls" technique allows the user to create dolls which are a miniature representation of existing virtual objects in the 3D environment. The selection of the objects is done by the "head crusher" technique, and the manipulation can then be done using the simple virtual hand.

1.2.5.2 Application control techniques

The application control groups together all the indirect manipulation techniques on the 3D environment. Application control is different from other universal interaction tasks, in the sense that the user uses the services offered by the application itself. Several application control techniques have been designed, these techniques is detailed as follows: [9]

- **Graphical menus:** is the 3D equivalent of 2D menus that have proven to be a successful system control technique in desktop UIs. The graphical menus can take the two forms: a 3D menu or 2D menu, The 2D menu will be placed in 3D virtual space. Thus, the manipulation of the buttons will happen through the 3D selection/manipulation techniques.

- **Voice commands:** the issuing of voice commands can be performed via simple speech recognition to control the application. Speech recognition techniques are commonly used to initialize or order the system to perform a single operation.

- **Gestural commands:** control commands are associated with very specific gestures or posture, so the system can be controlled through the movement or the configuration of the hand.

- **Tools:** the system control can be done by other virtual or physical tools with which the user can interact with the environment. The use of physical devices can lead to increased usability. These tools are classified in three categories: physical tools, tangibles, and virtual tools.

1.3 Virtual Reality and Augmented Reality

1.3.1 Virtual Reality

Virtual Reality (VR) is an immersive computing technology that aims to provide a natural and intuitive interface to an information world. Ellis [28] considers "*VR as an advanced human-computer interface that simulates a realistic environment and allows participants to interact with it*".

1.3.2 Augmented Reality

Augmented Reality (AR) is a real-time direct or indirect view of a physical real-world environment that has the ability to superimpose virtual elements onto the real world to enhance the user's view. Azuma [5] defines the essential components of Augmented Reality with the following three characteristics:

- it combines the real and the virtual.
- it is interactive in real time.
- it is registered in 3D.

1.3.3 Similarities Between Virtual Reality and Augmented Reality

There are several similarities between augmented reality and virtual reality systems. Due to these similarities between the two technologies, the both concepts require a powerful viewing device to immerse the user into a digital world where the user is able to interact with the virtual environment. These characteristics allow both technologies to be interactive and immersive while providing information sensitivity.

1.3.4 Mixed Reality

Paul Milgram [55] defines Mixed Reality (MR) as the interval between the real environment and the virtual environment (figure 1.5). MR is composed of reality, Augmented Reality (AR), Augmented Virtuality (AV) and Virtual Reality (VR).

Figure 1.5: Milgram Reality-Virtuality Continuum

1.3.5 Virtual Environment

Virtual environment is a completely synthetic world, which may mimic the properties of some real-world environments, either existing or fictional [55], where the participant observer is able to interact with it. There are three types of environments [61]: Non-Immersive Virtual Environment (NIVE), Semi-Immersive Virtual Environment (SIVE) and Fully-Immersive Virtual Environment (FIVE).

1.4 Summary

The main goal of this chapter is to provide an overview of the user interfaces. The major terms related to the study are explained and defined. The existing interaction techniques related to the field of virtual reality and augmented reality also discussed and detailed.

In the following chapter we will focus on our main topic, the study of the natural user interface as well as the technologies used in this field.

CHAPTER 2

STATE OF THE ART

Computers serve a central role in our daily life, but most of our screens and interaction with the digital content is still done in two dimensions. Natural user interface wants to change that by providing to the user different ways of interaction to have the ability to communicate with a three dimensional space.

This chapter summarizes and introduces the concept of NUI and gives definitions of the important terms that are related to this field, and gives an overview of the different modalities that can be used to interact with the content. NUIs technologies are described as well. We finish by providing a classification of NUIs in Virtual and Augmented Reality applications.

2.1 Natural user interface

The natural user interface (NUI) is a new manner that makes the user interact with computing devices in a natural and initiative way by using different input modality [14]. NUI focused on interaction style such as tangible interaction, surface based interaction as well as gesture and voice [42].

NUIs enables users to interact with computers using natural languages and behaviours such as intuitive interactions which respond to ambient cues and intentional movements to create empathetic, personalised experiences [44].

NUI allows users to interact with the technology without an intermediary device for the user interface, these interactions happen directly using hand gestures, body movements or any other interaction modality [40].

2.2 Natural interaction modalities

A modality is defined as a couple (physical device, representational system), where the representational system is a conventional structured system of signs ensuring a communication function (e.g. a pseudo natural language) [57]. However, there is no modality that can be adapted to all interaction scenarios. For example, touch-based interaction paradigms are appropriate when playing video games on a smartphone, while it may distract users attention when driving. Each modality relies on different perceptual abilities so that they can be used for different purposes. Turk [75] classifies those modalities in four main categories: touch, auditory, visual and Other sensors.

2.2.1 Tactile interaction

Direct interaction by hand is considered to be the most natural way of interaction. Tactile interaction performed through direct input devices such as digital touchscreen or a physical augmented surface. This interaction allows the use of direct hand or fingers contact on a touchpad. Ghomi [33] identified two types of contact: the postures of the fingers or the hand (static contact), and the movements of finger and hand (dynamic contact) .

Pokemon Go, one of the most successful mobile augmented reality games of 2016 (Figure 2.1), lets a user toss virtual Poke balls to capture Pokemon, which appear in the AR scene, through the touchscreen of his/her smartphone. [2]

Figure 2.1: Augmented Reality Pokemon on the Street[1].

[2]https://www.pokemongo.com/en-us/

2.2.2 Tangible interaction

Tangible interaction enables users to interact with virtual objects by manipulating physical objects. [3].

3D-Human on a Box is a demo prototype with an educational purpose to explain the structure of the human body [82]. The physical cube allows users to explore muscles, nerves and borne of the human. The manipulation of the physical cube let the user do several task such as move, rotate and pick menu (Figure 2.2).

Figure 2.2: The human body can be manipulated by cube [82].

Let's Cook is another application using tangible interaction [62], this game teaches cognitive impairments children how to cook. Let's Cook uses cards as physical objects to manipulate virtual food (Figure 2.3).

Figure 2.3: Let's cook [62]

2.2.3 Gesture

A gesture is a sequence of postures connected by motions over a short time span [2]. Human gestures come in many forms, such as hand gestures and general body gestures. There are mainly six types of gestures [13]:

- **Symbolic gestures**: gestures with single meaning.
- **Deictic gestures**: gestures that are used for pointing or directing the users attention to specific events or objects in the environment.
- **Iconic gestures**: gestures which are used to convey information about the size, orientation of the object.
- **Pantomimic gestures**: gestures which are typically used in showing the use of movement of a tool or object.
- **Beat gestures**: gesture of hand moves up and down with the rhythm of speech and looks like it is beating time.
- **Cohesive gestures**: variations of iconic, pantomimic or deictic gestures that are used to tie together temporally separated but thematically related portions of discourse.

A gesture can deliver more natural and intuitive methods for communicating with computers or the environment[3]. When a person uses gestures to convey some information, this gesture is recognized by tracking the motion of hands, arms, face or body. Then it's interpreted and mapped to a meaningful action [40].

2.2.3.1 Freehand interaction

Freehand interaction has been explored to deliver natural, intuitive and effective interaction. There are many input devices that track hand gestures such as fiducial markers or digital gloves [36] [12], Kinect to detect hand poses and movements for freehand menu selection [23] and object manipulation [56], Image-based methods proposed to detect and recognize hand gestures in closed and public environments using image processing technology.

Figure 2.4 showing a natural designed workspace. This workspace allows The user to modify complex computer aided design models using hand gestures via leap motion sensor. The user can use several bimanual gestures for rotate, pan, zoom and explode the model in a 3D virtual environment.

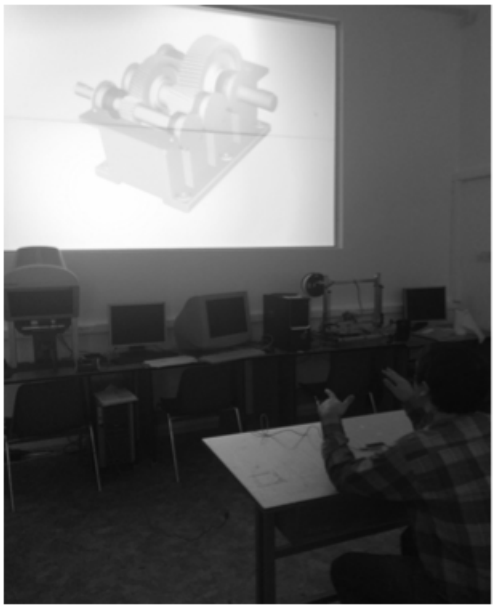

Figure 2.4: Freehand Gesture interaction [31]

2.2.3.2 Body movements

Similar to hand movements. Users interact with the system by making metaphoric gestures or moving their body parts such as head, arm, hand and feet in 3d space to select and manipulate virtual objects or to reproduce those movements through an avatar [79].

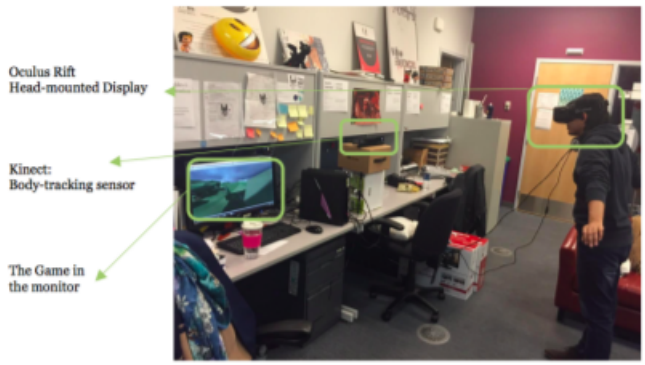

Figure 2.5: Player move their body to control their virtual movements [74].

Beyond is a Virtual Reality game [74], which allows players to control their flying movements using body gestures captured by a Kinect. Players can do multiple actions such as fly above a forest, finding and collecting five hidden items in the forest (Figure 2.5).

2.2.3.3 Sign language

Sign language is the primary way of communication for hearing impaired people. The communication is performed through the use of hand gestures, movements of arm/body, facial expressions, and lip movements. every word is represented using a meaningful gesture. [43].

Figure 2.6: Sign language recognition framework [43].

2.2.4 Facial expression

Facial expressions play a major role in how people communicate, so they make behaviour more understandable and they supplement verbal communication.

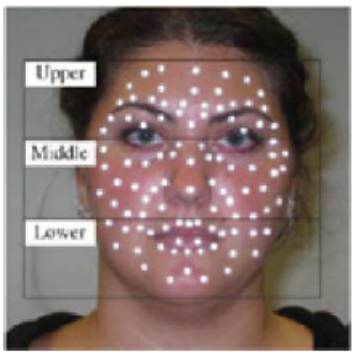

Figure 2.7: Subdivision of the facial surface in three levels, the expression is tracked using markers attached to the face [17].

Facial expressions or emotion recognition are the facial changes in response to a person's internal emotional states, intentions or social communications [2]. It can be divided into eye blinking, mouth and eyebrow movements [52] (Figure 2.7).

2.2.5 Lip movement

The lip movement is generally used to complement or to improve recognition of other modality such as speech. This movement produces silent speech that allows users to communicate and interact with the system without making sounds [73].

According to Malcangi [51] the lips are the most externally visible speech articulators that provide crucial visual feature information regarding the speech.

Lip-Interact [73] is an interaction technique that allows users to issue commands on their smartphone through silent speech. Lip-Interact is using camera and deep learning to capture and recognize lips movements (Figure 2.8).

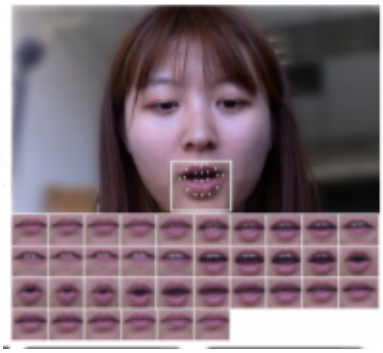

Figure 2.8: Command smartphone through the mouth movement [73].

2.2.6 Gaze

Human eyes can not only be used as visual information receptors, but their movements can act directly as a cursor in the virtual environment. It can be used for basic interactive tasks such as pointing or navigating, by tracking eye movements we obtain eye positions, gaze positions and eye movements. Eye position is the position of the physical eye and gaze position is the position on where the user is looking. So with this information, the computer can identify the types of the eye movement [24].

Some studies have exploited the use of the gaze by combining it with other modalities, to improve the accuracy or predict the future action of the user. In the work of [64] they added GazeButtons for a text editing tool on a multitouch tablet, this button allows the text cursor position to be set as users look at the position (Figure 2.9).

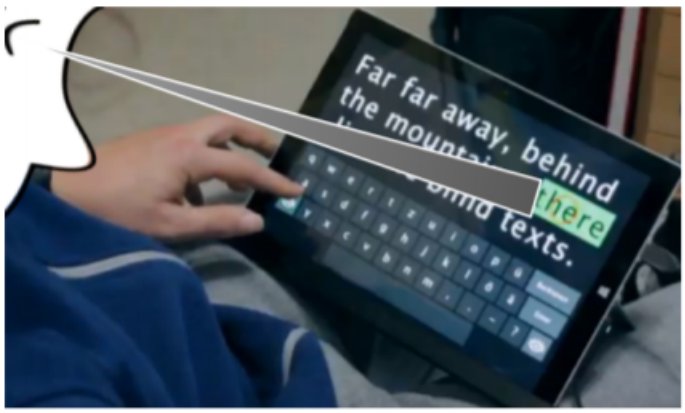

Figure 2.9: Highlighting/Selecting a word with gaze [64].

