Tiago Miguel Martins Coelho
Licenciado em Ciências da Engenharia Electrotécnica e de Computadores

Avatar modeling: a telepresence study with natural user interface

Dissertação para obtenção do Grau de Mestre em Engenharia Electrotécnica e de Computadores

Orientador: Yves Rybarczyk, PhD, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
Co-orientador: Tiago Oliveira Machado de Figueiredo Cardoso, PhD, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri:

Presidente: Doutor Luís Filipe Figueira de Brito Palma
Arguentes: Doutor Paulo Ignácio Noriega
Doutor Vitor Emanuel Pereira Duarte dos Santos
Vogais: Doutor Yves Philippe Rybarczyk
Doutor Tiago Oliveira Machado de Figueiredo Cardoso

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2013

Avatar modeling: a telepresence study with natural user interface

Modelação de avatares para estudos de telepresença com recurso a interfaces de utilizador naturais

Acknowledgement

I would like to thank my supervisor Professor Yves Rybarczyk for the opportunity of doing this project, but also for his guidance, knowledge and support. I also want to thank Professor Tiago Cardoso for his help, attention and support.

I also wish to thank Professor Rita de Oliveira for all the help she provided, and above all, her support and encouragement.

I want to thank all the participants who took place in the experiments, for they have done it with the only reward of helping a scientist colleague contributing to the progress of science.

I want to thank all the friends who were there in the good and bad moments over the years. Many came and many have gone, but it was a pleasure to have stood by their side all this time. The brotherhood I shared with them will always remain. The fellowship I experienced at the Republic will always be special and unique. To them I owe my mental sanity.

At last, but not least, I want to thank my family, who always supported and encouraged me, no matter how far.

Abstract

Virtual environments are an increasing trend in today's society. In this context, the avatar concept appears as the representation of the user in the virtual world. Nevertheless the relationship between avatars and human beings lacks on empirical studies in what concerns their interaction. Based on this motivation, this work aimed at studying how the morphology's modeling and dynamics affect the control between the avatar and its user. An experiment was conducted to measure telepresence and ownership on the participants while using a natural user interface to control the avatar. In that experiment, affordances were used as behavioral assessment on the virtual environment as the user controls the avatar when it passes through apertures of various sizes. The results show that in virtual environments, the feelings of telepresence and ownership are greater when the kinematics and the avatar proportions are closer to the user.

Keywords:

Avatar, telepresence, ownership, affordances, natural user interface, virtual environment.

Sumário

Os ambientes virtuais são uma tendência em crescimento nos dias de hoje. Neste contexto, o conceito de avatar aparece como a representação do indivíduo no mundo virtual. No entanto, essa relação carece de estudos empíricos relativamente à natureza da interação entre avatares e seres humanos. Neste trabalho foi estudado como a modelação da morfologia e do dinamismo do avatar afetam o seu controlo por parte do utilizador. Foi realizada uma experiência para medir a telepresença e apropriação nos participantes enquanto estes utilizavam uma interface de utilizador natural. Nessa experiência foram usadas possibilidades de acção para uma avaliação comportamental do desempenho do indivíduo enquanto este devia guiar o avatar através de aberturas de várias dimensões. Os resultados mostram que em ambientes virtuais, os sentimentos de telepresença e apropriação são tanto maiores quanto mais a cinemática e as proporções do avatar se aproximarem daqueles do utilizador.

Palavras-chave

Avatar, telepresença, apropriação, possibilidade de acção, interface de utilizador natural, ambientes virtuais.

Index

Figure Index

Table Index

1 Introduction

In this chapter it will be made an introduction on the subject of analysis. It will also be revealed the importance of the investigation on this area of knowledge regarding the relationship between a user and his/her avatar in a virtual environment.

1.1 Motivation

Nowadays, the virtual environments are emerging in multimedia applications and video games. In these kinds of systems involving virtual environments, there is usually a virtual representation of the user, which is his/her avatar. The avatar is the link between the user and the virtual environment. Nevertheless, this bond created between the user and the avatar lacks on empirical knowledge. There are some characteristics about the avatar that might enhance that relationship, such as the interface between the user and the virtual environment, or even the visual aspect, morphology and the dynamic of the avatar. All of these characteristics may have an important role in making the user feel more comfortable when he/she is represented by the avatar in the virtual world. In other words, if these characteristics were to be exhaustively studied in order to have an almost “perfect” avatar, it would be easier for the user to feel the avatar as him/herself. This would lead to a different specific avatar for everyone, but would have identical characteristics for all the avatars, such as being scaled-to-user, visually resembled, dynamic and with real-time response. When the users can be completely immersed in the virtual environment, the experience in virtual environments would seem as real as the real world.

1.2 Goals and work done

The main goal of this thesis is to figure out some of these aspects that may tighten the relation between the user and the avatar. For that, it will be tested how the morphology and movements of the avatar influence the immersion of the user on the virtual environment. Two different types of avatar were created: one was tailored to the anatomical proportions of each user and a more generic avatar that was the same for every participant. While the anatomically proportional avatar replicated every move the user made, the generic version could only rotate upon himself and move sideways with an animation of a sidestep. The interface was performed using a Kinect, which allowed natural movements to control the avatar. A prototype containing a virtual environment was created. The goal was to study several parameters of the avatars towards testing how immerse the users are when facing situations with each one of the avatars created and compare the results afterwards.