2.2.7 Voice

Voice based interaction can be for most cases trustable, helpful than visual signals [2]. A speech consists of some words spoken to interact with the computer with natural language. Users use voice commands to select and manipulate virtual objects or do precise action. Speech modality gives better user experience to the application. There are many applications that use this modality such as Located-Based augmented reality [1] which allow users to search for specific locations via voice command.

Mobile Augmented Reality game employs the voice modality [18], where the user can interact with a virtual dog by using speech commands. This application analyzes the user's speech, and allow the virtual character to perform the corresponding actions (Figure 2.10).

Figure 2.10: AR game interact by speech [18]

2.2.8 Brain activities

Brain Computer Interfaces(BCI) are communication systems conveying messages through the brain [37]. This modality allows the user to interact with the computer using only the power of thought. Several research studies prove that BCI can be used to interact with 3D virtual environments supporting navigation, selection and manipulation [8, 45, 47].

An example of BCI systems is "use the force" was inspired by a sequence in Star Wars [47]. The user must lift a virtual spaceship by imagining foot movements (Figure 2.11).

Figure 2.11: Controlling a virtual spaceship using BCI [47].

2.3 Natural multimodal user interfaces

2.3.1 Multimodal user interface

Multimodal interaction defines the manner to employ natural modes of senses such as vision, sound, touch, smell, taste and proprioception, both sequentially and in parallel to communicate with the environment [75].

Multimodality is the combination of two or more input stream modality in coordination manner, to cover the weakness of one modality by another or to improve the quality of interaction to satisfy the new communication needs [26].

2.3.2 Multimodal integration

To create a multimodal natural interface we need to combine several modalities in a well coordinatied manner, but this modalities do not all have similar characteristic: [75]

- *Some modalities provide information at discrete points in time, while others generate continuous but less time-specific.*
- *Some modal combined are intended to be interpreted in parallel, which others may typically be offered Sequentially.*

To ensure the availability of the combined modalities several frameworks are proposed such as TYCOON, CARE and CASE Properties.

2.3.2.1 TYCOON

TYCOON is a theoretical framework created by Martin [53] they conceptualize the different possible relationships between input and output modalities of a multimodal interface system. TYCOON is the abbreviation of TYpes and goals of COOperatioN between modalities. In this framework there's five types of cooperation between the modalities :

- **Equivalence:** describes the choice between several modalities having equal capability in processing the information.
- **Specialization:** a given modality is always used for processing specific information.
- **Redundancy:** more than one modality processed the same information.
- **Complementarity:** describes the case that needs to merge several modalities to reach a given stats.
- **Transfer:** the information or interaction produced by a modality is used by another modality.

2.3.2.2 CARE Properties

It's proposed by Coutaz et al [22] in 1994, this model contains the four properties that can be occurred between several modalities:

- **Equivalence:** possibility to choose between several modalities to reach a given state.
- **Assignment:** means that we need to use only one modality to reach a given state.
- **Redundancy:** several modalities can reach the given state.
- **Complementarity:** means that we need to use all the modalities of the system to reach a given state.

2.3.2.3 CASE Properties

Fusion of multiple modalities can be different from one application to another, for this reason Nigay and Coutaz [7] classified these multimodal interfaces depending on the fusion method and the use of modalities (table2.1). CASE is the abbreviation of concurrent, alternate, synergistic and exclusive. This model concentrates on the temporal availability of modalities.

		Use of modalities	
		Sequential	Parallel
Fusion	Combined	Alternative	Synergistic
	Independent	Exclusive	Concurrent

Table 2.1: A classification of multimodal interfaces systems[7].

According to Turk [75] we have four Multimodal interfaces types:

- **Exclusive multimodal system**: the modalities are used sequentially.
- **Alternative multimodal system**: modalities are used sequentially and they are Integrated.
- **Synergistic multimodal system**: modalities are used in parallel and they are fully integrated.
- **Concurrent multimodal system**: modal information is available in parallel and not integrated by the system.

The difference between CASE and CARE, is the CASE model focusing on modality combination at the fusion engine level while the other one at the user level [27].

2.3.3 Multimodal Fusion Level

The goal of fusion is to extract meaning from a set of input stream modalities and pass it to a system. Fusion of different modalities is a delicate task, which can be executed at three levels: [27]

- **Data-level fusion:** is used when data coming from a similar modality, this signal of data is directly processed.
- **Feature-level fusion:** is used when we have data coming from closely coupled modalities. Different architectures used in this type of fusion such as artificial neural networks, gaussian mixture models or hidden Markov models.

- **Decision-level fusion:** is used when we have loosely coupled modalities, this level uses frame(attribute-value pairs) to represent the data coming from various modalities, and recursively merging these structures to obtain a logical meaning. This fusion has the major benefit of improving reliability and accuracy of semantic interpretation.

2.3.4 Machine Learning for Multimodal Interaction

Modality recognizers make extensive use of machine learning such as speech recognition, face detection, face recognition, facial expression analysis, gesture recognition or eye tracking. Aside from modality handling, machine learning has been applied for fusion of input recognizers' data, at the feature level and decision level [27].

2.3.5 Collaborative and Multimodal Natural user interface

Collaborative interfaces allow multiple users to interact in a virtual environment [20]. By sharing the replicated virtual space, all users can work together from anywhere in the world.

An example of collaborative interfaces is TinkerLamp (Figure 2.12). TinkerLamp is a tabletop learning environment for logistics apprentices that allows users to quickly build and evaluate small scale warehouses. TinkerLamp supports a multi-touch and tangible interface.

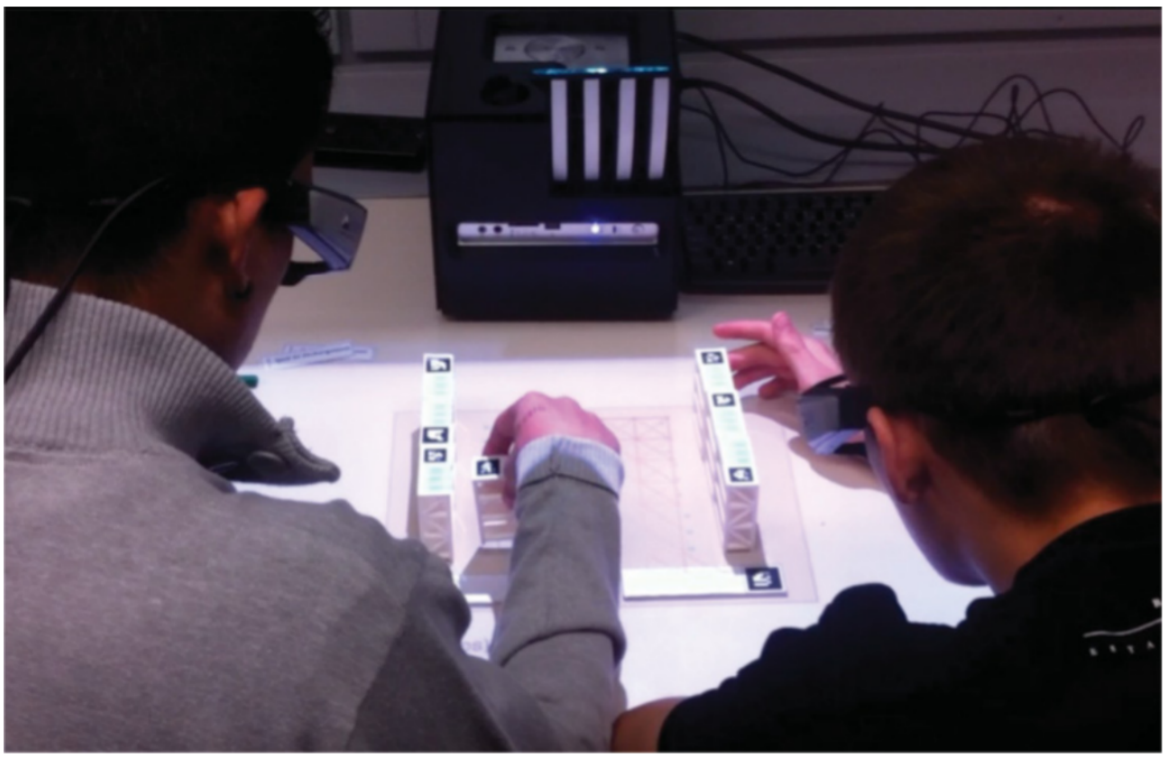

Figure 2.12: Two Students learning to solve optimization problems with the TinkerLamp [25].

An example of collaborative natural interface is puzzle game on a multi-touch tabletop (Figure 2.13). In this game each puzzle piece is represented as a cube, with a picture on one of its sides. However, this game sometimes requires two users to collaborate tightly. Each user is assigned a colored disk, so not all pieces behave in the same way [77].

Figure 2.13: Player interacts with a puzzle piece while her partner reads the help[77].

Collaborative virtual environments can be classified as follows [39]:

- The environments in which users perceive the co-presence through their avatars, but each user can independently interact with objects. Any change to the attribute is visible to all collaborators.

- The environments which allow multiple users to interact, but not all of them at the same time, the users need to wait their turn to manipulate the object.

- The environments in which multiple users can manipulate the same object.

2.4 Natural user interfaces I/O technologies

Term NUI stands for the ways of interaction with a device based on methods other than a mouse and a keyboard. This technology must provide a natural and intuitive interaction for a human.

2.4.1 Display devices

A physical device that serves as the medium of communication between the human and system. Display devices are able to present information to one or more of the user's senses such as visual or auditory information.

2.4.1.1 Visual display

Visual displays are the most common displays used in the 3D user interface. This device converts data and digital content into visual information that can be understood by the human visual system. They are commonly used today in smartphones, tablets, head-mounted display and monitors.

2.4.1.2 Auditory display

Auditory displays have great impact in increasing the filling of presence in the virtual environment. This device generates sound to enable the user to use his auditory localization capabilities, by searching the location and direction from which a sound is occured. Two main approaches are used for displaying the sounds signals: headphones and speakers.

2.4.1.3 Haptic display

Haptic displays provide the user the feeling of touch and force by simulating the physical interaction between the user and the virtual objects. This device is generally coupled with an input device to provide a fast feedback between the force used by the user and haptic feedback transmitted back to the user.

2.4.2 Sensor devices

These devices are used as sensors to capture user input such as voice, touch or movement. Currently there's five types of sensors[67]: touch surface, microphone, camera, motion sensors and brain-computer interface.

2.4.2.1 Touch surface

Touchscreen surfaces have capacitive or resistive sensors[67]. They are used to capture touch gestures such as fingertip or hand motion on a 2-dimensional(2D) surface. The coordinates of the

touch is processed by a touch screen controller. They are commonly used today in smartphones, tablets.

2.4.2.2 Microphone

A microphone captures voices of users and allows free-air interaction with a device. The data captured is recognized through voice recognition algorithms. These algorithms understand a wide range of vocabularies, languages and accents. A microphone creates communication channels between users and devices such as smartwatches, tablets and smart-home units.

2.4.2.3 Camera

A camera can serve as a sensor to capture either an image or a video representing a variety of input stream modalities, free-hand gesture, lip movement, body movement or gaze direction. Those information are used as an interface to execute various tasks including navigation as well as selection and manipulation. There's another type of camera called a depth or 3-dimensional(3D) camera, this device is used as a 3D control interface. a 3D sensing camera can capture precise volumetric information from gestural interaction. Such devices as:

- **Leap Motion:** a controller device with two infrared cameras and three infrared sensors (Figure 2.14) [8]. They allow tracking the movements and positions of the all 10 human fingers simultaneously in space above the device itself, which gives the user the opportunity to move his hand in 3D virtual space.

Figure 2.14: Leap Motion Controller [3].

- **Stereolabs ZED:** is a binocular vision system used to provide a 3D perception of the world [60]. The ZED camera is not using an infrared sensor to measure depth. Instead, it's mimicking the way the human eyes work, it has two high-resolution cameras that capture images at the same time and transmit. This camera perceives the depth of objects in opened and closed environments (Figure 2.15).

[3] https://www.ultraleap.com/product/leap-motion-controller/

Figure 2.15: Stereolabs ZED camera [4].

- **Kinect:** a depth sensor device is able to recognize and track the movement of a many people at the same time as well as recognize their voice commands. Kinect also provides the information about the distance between the user and the device by infrared projector. The information about the depth is determined on the basis of stereoscopic triangulation [34]. The new version of kinect is Azure Kinect (Figure 2.16) . This new device has a DK camera system and new sensor SDK, body tracking SDK, vision APIs, speech service SDK.

Figure 2.16: Microsoft azure Kinect[5].

- **VicoVR:** is a Wi-Fi accessory that provides wireless full body and positional tracking to Android and iOS ecosystems - without a PC, wires, or wearable sensors [50]. This sensor uses depth sensing technology and the body tracking SDK. The difference between kinect and VicoVR is that the data captured is executed inside the sensor.

[4]https://www.stereolabs.com/zed/
[5]https://azure.microsoft.com/en-us/services/kinect-dk/

Figure 2.17: VicoVr camera [6].

2.4.2.4 Motion sensor

Accelerometers and gyroscopes are considered as Motion sensors. These sensors can be used together to determine the orientation, movement and position of a device in 3-D space. An accelerometer is deployed to detect linear acceleration of the device, while a gyroscope is designed to detect its rotation around a fixed axis. However, the two sensor types are embedded into various devices such as :

- **Sony Move**: is a controller based on 3 gyroscopes and 3 accelerometers which allows for determining the inclination towards each axis as well as the movement [34]. The controller communicates with PlayStation Eye, it is camera able to detect the location of the controller and user as well as distance between the player and the device (figure 2.18).

Figure 2.18: Sony Move Controller [7].

- **Nintendo Wii Remote**: Wii Remote controller is a wireless device. This controller is equipped with an accelerometer that allows it to determine movement in three planes x, y and z as well as rotation [34]. Wii Remote communicates with Sensor Bar. The Sensor Bar contains a camera with infrared filter, which is able to determine the distance between the Sensor Bar and the user (Figure 2.19).

[6] https://vicovr.com/
[7] https://www.playstation.com/en-us/explore/accessories/vr-accessories/playstation-move/

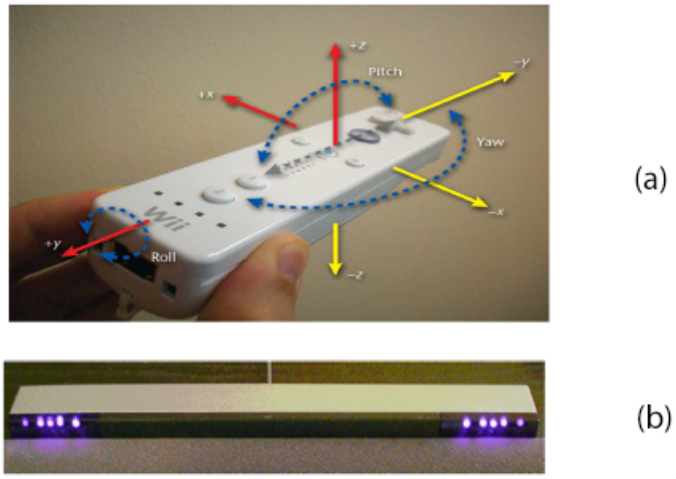

Figure 2.19: Nintendo Wii Remote(b) and Sensor Bar(a) [34].

2.4.2.5 Brain-computer interface

Brain-computer interface does not require any physical movement to interact with a machine. BCI is considered as communication channel between a user and computer. This form of communication uses electrical signals of the brain's activity. The electroencephalography (EEG) signal is a multi-dimensional waveform captured by a set of electrodes operating.

An example of BCI technology is Emotiv EPOC+ (Figure 2.20), is a multi-channel electroencephalography device for research. The EPOC + is a wireless headset that records 14-channel EEG, the measured data is transmitted through Bluetooth. This device uses saline based wet sensors unlike other EEG systems that use sticky gels.

Figure 2.20: Emotiv EPOC+ NeuroHeadset [8].

[8]https://www.emotiv.com/epoc/

2.4.3 Hybrid devices

Hybrid technology is the use of more than one sensing and displaying technology together to provide a better overall 3D interaction experience. The use of different sensing technologies enables the device to attain more degree of freedom (DOF) and increases the tracking effectiveness [46].

Sensor fusion algorithms are often required to combine the different sensors together to ensure that each sensor type is used properly [46].

HoloLens

HoloLens is an augmented reality device created by microsoft (Figure 2.21). HoloLens is a head-mounted wearable computer. This device has the capability to integrate the virtual and real world seamlessly together [35]. This holographic computer supports different kinds of modalities input such as voice, gaze and hand gestures.

Figure 2.21: The Microsoft HoloLens [9].

The Microsoft HoloLens use a variety of sensors: four tracking cameras and depth camera for spatial mapping (SLAM), as well as accelerometers and gyroscopes working together for environment understanding [19]. SLAM is an algorithm that uses computer vision to compare data with gyroscopes and accelerometers.

Hololens is an audiovisual display device as well. It has two small projector located at the bridge of the nose that have the capability to make the distributed light ray appear as a real object in the point of view of the eye. 3D audio speakers are also used to support fully immersive 3D audiovisual effects [19].

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

Oculus Rift

Oculus Rift is a virtual reality headset created by Oculus VR. The HMD is equipped with two screens and integrated headphones to provide audiovisual information. Oculus Rift uses an external camera to track IR light of the headset and the two hand controller [29]. This IR tracking system is able to identify the position of the user head. The controller has five buttons that offer a touch interface (Figure 2.22).

Figure 2.22: Oculus Rift [10].

PlayStation VR

PlayStation VR is a virtual reality headset developed by Sony. PSVR uses a combination of the PlayStation Camera, LEDs and accelerometers/gyroscopes. The PlayStation Camera tracks the visual light and the motion of the headset, as well as the orbs light on the controllers. This data is translated into game movement by the console. The gyroscopes and/or accelerometers are also used to track the precise movement (Figure 2.23).

Figure 2.23: PlayStation VR [11].

[10] https://www.oculus.com/rift/
[11] https://www.playstation.com/en-us/explore/playstation-vr/

2.5 Natural user interface in virtual and augmented reality

2.5.1 Immersive analytics

The availability of high-quality immersive hardware and software technology led to the rise of system proposals and scientific contributions in the domain of immersive analytics. Immersive analytics (IA) is a new term referring to the use of immersive technologies to support data understanding and decision making (Figure 2.24) [32].

Figure 2.24: Distribution of a visual analytics system across different display geometries [16].

A.Fonnet et al. describe in his survey all the immersive technologies focus on its fidelity aspect, data representations and sensory mapping, and low-level interaction modalities that have been implemented in IA systems over the last three decades, as well as how these systems have been evaluated [32].

A.Fonnet et al. also mention that the types of interactions are limited in the current IA systems, so that may be a challenge for the future to create a new interaction paradigm to the IA field [32].

2.5.2 Immersive VR for Education

Immersive VR has the ability to maintain an illusion of presence, such that users or learners feel their bodies are inside the virtual environment, that can have positive effects on attention

and enhance student learning and engagement [38].

Johnson,et al have explained the two affordances associated with VR for educational purposes: the sensation of presence and the embodied affordances of gesture in a 3D learning space, and introduce new graphic cube to visualize the amount of embodiment in immersive educational lessons [38].

The Google Expeditions series is one of the Immersive VR systems that have been used in education. It is used to help students visualise information in a new way or explore history, science, the arts and the natural world. This experience will have a great impact in the ability of the student to retain the information (Figure 2.25).

Figure 2.25: The Google Expeditions series [12].

2.5.3 Olfactory Interfaces

The sense of smell is the most complex and challenging human senses, that is why it is considered as one of the least appreciated systems in the world of technologies. Compared to other modalities, The olfactory has direct connections to the two brain areas that control emotions and memories[4].

Smell plays an important role for enhancing the sense of immersion to augment digital experiences and creates a more complete sensorial experience[58].

[12]https://edu.google.com/products/vr-ar/expeditions/

"Essence" is a wearable olfactory display that can be controlled through a smartphone to influence the emotion of the user [58] (Figure 2.26). This portable device also can change the smell automatically based on biometric or contextual data of the user. Essence enhances user experience in virtual and augmented reality environments and can be used as a way of communication among remote users.

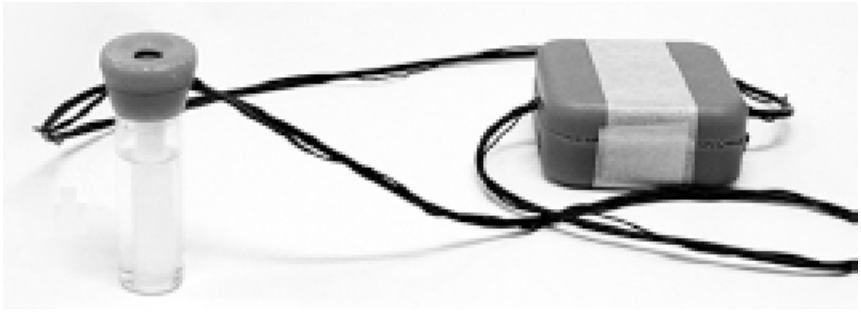

Figure 2.26: The Essence necklace [4].

2.5.4 Identify the common characteristics of NUIs

Natural interface aims to enhance the way of interaction of the user in a Virtual and augmented reality environment based on real-world experiences. For that the interface must meet a certain set of requirements and conditions:

(C1) Understand natural command: is the system's ability to understand verbal or nonverbal communication of the user. Affirm that the interaction is based on previous user knowledge.

(C2) Easy to use: the interface must reuse existing simple skills of the users, to make the interaction with the content more natural and intuitive.

(C3) Easy to learn: the interface must allow your users to learn and evolve progressively, from beginners to expert level, even if some of the skills are completely new.

(C4) Direct interaction: it is the possibility of the user to have direct interaction with the technology or content without intermediary device.

(C5) Transparency: the virtual environment should appear as an extension of the real world.

(C6) Flexibility: is the case where the interface allows simultaneous interaction of a group of users in public environments, the functioning of the system must not be impaired when used by a huge number of users.

(C7) Real-Time interaction: the system should immediately give feedback to the user like natural communication between humans.

(C8) Limited use: is the system has limited utility and requires other modalities to perform 3D interaction tasks in virtual, augmented and mixed reality environments.

2.5.5 Classification according to common characteristics identified

This classification show the satisfaction of each characteristics previously identified, by the different modalities of natural interaction in virtual, augmented and mixed reality applications (table 2.2). Different publications are used to cover all the the different modalities (tangible and tactile interaction [6, 10, 11, 21, 62, 68, 71, 82], freehand gesture and body movement [31, 54, 59, 74, 81, 82], facial expression [2, 17], lip movement[73], gaze [41, 64, 71, 82], sign language [43, 69], voice [18, 82], brain activities [47–49, 59]).

			C1	C2	C3	C4	C5	C6	C7	C8
Touch	Tactile		Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Visual	Tangible		Yes	Yes	Yes	Yes	Yes	No	Yes	No
	Gesture	Freehand	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
		Body movements	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
		Sign language	Yes	No	No	Yes	Yes	Yes	Yes	Yes
	Lip movement		Yes	No	Yes	Yes	Yes	No	Yes	Yes
	Gaze		Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	Facial expression		No	No	No	No	No	No	No	Yes
Auditory	Voice		Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Other Sensor	Brain activities		Yes	Yes	Yes	Yes	Yes	No	Yes	No

Table 2.2: Classification of natural interaction modalities according to common characteristics.

2.6 Summary

In this chapter, we have covered definitions of several terms and concepts which are relative to our research topic. We have presented different definitions of natural user interfaces and several potential modalities that can be used to develop these interfaces. The NUIs input/output technologies were also discussed. Then, we have presented several studies that have explored how to use these modalities independently or in collaboration to develop multimodal user interface. Finishing by proposing a classification of NUIs for VR and AR application according to common characteristics identified.

In the next chapter, we will present a proposed design model of NUIs for Virtual and Augmented Reality (VR / AR) applications.

CHAPTER 3

PROPOSE A CONCEPTUAL MODEL OF NUIS FOR VR/AR APPLICATIONS

In this chapter, we will propose and describe a conceptual model of natural user interfaces for virtual and augmented reality applications. Different layers of the model are detailed and explained.

3.1 The structure of the conceptual model

3.1.1 Problematic and objectives

As the technology moves forward, the work on natural user interfaces for virtual reality and augmented reality application is becoming more and more widespread without leaving room for design work. This rapid asynchronous progress makes the void larger between the designers and technology. Thus, the designers of natural interaction techniques are facing many problems:

- How to increase the ease of interaction with the user?
- What evaluation techniques should be used?
- how to transform metaphorical interfaces to direct and intuitive interfaces?

The main objective of this research is to propose a conceptual model of NUIs for virtual and augmented reality (VR / AR) applications based on the state of the art of the previous chapter.

This model summarizes essential elements for the realization of a 3D interaction technique, in order to ensure a natural and intuitive interaction between human and system. Finishing by evaluating the interaction model based on the characteristics specific of NUIs.

3.1.2 Global approach for the construction of the model

3.1.3 The overall structure of the model

We propose a model that contains seven distinct layers that can ensure natural and intuitive interaction when integrating a natural interface in VR / AR applications (Figure 3.1):

1. Choose the technology.
2. Specify the environment.
3. Define user characteristics.
4. Choose 3D interaction tasks.
5. Identify natural interaction modalities.
6. Create interaction model.
7. Apply evaluation model.

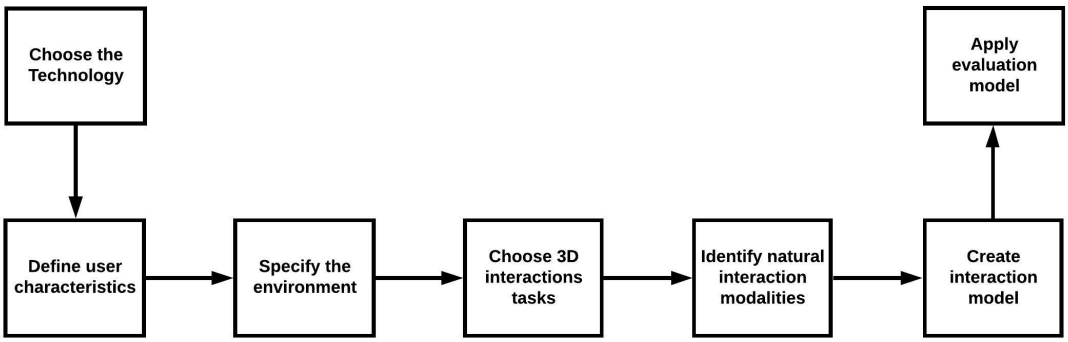

Figure 3.1: Design model of NUIs for AR/VR applications.

3.1.3.1 Choose technology

This layer focuses on the technologies since multiple kinds of technology offer different DOF control inside the application. Mainly, there are three categories: the sensor is the first category. Those devices are used to capture the user input such as voice, gestures, touch, brain activity. While the second is hardware that is used to display (auditory, visual or haptic). The last device category is the hybrid technology, which is used as input and output devices at the same time. (Figure 3.2).

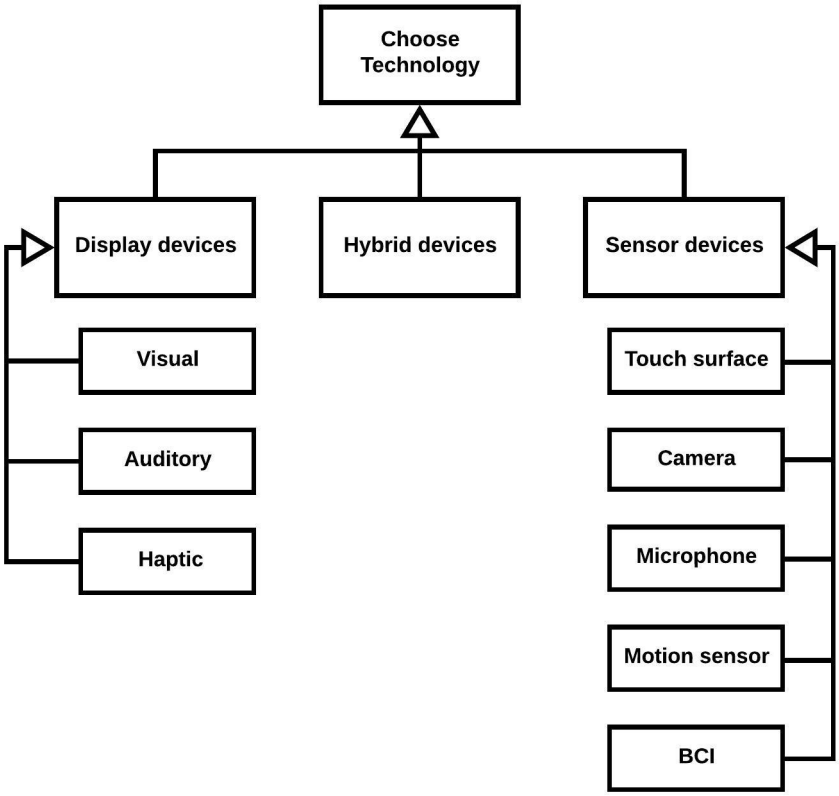

Figure 3.2: Class diagram represents NUIs technologies.

3.1.3.2 Specify the environment

The environment can change definitely the way of interaction between user and application (Figure 3.3):

- **Virtual environment:** in this virtual space, users can do several tasks such as navigation, selection and manipulation. To ensure this interaction we must specify the type (NIVE/SIVE/FIVE) and characteristics (collaborative/individual) of the virtual environment.
- **Real environment:** real space has an impact on the input received by the user such as noise. Those added informations may cause misinterpretation of the data, for this reason, we must specify if the application requires an opened or closed environment.

3.1.3.3 Define user characteristics

The information about targeted users are necessary to enable the evaluation of the developed system (Figure 3.4). This layer focus on the group of users who will actually be using the

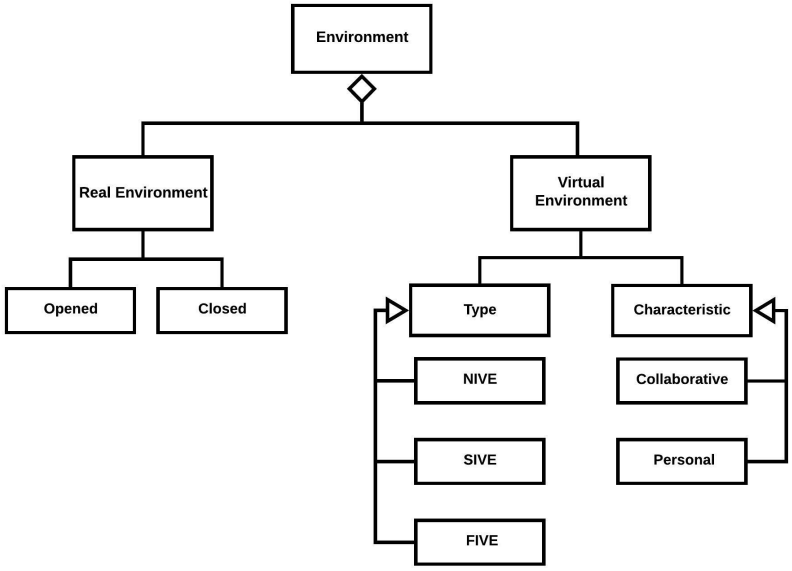

Figure 3.3: Class diagram represents environment.

application and their capabilities:

- **User skills:** it is about his mental and physical abilities that allow users to do several activities.

- **User categories:** is the age range of users that are targeted for the applications.

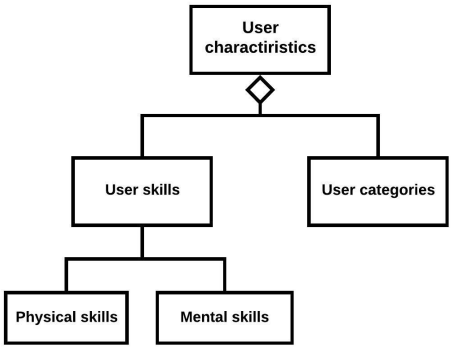

Figure 3.4: Class diagram represents user characteristics.

3.1.3.4 Identify 3D interaction tasks

3D interaction tasks represent fundamental tasks (Figure 3.5) that can be combined to create more complex interaction tasks. This tasks related to the objective of application (e.g. navigate in a virtual 3D surface while selecting or manipulating virtual objects). Our state of the art presents that 3D interactions tasks are not achievable with all modalities (e.g. facial expressions and navigation).

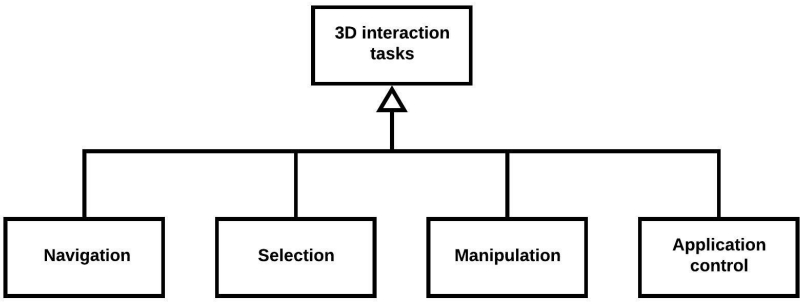

Figure 3.5: Class diagram represents 3D universal interaction tasks.

3.1.3.5 Choose natural interaction modalities

To keep 3D interaction tasks possible to perform through the chosen modalities. We must choose judiciously the appropriate modalities based on the four previous layers especially the information of the target user. The future application requires physical capabilities of users to interact with content.

3.1.3.6 Create interaction model

Interaction model is a set of composed tasks that are varied from simple to complex (Figure 3.7). Those tasks are performed through the use of 3D interaction tasks with interaction modalities. These tasks are used to interact and communicate with the virtual environment. The simple tasks can be performed through one modality, but complex tasks require several subtasks to be finished. Those subtasks may use one or several modalities, and two or more 3D interaction tasks to achieve the goal of the application.

Interaction model is different from an application to another, it can be defined as a combination of 3D interaction tasks and modalities. This combination creates a new way of communication that satisfies user needs.

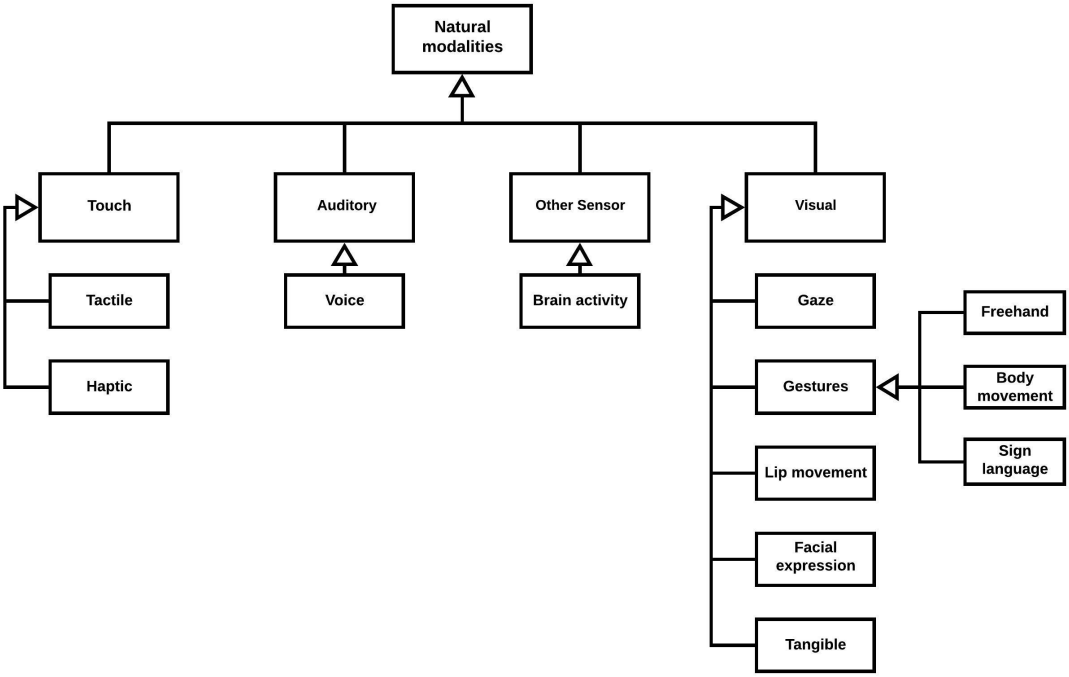

Figure 3.6: Class diagram represents 3D interaction tasks.

In order to create a good interaction model, the designer can use the guidelines of other designers whose results have been published and validated (Figure 3.7). These guidelines can be used to provide further insights. The designer guidelines are varied depending on the technologies, the modalities and the field of the future system.

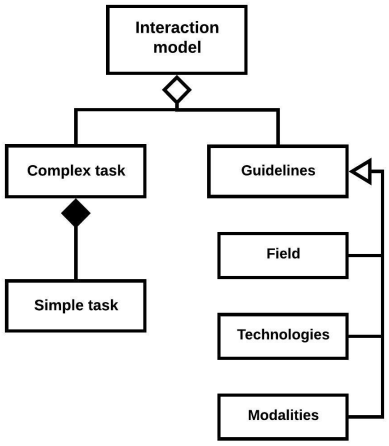

Figure 3.7: Class diagram represents interaction model.

3.1.3.7 Apply evaluation model

This layer focuses on how we can evaluate and ensure that the 3D interaction technique created in the interaction model of the previous layer is natural. Many evaluation metrics have been proposed to evaluate the technique: First, the interaction technique will evaluate according to common characteristics specific to NUIs proposed in chapter one, namely: Understand natural command, Easy to use, Easy to learn, Direct interaction, Transparency, Flexibility, Real-time interaction and limited use (Figure 3.8).

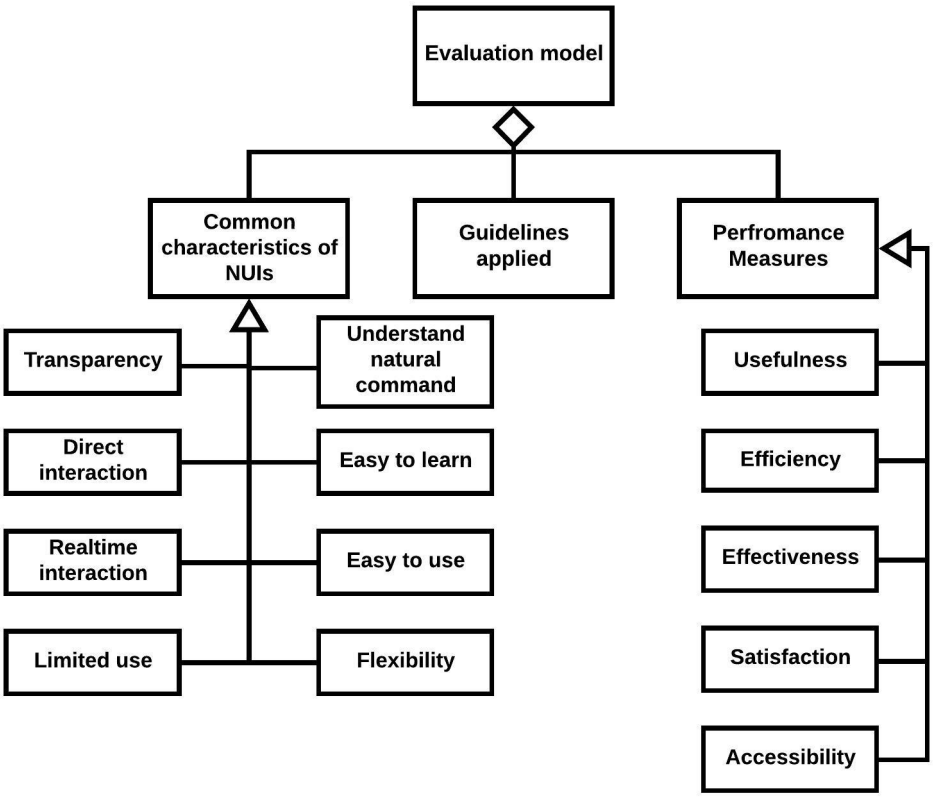

Figure 3.8: Class diagram represents evaluation model.

The interaction techniques will then be evaluated according to the guidelines design that are used in the interaction model to ensure that this technique is the optimal choice for the model and the application. This part of evaluation helps the designer to accept or reject the designed combinations in the model. This evaluation will then be able to provide new design ideas to cover the holes that can appear between the combinations.

Finally, The performance of the techniques will be evaluated to measure the usability of

the interaction technique according to four aspects. These aspects are part of the definition of usability provided by Rubin & Chisnell: [66]

- **Usefulness:** concerns the degree of which the user can complete the task the application is intended to solve or achieve. The system will not be used if it does not achieve the specific goals, Regardless of how easy to use and learn, and even satisfying the user's needs. Interestingly enough, usefulness is probably the element that is most often overlooked during experiments, but it is an obvious factor to consider and make sure that the users actually can achieve their specific goal with the product.

- **Efficiency:** is how quickly a user can finish accurately and completely task, commonly is determined by measuring the time taken by the user to accomplish a given task. Efficiency can be considered as the level of mastery that the user has achieved. The measure of the level of expertise of the user is important to display the complexity of the system usability.

- **Effectiveness:** is "*the extent to which the product behaves in the way that users expect it to and the ease with which users can use it to do what they intend*". Commonly, the quantitative errors rate of some percentage of total users is used to measure a product's effectiveness. As an example "60 percent of all users push a button on a user interface to save their current work, but instead it deletes it, the system has clearly behaved unexpectedly from the user's point of view. Effectiveness is defined also as the number of the errors in the system. An effective system should not have catastrophic errors, and the errors rate should be low and easy to recover. This kind of error must only slow the process of completing the task and not stopping the system.

- **Satisfaction:** is the users subjective opinions and feeling to the system, if they like it or not. The satisfaction of the system can be determined by averaging the answers of many subjective users. Usually this feedback is captured through interviews or surveys. Users are asked to rate the application that they try, and this can often reveal causes and reasons for problems that occur. satisfaction is differ from a system to another depending on the objective of the application. In a leisure system, the user can be satisfied from the experience offered by the system and this satisfaction doesn't require an effective system and easy to learn. In contrast, the labor system requires the opposite, since the user focuses on accomplishing the task in specific time. The satisfaction of the user can be measured by other alternatives such as measuring the blood pressure, heart rate, Electroencephalography in case of BCI system or the emotion of the user face while using the system to accomplish tasks.

- **Accessibility:** is the ability to have access to the system to accomplish the tasks needed. The most important thing of this term is how to make people with disabilities or who are facing temporary limitations accomplishing their goals. The designer eliminates the design

problems and minimizes disappointment for users. This hard work may have small effects on the user experience while using the system.

3.2 Summary

In this chapter, we have presented our conceptual model of NUI for VR and AR application, this model will help and simplify the decision-making for designers to choose the appropriate technique and tools, with these formalisms, the designer could also provide solutions for their specific problems.

In the next chapter, we will present our multimodal virtual reality application, the tools and frameworks are mentioned as well.

CHAPTER 4

IMPLEMENTATION OF MULTIMODAL VIRTUAL REALITY EDUCATIONAL APPLICATION

In order to come up with a new 3D multimodal interaction paradigm of NUI for VR and AR based on the proposed conceptual model of the previous section, we built an Multimodal Virtual Reality educational game, in which a user or a learner can navigate or select and manipulate objects in three dimensional space. This immersive experience afforded by the 3D technology gives a learner unprecedented personal control over the learning environment.

4.1 System overview

The learner listens to the sound that speaks the word and tries to form that word with cubes with alphabetic characters in their faces Scattered in the ground. The user can use several modalities to interact with the virtual learning environment such as hand gestures or speech for select and manipulate the cubes, as well as using his head or hand for navigating throw virtual space.

4.2 System design

From our first chapter, we have found the assets and liabilities of the NUI for existing systems. In our proposed one, we have taken some of the general guidelines provided by Jonson to build immersive VR educational applications. Using those guidelines and our proposed conceptual model of NUIs for VR/AR application of section 3.1, we have designed our proposed system. This section deals with the system's design.

4.2.1 System Functionality

Below Figure 4.1 shows the Use case diagram for our Natural interface application , we can observe that the user performs operations by moving, closing or opening his hands facing the camera, and can change his point of view by moving his head. The user also has the possibility to control the interface by using his voice commands.

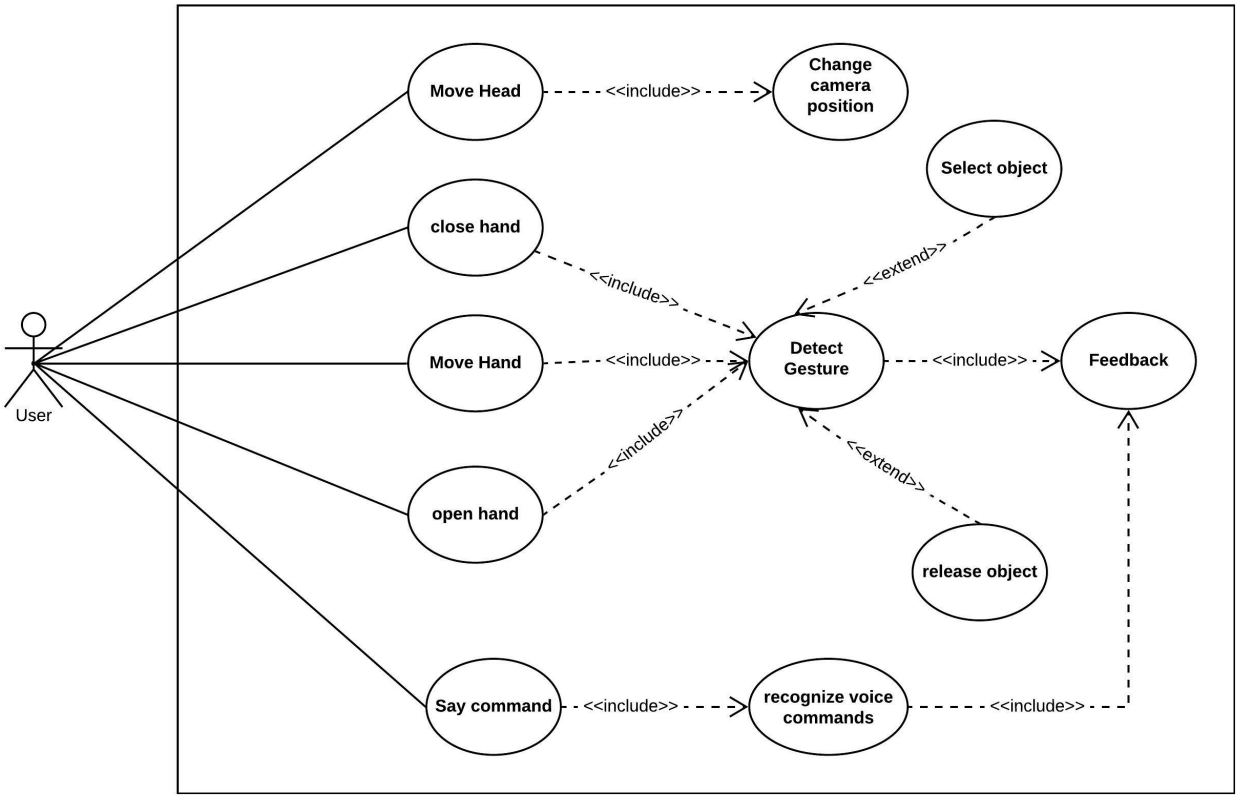

Figure 4.1: Use case diagram for our multimodal virtual reality application.

4.2.1.1 Functional requirements

- **Move Head:** user moves his head to change his point of view within the scene.
- **Change camera position:** the position of the virtual camera will change according to the user's head position in the real space.
- **Close hand:** user closes his hand to select and activate the operation of grasping the virtual object.
- **Move hand:** user moves his hand to move the virtual hand around the virtual space.

- **Open hand:** user opens his hand to allow the virtual hand to release the virtual object.
- **Detect gesture:** once the user starts to move his hands using it also performs gesture recognition. If the gesture is recognized, then the user can do further operations.
- **Select object:** once the user hand gesture is recognized, the designed operation will be performed. In case the hands are not collided with any object the operation will not be activated.
- **Release object:** to activate this event, the user should hold the object on his hands before the gesture is recognized by the system.
- **Feedback:** it is a way of representing the activation of any event, example, selecting and realising virtual objects, allowing virtual hand to follow real hand, clicking buttons.
- **Say command:** user annonces some voice commands to control the interface.
- **Recognize voice commands:** once the user starts speaking, the system starts recording the command and activates phrase recognizers. If the command is recognized, then the user can do further operations.

4.2.1.2 Non-functional requirements

- Computer must have USB 3.0 input slot to run ““Kinect ””.
- The device should be placed away from the user with 1.2m.

4.3 Implementation frameworks and tools

In order to come up with a new 3D multimodal interaction paradigme of NUI for VR and AR, a personal computer with Intel (R) Core(TM) i7-4600U. CPU 2.70 GHz and 8GB of RAM is used to build the application with Unity engine under Windows 10 and kinect V2.

4.3.1 Kinect

Kinect is motion sensing input devices produced by Microsoft. Kinect sensor includes an RGB camera, infrared projectors and detectors, and a multi-array microphone. This device is able to perform real-time gesture recognition, speech recognition and body skeletal detection.

This device was originally created for gaming purposes (xbox 360 and xbox one), but nowadays this technology is applying to real world applications (Azure Kinect 2019). The kinect used in our application is kinect xbox one.

4.3.2 Unity

The virtual learning environment is built in Unity 2019.4.0f1 (64 bit). Unity is a cross platform game engine developed by Unity Technologies. This engine is used Create 3D, 2D VR & AR visualizations for Games, Auto, Transportation, Film, Animation. Unity use c# and javascript for scripting. The code is written in our VR game entirely in C# based on Microsoft's Net 4.7.1 API level.

4.3.3 C sharp

C# is an object-oriented programming language developed by Microsoft that enables developers to build a variety of secure and robust applications that run on the .NET Framework such as web applications, desktop applications, mobile applications, games ...etc. The most recent version of c# is 8.0, which was released in 2019 alongside Visual Studio 2019.

4.3.4 .NET Framework

.NET is a software framework developed by Microsoft that supports building and running Windows applications and web services. This framework is designed to provide a code-execution environment that eliminates the performance problems of scripted or interpreted environments.

4.3.5 Visual Studio

Microsoft Visual Studio is an integrated development environment (IDE) created by Microsoft. It is used to develop a variety of applications. Visual Studio supports 36 different programming languages. The most recent version of visual studio is visual studio 2019 version 16.3.

4.3.6 Kinect for Windows SDK 2.0

The Kinect for Windows Software Development Kit (SDK) 2.0 enables developers to create applications that support gesture and voice recognition, using Kinect sensor technology on Windows. This sdk also includes a few Code samples.

4.3.7 Kinect for Windows Runtime 2.2.1811

The Kinect for Windows Runtime[1] provides the drivers and runtime environment required by Kinect for Windows applications using Kinect sensor technology. So it is used to provide for the purpose of the pre-installing Kinect 2.0 support on system images.

[1] https://www.microsoft.com/en-us/download/details.aspx?id=57578WT.mc$_id = rss_windows_allproducts$

4.4 Determination of user characteristic

According to Johnson's guidelines for building effective educational VR games, it is necessary to make the content level match the user's skill. Thus, to achieve high efficacy the learner must be greater than 10 years old and have basic knowledge of how to pronounce alphabetic letters.

In addition, active learning has been shown to increase STEM[2] grades by up to 20%. So the learner must have the ability to move its hand and head to maintain a control over the learning environment.

4.5 Identification of 3D interaction tasks

In our application, the four 3D interaction tasks are defined:

- **Navigation:** the user flies around the virtual environment and changes his point of view by moving his hands or head. The view in our virtual learning environment is under the real-time control of the user.
- **Selection:** before manipulating the 3D cube in a virtual environment, the virtual object must be designated and selected. two methods are implemented to select cubes: by gestures or speech command.
- **Manipulation:** two operations are performed through the virtual environment : moving the position or changing the orientation of the 3D object.
- **Application control:** hand gestures and voice control are used as a control system. The last one is very useful when the hands are already busy with something else.

4.6 Interaction modalities

The interaction tasks identified in our 3D multimodal interface are the ones achievable by voice, hand gestures and head movement (Figure 4.2). These modalities can be used sequentially or in parallel, depending on the physical ability of the user(if he can move his head and hands or not) .

4.6.1 Hand gesture based Interaction

As we mentioned earlier in the state of the art, there are six types of gestures. Here, we focus on symbolic gestures, the virtual hands avatar in our application are designed to follow learner hands in order to navigate in the virtual space or perform the corresponding actions (grab or

[2]STEM: abbreviation of Science, Technology, Engineering, and Mathematics.

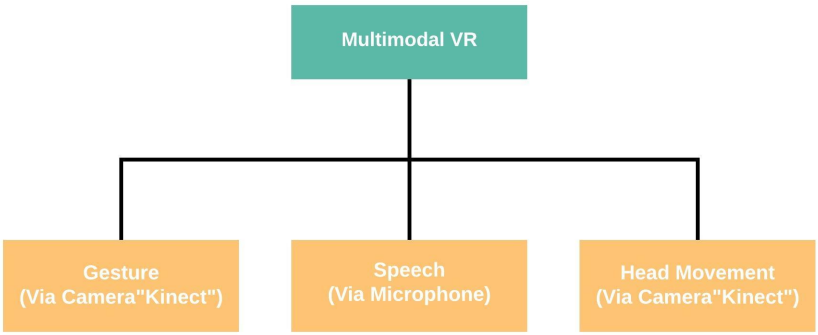

Figure 4.2: Overall structure of our multimodal VR application. Our VR application is focused on gesture, speech and head movement.

release virtual objects).

Two gesture-action are currently implemented (Figure 4.3). A follow gesture is designed to allow the virtual model hand to follow user hands movements. Thus, our users are able to navigate in our virtual learning environment without any need to move the rest of the body parts in the real space.

In addition, the cube will be grabbed or released when the learner closes one of his hands or opens it. The user can hold one cube in each hand.

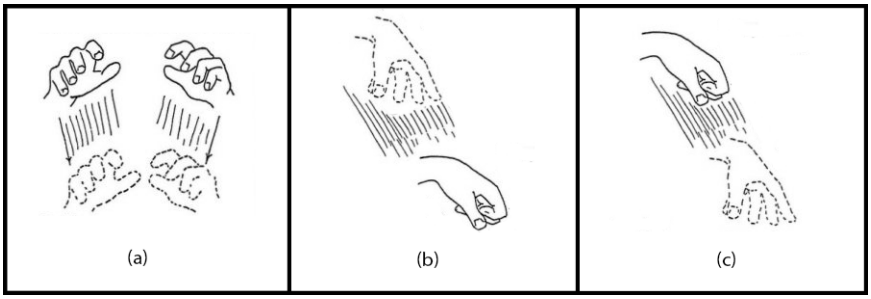

Figure 4.3: Hand-Gestures: (a) follow user hand. (b) hold objects. (c) release object.

4.6.2 Speech based Interaction

Adding more modalities to the act of learning will continue to increase the strength of the memory trace. Thus, we believe the incorporation of speech modality will bring better user experience in VR for learning. Our application analyzes the learner speech, and lets the virtual hands perform the corresponding actions.

There are mainly seven speech commands shown in table 1. The first command is used to select cubes that are located in the ground. The question mark in table 1 refers to one of the 26-characters that can be pronounced by the user. The user can select only one cube each time. If the two hands are holding cubes and the user wants to activate another voice command of selection, a message of notification will appear to tell the user to validate or free at least one hand to enable the operation of selection.

The second command is for manipulating the cubes and checking if the selected cubes are correct or not. This checking command can be applied in two methods: by using one hand each time (left or right) or the two hands together (right then left). There are few cases where the user makes mistakes and selects wrong cubes, and he or she wants to return the cube to the ground. In this case the user can use the last three commands to release the selected cube, also there are two methods for releasing the grabbing cube: by saying the command "free" plus the hand which hold the cube (right or left) or by saying "free my hands" to free his hands together at the same time.

Speech Orders		Tasks
say "Select box ?"		make the virtual hand hold the spoken alphabet (i.e select box B).
say "check"	"my hands"	make the virtual hands throw the held alphabet to white hole to check if they are correct or not.
	"left hand"	make the virtual left hand throw the held alphabet to white hole to check if they are correct or not.
	"right hand"	make the virtual right hand throw the held alphabet to white hole to check if they are correct or not.
say "free"	"my hands"	make the virtual hands release the held alphabet.
	"left"	make the virtual left hand release the held alphabet
	"right"	make the virtual hand release the held alphabet.

Table 4.1: The set of voice commands for controlling the game objects.

In addition, there are 15 commands implemented for controlling the game menu detailed in table 4.2. We can see that the commands are composed from one or two words which makes it easy to understand and pronounce.

The table 4.3 show the commands that have been used to control the system while playing. To give more freedom since the audience of our multimodal application are generally more active.

Speech Orders	Tasks
Start	This command is used to enter the game menu.
General Mode	This command is used to start a normal mode that supports the three integrated modalities (voice, hand gestures and head movement).
Speech Mode	This command is used to start a special mode that supports only voice commands.
Examples	This command is used to move to a new panel that gives few examples on how the control can be performed.
Help	This command is used to give a few explanations to the teacher or the parent of the learner about the game and control tips.
Exit	This command is used to exit from the game.
Settings	This command is used to open the settings panel to make a few configurations such as reduce or increase the voice and maximize or minimize the screen size.
About	This command is used to open new panels which contains the information of the developer to contact in case the user detects new bugs .
Back	This command is used to return to the previous panel
Easy	Those commands are used to choose the degree of difficulty of game. The difficulty level is defined depending on the number of alphabetic letters in each level. It starts from 3 letters to +7 letters.
Medium	
Hard	
Very Hard	
Yes	These commands are used to confirm or cancel the exit action command from the game.
No	

Table 4.2: The set of voice commands for controlling the game menu.

Speech Orders	Tasks
Pause	This command is used to pause the game.
Restart	This command is used to restart the level in case of loss to get another chance or just restarting the level for playing again .
Word	This command is used to pronounce the word of the current level.
Follow	This command is used to allow the virtual camera to follow the head of the learner to change his point of view.
Stop	This command is used to disable the head tracking technique and the virtual camera will return into its initial position.
Exit	These commands are used to cancel the action or confirm the return to the game menu.

Table 4.3: The set of voice commands for controlling the UI while playing.

4.6.3 Head movement based interaction

This technique is based on tracking the user's head position . Head tracking technique has the advantage of being intuitive. By this way of navigation, the point of view within the scene can change by following the user head movements.

In the application, three Head-Tracking movements are implemented (Figure 4.4). First movement is moving the head up or down to update the position of the virtual camera between top and bottom. Second, moving the head backward or ahead to get closer to or far from the cubes in the virtual learning environment. The last movement is to move the head in the left or the right to get lateral views. This method was inspired by the work [65].

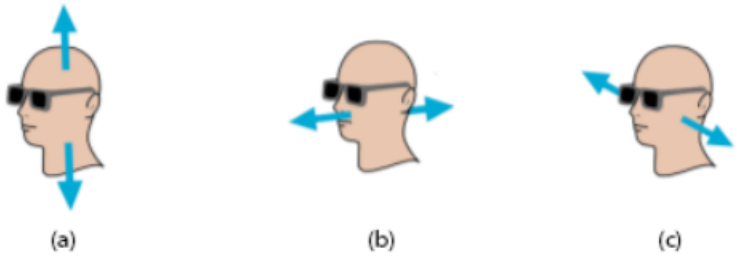

Figure 4.4: Head-Tracking movements: (a) Get up and down views. (b) Get closer to/far from the objects. (c) Get lateral views.

4.7 Definition of an interaction model

4.7.1 Interaction techniques

In what follows, we will give a definition of the interaction techniques that have been used with the 3D interaction tasks to create interaction model of our virtual reality application.

4.7.1.1 Go-Go

Go go is an interaction technique suggested by Poupyrev et al. (1996) [63]. This technique is based on the metaphor of grabbing, where the user can stretch out their virtual hand longer than their real arm length (Figure 4.5). The technique in its original form makes use of a nonlinear mapping, at the beginning the virtual hand follows the real hand position, but when the distance between the user and its hand becomes too long, the virtual hand shoots out further than the real hand could [46].

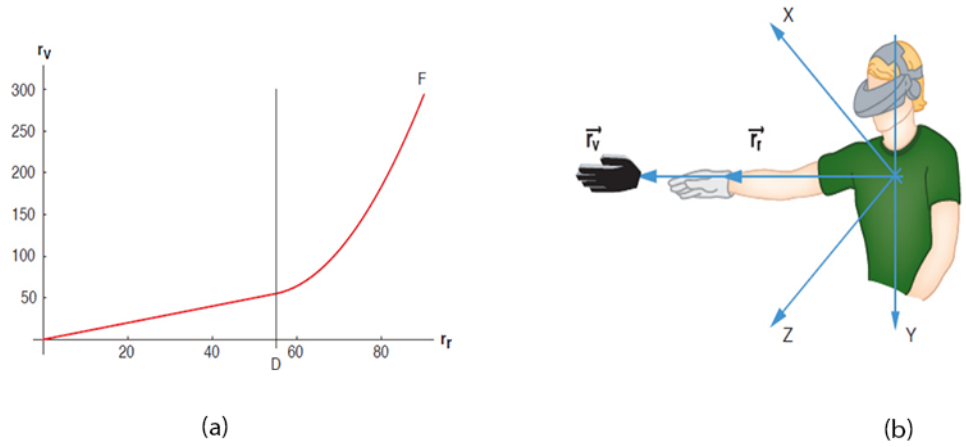

Figure 4.5: Go-Go interaction technique: (a) mapping function F. (b) egocentric coordinate system. [46]

4.7.1.2 FishTank VR

The concept of FishTank VR was introduced by Ware et al.[78]. Fish Tank Virtual Reality is a stereo image of a 3D scene viewed on a monitor placed very close to the eyes of the observer. This projection followed the head position of the user to provide an immersive experience at an affordable cost [32].

4.7.2 Navigation

After an analytical work focused on the interaction possibilities offered by the technologies we have, we thought of using the new interaction technique. This technique is based on GoGo Interaction Technique and FishtankVR . Its operation is as follows: the head tracking technique of FishtankVR is used to control virtual camera to change the point of view of the user within the scene in several ways (see section 4.6.3). Then, the gogo gives the possibility to the camera to extend forward far beyond the reach of any normal virtual camera being. This new technique minimizes the cost of energy used by the body and gives the user full control of his view within the scene. This metaphor adds a lot of immersion when playing the game, and helps the virtual content to appear more naturally within display.

The interface in Figure 4.6 is showing the learner using his head to change his point of view within the scene, to get closer to or far away from the virtual cubes.

Figure 4.6: User move his head sideways to get lateral view.

4.7.3 Selection

We propose two different solutions for performing the selection: the speech and the gesture modalities. Regarding the gesture modality, different kinds of information can be extracted from hand gesture modality during the system design such as the hand(s) position, movement and posture. This information gives the possibility to perform precise selection in the virtual space.

On the other side, linguistic features of the input speech are used with phrase recognition algorithms to select and grab the virtual object.

4.7.3.1 Selection with hand gesture

The selection is made by the gesture of the user's hand, as shown in Figure 4.7. For this, Go-Go interaction technique was implemented in the same manner as in the original work to select objects. This technique gives the ability to the user to reach the object that is far away from him by extending his/her arm. GoGo interaction techniques mimic the grabbing motion in the real world which give more feeling of presence to the user. This grabbing motion keeps the learner physically active inside the virtual learning environment and allows him to activate more motor neurons which should strengthen memory traces.

The operation of selection will be activated when the user closes his hand. this gesture will be captured and recognized by the kinect. The kinect will give hand joint position in three axes X, Y, Z and the orientation W. the value of x,y and z is limited between 0 and 1. Then, this position of hand will be projected onto the virtual environment. The measured point will be changed

depending on the difference of distance between the hand and body, same as mentioned in GO-GO technique.

Figure 4.7: User selecting boxes with his hand.

Figure 4.8 shows the state diagram for the gesture input in the selection command. The diagram is simplified by removing the states of error counting. Assuming that there is no hand detection yet, the system starts in state Idle. As soon as the hand is recognized by Kinect, the system goes to state follow, which allows the virtual hands to follow the movement of the real hands, and waits until the user closes his hand to proceed to the third state "selection". In this state the system checks if the virtual hand is collided with an object to proceed to the final state. However, in the case of failure, the system comes back to state follow. There are cases where the selection is

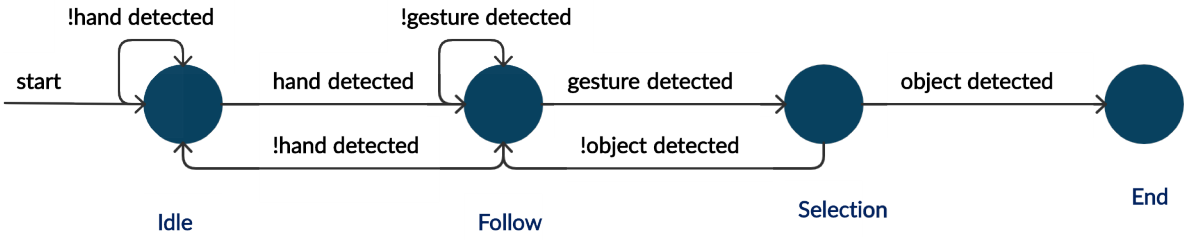

Figure 4.8: State diagram of selection with gesture.

hard to be achieved, especially when targeting an object partially occluded by another. Due to the previous problems, navigation technique is implemented to change the point of view within the

scene. This technique follows the head of the user to maximize the visible separation of the objects.

4.7.3.2 Selection with speech commands

Once the recognizer detects an input speech, the recognizer compares the input with each phrase in the grammar of selection, if this input matches the grammar, the speech recognizer can start the event of selection which allows the virtual hand to move and grab the cube that the learner pronounced.

Figure 4.9: User selecting box O by voice command.

After successful speech detection, the system goes to the second state and waits until the speech recognizer algorithm recognizes the voice command to proceed to the state of verification to compare if the input matches the grammar of selection or not. Then, the system passes to the state of selection. In this state we test if the object exists to proceed to the End state. In case the object does not exist or the audio input does not match the grammar of selection, the program goes back to state Idle regardless of the speech recognition result.

4.7.4 Manipulation

Same as Selection, the manipulation is performed through speech command and hand gestures. In our application, the user is allowed only to move the cubes. The rotation is not implemented since the cube has the alphabetic character in all the faces.

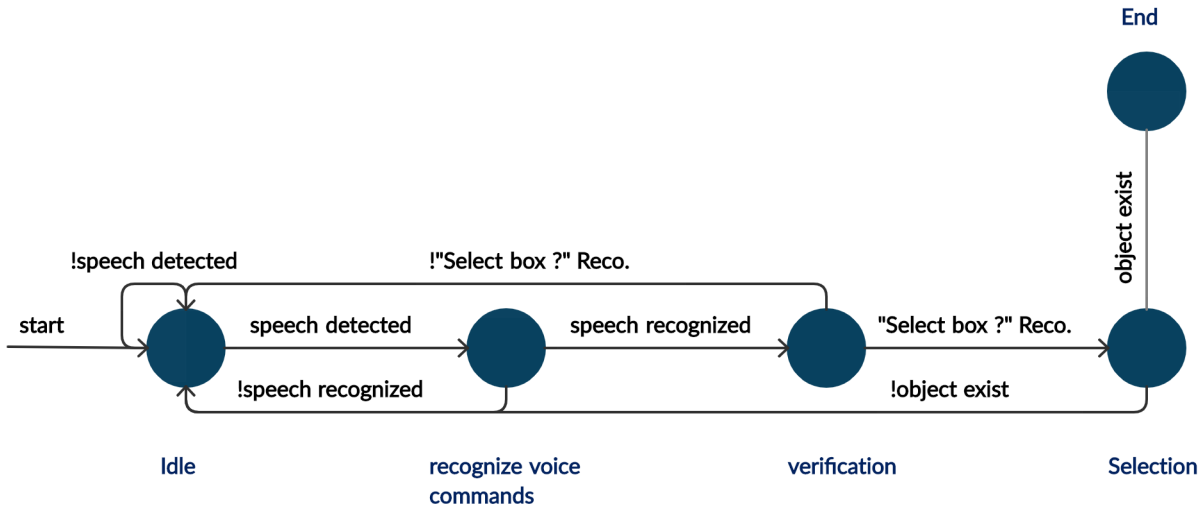

Figure 4.10: State diagram of selection with speech

4.7.4.1 Translation with hand gesture

Manipulation is the natural step after selection, allowing the user to move the object closer or away from themselves around the virtual learning environment. The operation of moving 3D objects will translate to 3D translations in the virtual environment. A translational movement of the real hand in the real space will be reproduced by the virtual hand. The movement of the selected 3D object attached to the virtual hand will be used to define the value and the direction of the translation vector.

In the end of manipulation, the 3D object will be released. The released object if it was in the space they will go down to ground due to the gravity, if the box collides with the white hole and the alphabet printed in cube is in the correct order, the box will be destroyed and the correct sound will be appeared to inform the learner that he is correct.

Figure 4.12 shows the state diagram of the operation of translation with gesture modality. It is very similar to the selection state diagram, but there are two differences. First, the selected object will keep attached to the virtual hand. it means that the position of the virtual cube will follow the position of the closed virtual hand. Second, the operation is terminated by opening the hand to release the object. This action will be detected by the system as a new gesture to drop the held object. The virtual object will go down due the gravity applied in the environment.

Figure 4.11: User Manipulating two boxes with his hand.

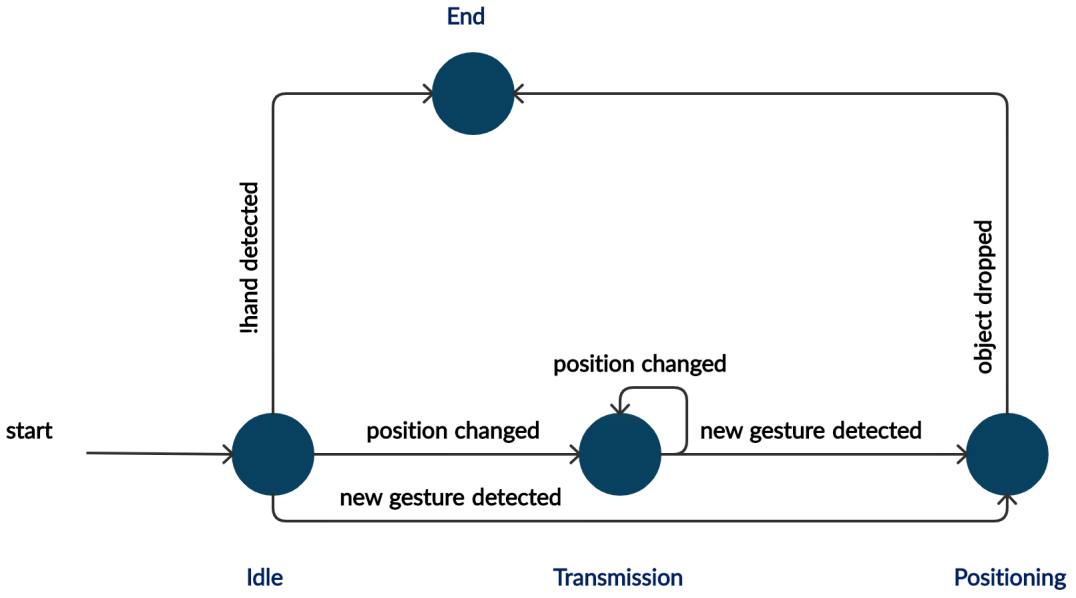

Figure 4.12: State diagram of translation with hand gesture.

4.7.4.2 Translation with speech command

Speech modality is used to determine the command types (validate, move or release the virtual cube). In order to make an action to translate the virtual object, a speech recognizer matches

the audio input against the grammar of manipulation. Once this audio input is recognized, the operation starts to be performed.

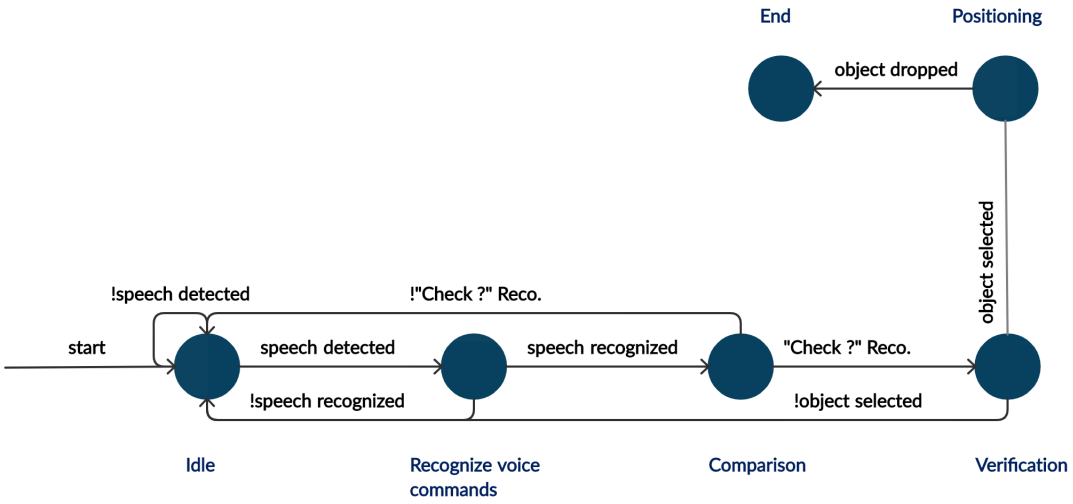

Figure 4.13: State diagram of translation with speech.

The state diagram shown in Figure 4.13 shows the two operations of the translation with speech modality. After successful word recognition and speed comparison of the audio inputs with grammar of manipulation. The system passes to the verification state to check if the virtual hand is holding an object or not. Then, the program goes to Translation state to drop the object in the white hole in case of "check ? "command or send it back to the ground with 'free ?' command.

4.7.5 Application Control

The application control is different from other universal interaction tasks, since the user uses the services offered by the application. In our Application we implemented two application control techniques: voice command and Gestural command.

4.7.5.1 Speech

Voice command is used to order the system to perform a specific operation. The user uses simple speech to initialize and execute the services that are offered by the application (see table 4.2 and table 4.3 of section 4.6.2).

Voice command technique is implemented in the game menu and in each level since it is very useful when the hands of the user are already busy with grabbing and holding cubes.

4.7.5.2 Gesture and posture

The system is controlled through the movement or the configuration of the hand. The learner uses his hand to move the virtual hand through the 2D menu. The click action is performed through the posture of the closed hand.

4.8 Summary

In this chapter, we have presented our multimodal virtual reality game for educational purposes. This application could help students to learn how to write words, especially at the introductory level.

More specifically, we have detailed and explained the interaction techniques implemented in the application which are created based on our conceptual model. The interface of the game is also illustrated by screenshots to describe to learners how to use our application.

Conclusion and future work

This thesis introduces the transaction of interface paradigms from graphical to natural user interfaces in order to facilitate the interaction between humans and machines.

A conceptual model of NUIs for virtual and augmented reality applications is proposed to aid the designer to plan their design project for a natural 3D interaction technique. The model is applied to a Virtual reality game to define a natural and multimodel 3D interaction technique, based on hand gesture, voice and head movement modalities by identifying the characteristics of each technique and apply it properly in the proposed system.

As for future work and perspectives, we intend to extend the developed application and make it accessible in the field of medicine and operating rooms rather than educational systems only. we will also try to add facial expression modality to our application and study the impact of these combinations.

BIBLIOGRAPHY

[1] R. Z. ABIDIN, H. ARSHAD, AND S. A. A. SHUKRI, *A framework of adaptive multimodal input for location-based augmented reality application*, Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9 (2017), pp. 97–103.

[2] P. ADKAR, *Unimodal and multimodal human computer interaction: a modern overview*, Int. J. Comput. Sci. Inf. Eng. Technol, 2 (2013), pp. 1–8.

[3] J. ALIPRANTIS, M. KONSTANTAKIS, R. NIKOPOULOU, P. MYLONAS, AND G. CARIDAKIS, *Natural interaction in augmented reality context.*, in VIPERC@ IRCDL, 2019, pp. 50–61.

[4] J. AMORES AND P. MAES, *Essence: Olfactory interfaces for unconscious influence of mood and cognitive performance*, in Proceedings of the 2017 CHI conference on human factors in computing systems, 2017, pp. 28–34.

[5] R. T. AZUMA, *A survey of augmented reality*, Presence: Teleoperators & Virtual Environments, 6 (1997), pp. 355–385.

[6] B. BACH, R. SICAT, J. BEYER, M. CORDEIL, AND H. PFISTER, *The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality?*, IEEE transactions on visualization and computer graphics, 24 (2017), pp. 457–467.

[7] S. BALBO, J. COUTAZ, AND D. SALBER, *Towards automatic evaluation of multimodal user interfaces*, in Proceedings of the 1st international conference on Intelligent user interfaces, 1993, pp. 201–208.

[8] J. D. BAYLISS, *Use of the evoked potential p3 component for control in a virtual apartment*, IEEE transactions on neural systems and rehabilitation engineering, 11 (2003), pp. 113–116.

[9] A. BELLARBI, *Vers l'immersion mobile en réalité augmentée: une approche basée sur le suivi robuste de cibles naturelles et sur l'interaction 3D*, PhD thesis, 2017.

[10] L. BESANÇON, P. ISSARTEL, M. AMMI, AND T. ISENBERG, *Hybrid tactile/tangible interaction for 3d data exploration*, IEEE transactions on visualization and computer graphics, 23 (2016), pp. 881–890.

[11] ——, *Mouse, tactile, and tangible input for 3d manipulation*, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 4727–4740.

[12] M. BILLINGHURST, A. CLARK, G. LEE, ET AL., *A survey of augmented reality*, Foundations and Trends® in Human–Computer Interaction, 8 (2015), pp. 73–272.

[13] M. BILLINGHURST ET AL., *Chapter 14: Gesture based interaction*, Haptic Input, Aug, 24 (2011).

[14] J. BLAKE, *The natural user interface revolution*, Natural User Interfaces in. Net, (2012), pp. 1–43.

[15] D. A. BOWMAN, *Interaction techniques for common tasks in immersive virtual environments*, Georgia Institute of Technology, (1999).

[16] W. BÜSCHEL, J. CHEN, R. DACHSELT, S. DRUCKER, T. DWYER, C. GÖRG, T. ISENBERG, A. KERREN, C. NORTH, AND W. STUERZLINGER, *Interaction for immersive analytics*, in Immersive Analytics, Springer, 2018, pp. 95–138.

[17] C. BUSSO AND S. S. NARAYANAN, *Interrelation between speech and facial gestures in emotional utterances: a single subject study*, IEEE Transactions on Audio, Speech, and Language Processing, 15 (2007), pp. 2331–2347.

[18] Z. CHEN, J. LI, Y. HUA, R. SHEN, AND A. BASU, *Multimodal interaction in augmented reality*, in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2017, pp. 206–209.

[19] S. CHIDAY, *Developing a kinect based holoportation system*, (2018).

[20] E. F. CHURCHILL, D. N. SNOWDON, AND A. J. MUNRO, *Collaborative virtual environments: digital places and spaces for interaction*, Springer Science & Business Media, 2012.

[21] M. CORDEIL, B. BACH, Y. LI, E. WILSON, AND T. DWYER, *Design space for spatio-data coordination: Tangible interaction devices for immersive information visualisation*, in 2017 IEEE Pacific Visualization Symposium (PacificVis), IEEE, 2017, pp. 46–50.

[22] J. COUTAZ, L. NIGAY, D. SALBER, A. BLANDFORD, J. MAY, AND R. M. YOUNG, *Four easy pieces for assessing the usability of multimodal interaction: the care properties*, in Human—Computer Interaction, Springer, 1995, pp. 115–120.

[23] J. CUI, A. KUIJPER, AND A. SOURIN, *Exploration of natural free-hand interaction for shape modeling using leap motion controller*, in 2016 International Conference on Cyberworlds (CW), IEEE, 2016, pp. 41–48.

[24] S. DENG, *Multimodal interactions in virtual environments using eye tracking and gesture control.*, PhD thesis, Bournemouth University, 2018.

[25] S. DO-LENH, P. JERMANN, A. LEGGE, G. ZUFFEREY, AND P. DILLENBOURG, *Tinkerlamp 2.0: designing and evaluating orchestration technologies for the classroom*, in European Conference on Technology Enhanced Learning, Springer, 2012, pp. 65–78.

[26] B. DUMAS, *Frameworks, description languages and fusion engines for multimodal interactive systems*, PhD thesis, Université de Fribourg, 2010.

[27] B. DUMAS, D. LALANNE, AND S. OVIATT, *Multimodal interfaces: A survey of principles, models and frameworks*, in Human machine interaction, Springer, 2009, pp. 3–26.

[28] S. R. ELLIS, *What are virtual environments?*, IEEE Computer Graphics and Applications, 14 (1994), pp. 17–22.

[29] M. ERIKSSON, *Reaching out to grasp in virtual reality: A qualitative usability evaluation of interaction techniques for selection and manipulation in a vr game*, 2016.

[30] C. FALCAO, A. C. LEMOS, AND M. SOARES, *Evaluation of natural user interface: a usability study based on the leap motion device*, Procedia Manufacturing, 3 (2015), pp. 5490–5495.

[31] G. FLORIN, G. TEODORA, AND B. SILVIU, *Design review of cad models using a nui leap motion sensor*, Journal of Industrial Design and Engineering Graphics, 10 (2015), pp. 21–24.

[32] A. FONNET AND Y. PRIÉ, *Survey of immersive analytics*, IEEE Transactions on Visualization and Computer Graphics, (2019).

[33] E. GHOMI, O. BAU, W. MACKAY, AND S. HUOT, *Conception et apprentissage des interactions tactiles: Arpege et le cas des postures multi-doigts*.

[34] G. GLONEK AND M. PIETRUSZKA, *Natural user interfaces (nui)*, (2012).

[35] P. HÜBNER, K. CLINTWORTH, Q. LIU, M. WEINMANN, AND S. WURSTHORN, *Evaluation of hololens tracking and depth sensing for indoor mapping applications*, Sensors, 20 (2020), p. 1021.

[36] W. HÜRST AND C. VAN WEZEL, *Gesture-based interaction via finger tracking for mobile augmented reality*, Multimedia Tools and Applications, 62 (2013), pp. 233–258.

[37] J. JANKOWSKI AND M. HACHET, *Advances in interaction with 3d environments*, in Computer Graphics Forum, vol. 34, Wiley Online Library, 2015, pp. 152–190.

[38] M. C. JOHNSON-GLENBERG, *Immersive vr and education: Embodied design principles that include gesture and hand controls*, Frontiers in Robotics and AI, 5 (2018), p. 81.

[39] S. KHALID, S. ULLAH, N. ALI, A. ALAM, I. RABBI, I. U. REHMAN, AND M. AZHAR, *Navigation aids in collaborative virtual environments: Comparison of 3dml, audio, textual, arrows-casting*, IEEE Access, 7 (2019), pp. 152979–152989.

[40] H. KHAROUB, M. LATAIFEH, AND N. AHMED, *3d user interface design and usability for immersive vr*, Applied Sciences, 9 (2019), p. 4861.

[41] M. KIM, J. LEE, C. JEON, AND J. KIM, *A study on interaction of gaze pointer-based user interface in mobile virtual reality environment*, Symmetry, 9 (2017), p. 189.

[42] W. A. KÖNIG, R. RÄDLE, AND H. REITERER, *Squidy: a zoomable design environment for natural user interfaces*, in CHI'09 Extended Abstracts on Human Factors in Computing Systems, 2009, pp. 4561–4566.

[43] P. KUMAR, H. GAUBA, P. P. ROY, AND D. P. DOGRA, *A multimodal framework for sensor based sign language recognition*, Neurocomputing, 259 (2017), pp. 21–38.

[44] N. KUNKEL, S. SOECHTIG, J. MINIMAN, AND C. STAUCH, *Augmented and virtual reality go to work: Seeing business through a different lens*, Tech Trends, (2016).

[45] E. C. LALOR, S. P. KELLY, C. FINUCANE, R. BURKE, R. SMITH, R. B. REILLY, AND G. MCDARBY, *Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment*, EURASIP Journal on Advances in Signal Processing, 2005 (2005), p. 706906.

[46] J. J. LAVIOLA JR, E. KRUIJFF, R. P. MCMAHAN, D. BOWMAN, AND I. P. POUPYREV, *3D user interfaces: theory and practice*, Addison-Wesley Professional, 2017.

[47] A. LÉCUYER, F. LOTTE, R. B. REILLY, R. LEEB, M. HIROSE, AND M. SLATER, *Brain-computer interfaces, virtual reality, and videogames*, Computer, 41 (2008), pp. 66–72.

[48] R. LEEB, F. LEE, C. KEINRATH, R. SCHERER, H. BISCHOF, AND G. PFURTSCHELLER, *Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 15 (2007), pp. 473–482.

[49] F. LOTTE, A. LÉCUYER, Y. RENARD, F. LAMARCHE, AND B. ARNALDI, *Classification de données cérébrales par système d'inférence flou pour l'utilisation d'interfaces cerveau-ordinateur en réalité virtuelle*, 2006.

[50] M. MA, B. J. MEYER, L. LIN, R. PROFFITT, AND M. SKUBIC, *Vicovr-based wireless daily activity recognition and assessment system for stroke rehabilitation*, in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 1117–1121.

[51] M. MALCANGI, K. OUAZZANE, AND P. PATEL, *Audio-visual fuzzy fusion for robust speech recognition*, in The 2013 International Joint Conference on Neural Networks (IJCNN), IEEE, 2013, pp. 1–8.

[52] A. MARGIENĖ AND S. RAMANAUSKAITĖ, *Trends and challenges of multimodal user interfaces*, in 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream), IEEE, 2019, pp. 1–5.

[53] J.-C. MARTIN, *Tycoon: Theoretical framework and software tools for multimodal interfaces*, Intelligence and Multimodality in Multimedia interfaces, (1998), pp. 1–25.

[54] R. MCNANEY, M. BALAAM, A. HOLDEN, G. SCHOFIELD, D. JACKSON, M. WEBSTER, B. GALNA, G. BARRY, L. ROCHESTER, AND P. OLIVIER, *Designing for and with people with parkinson's: A focus on exergaming*, in Proceedings of the 33rd annual ACM conference on Human Factors in Computing Systems, 2015, pp. 501–510.

[55] P. MILGRAM AND F. KISHINO, *A taxonomy of mixed reality visual displays*, IEICE TRANSACTIONS on Information and Systems, 77 (1994), pp. 1321–1329.

[56] T. NI, D. BOWMAN, AND C. NORTH, *Airstroke: bringing unistroke text entry to freehand gesture interfaces*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 2473–2476.

[57] L. NIGAY, *Modalité d'interaction et multimodalité*, PhD thesis, 2001.

[58] M. OBRIST, A. N. TUCH, AND K. HORNBAEK, *Opportunities for odor: experiences with smell and implications for technology*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, pp. 2843–2852.

[59] K. O'HARA, A. SELLEN, AND R. HARPER, *Embodiment in brain-computer interaction*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 353–362.

[60] L. E. ORTIZ, E. V. CABRERA, AND L. M. GONÇALVES, *Depth data error modeling of the zed 3d vision sensor from stereolabs*, ELCVIA: electronic letters on computer vision and image analysis, 17 (2018), pp. 0001–15.

[61] N. OURAMDANE, S. OTMANE, AND M. MALLEM, *Interaction 3d en réalité virtuelle-etat de l'art*, (2009).

[62] E. PAPADAKI, S. NTOA, I. ADAMI, AND C. STEPHANIDIS, *Let's cook: An augmented reality system towards developing cooking skills for children with cognitive impairments*, in International Conference on Smart Objects and Technologies for Social Good, Springer, 2017, pp. 237–247.

[63] I. POUPYREV, M. BILLINGHURST, S. WEGHORST, AND T. ICHIKAWA, *The go-go interaction technique: non-linear mapping for direct manipulation in vr*, in Proceedings of the 9th annual ACM symposium on User interface software and technology, 1996, pp. 79–80.

[64] R. RADIAH, Y. ABDRABOU, T. MAYER, K. PFEUFFER, AND F. ALT, *Gazebutton: Enhancing buttons with eye gaze interactions*, (2019).

[65] A. RICCA, A. CHELLALI, AND S. OTMANE, *Comparing touch-based and head-tracking navigation techniques in a virtual reality biopsy simulator*, Virtual Reality, (2020), pp. 1–18.

[66] J. RUBIN AND D. CHISNELL, *How to plan, design, and conduct effective tests*, Handbook of usability testing, (2008), p. 348.

[67] N. SAE-BAE, J. WU, N. MEMON, J. KONRAD, AND P. ISHWAR, *Emerging nui-based methods for user authentication: A new taxonomy and survey*, IEEE Transactions on Biometrics, Behavior, and Identity Science, 1 (2019), pp. 5–31.

[68] F. SASANGOHAR, I. S. MACKENZIE, AND S. D. SCOTT, *Evaluation of mouse and touch input for a tabletop display using fitts' reciprocal tapping task*, in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 53, SAGE Publications Sage CA: Los Angeles, CA, 2009, pp. 839–843.

[69] J. SCHIOPPO, Z. MEYER, D. FABIANO, AND S. CANAVAN, *Sign language recognition: Learning american sign language in a virtual environment*, in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–6.

[70] C. SCHWESIG, *What makes an interface feel organic?*, Communications of the ACM, 51 (2008), pp. 67–69.

[71] J. SONG, S. CHO, S.-Y. BAEK, K. LEE, AND H. BANG, *Gafinc: Gaze and finger control interface for 3d model manipulation in cad application*, Computer-Aided Design, 46 (2014), pp. 239–245.

[72] L. STERNBERGER, *Interaction en réalité virtuelle*, PhD thesis, Strasbourg 1, 2006.

[73] K. SUN, C. YU, W. SHI, L. LIU, AND Y. SHI, *Lip-interact: Improving mobile device interaction with silent speech commands*, in Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, 2018, pp. 581–593.

[74] X. TONG, S. PEKCETIN, D. GROMALA, AND F. MACHUCA, *Exploring body gestures as natural user interface for flying in a virtual reality game with kinect*, Electronic Imaging, 2017 (2017), pp. 60–63.

[75] M. TURK, *Multimodal interaction: A review*, Pattern Recognition Letters, 36 (2014), pp. 189–195.

[76] S. ULLAH, *Multi-modal Interaction in Collaborative Virtual Environments: Study and analysis of performance in collaborative work*, PhD thesis, Université d'Evry-Val d'Essonne, 2011.

[77] D. VANACKEN, A. BEZNOSYK, AND K. CONINX, *Help systems for gestural interfaces and their effect on collaboration and communication*, in Workshop on gesture-based interaction design: communication and cognition, Citeseer, 2014.

[78] C. WARE, K. ARTHUR, AND K. S. BOOTH, *Fish tank virtual reality*, in Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems, 1993, pp. 37–42.

[79] C. R. WREN, A. AZARBAYEJANI, T. DARRELL, AND A. P. PENTLAND, *Pfinder: Real-time tracking of the human body*, IEEE Transactions on pattern analysis and machine intelligence, 19 (1997), pp. 780–785.

[80] S. WU, *Study and design of interaction techniques to facilitate object selection and manipulation in virtual environments on mobile devices*, PhD thesis, 2015.

[81] F. ZHANG, S. CHU, R. PAN, N. JI, AND L. XI, *Double hand-gesture interaction for walk-through in vr environment*, in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), IEEE, 2017, pp. 539–544.

[82] C. ZIMMER, M. BERTRAM, F. BÜNTIG, D. DROCHTERT, AND C. GEIGER, *Mobile augmented reality illustrations that entertain and inform: Design and implementation issues with the hololens*, in SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, 2017, pp. 1–7.