1.3 Contributions

This thesis' contributions are divided in two categories: a study about avatar characteristics and a system architecture to do so. The system architecture consists in a virtual environment integrating a full body motion capture, so that a user can control an avatar with his/her own body. With a prototype based on the system architecture, it was possible to carry out a study on avatars' characteristics.

This project participated on eNTERFACE'13 conference and contributed with an article named "Body ownership of virtual avatars: an affordance approach of telepresence", in IFIP Advances in Information and Communication Technology – A Springer Series in Computer Science.

1.4 Structure of the dissertation

On the second chapter, there is the current state of the art concerning full body motion capture, game engines and modeling software. This chapter will also target topics such as telepresence, ownership, teleoperation, virtual environments and affordances.

The third chapter will refer the proposal resulting from this work. The proposal consists on the approach that was made for the study. It includes the description of the designed model and the implemented prototype.

On the fourth chapter there is a description of the methods and materials used to perform the experimental validation. In this section, there is also a brief explanation about the chosen tools for this project. Here all the results are presented and then discussed.

The fifth chapter will be about the conclusions achieved with the work. This chapter also refers to possible approaches to continue this study with some other parameters and applications.

2 State of the art

In this chapter it will be discussed two distinct subjects: one will be about the tools used to accomplish the developed prototype, and the other will be about important state of the art concepts that are used in the thesis.

Regarding the tools, a brief description of existing full body motion capture devices will be made, where it will be described the way they work and some of their biggest advantages and disadvantages. This section where tools are described will also contain information about the software used, such as modeling software and game engines.

The second section will contain important concepts such as telepresence, ownership and affordances. These and some related concepts are explained in order to better understand some procedures and conclusions.

2.1 Full body motion capture

Full body motion capture refers to any method in which the result is the capture of the human motion on a full body scale. The purpose of full body motion capture is to be able to capture the full movement of the human being. This technology has a wide range of applications which include the military, cinema, gaming, simulators, entertainment and media.

2.1.1 Mechanical sensors

This technology is based on goniometers, which are attached to a suit. Goniometers are sensors that measure angles with potentiometers. These sensors work by measuring the angles made on the joints of the individual and the data is transmitted to the apparatus responsible for the data processing in real time. The position of every joint is calculated using kinematics. The biggest disadvantage is that the equipment does not know the position of the floor, which means that if the user performs a jump, there will only be data referring to the joint movement disregarding the height. The advantages are the immunity to magnetic and light interference.

One of the most practical examples of mechanical sensors is a project by Yoxall, Heller and Chamberlain (2011), in which the system is used to help the elder community by allowing them to have a better perception over their bodies’ limitations.

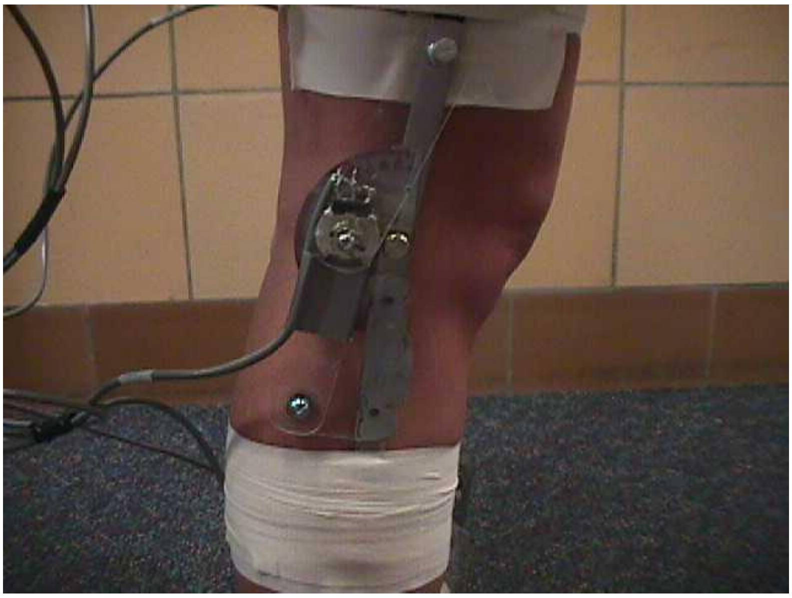


Figure 2-1- Example of a goniometer applied on a knee

Source- (Hoffman et al., 2006)

2.1.2 Magnetic sensors

Similar to goniometers, the magnetometers also require a suit, and the sensors are also placed over the joints of the individual. These sensors measure the proximity between them and a magnetic source near the equipment. The whole suit and equipment must be connected by a cable to the receiver. These types of sensors allow for six degrees of freedom, which track the movements upon the three coordinated axis and the rotation over those same axis. This system allows precision on position and rotation of every joint. The biggest disadvantage of this method is the interference with the proximity of metal. Buildings for example may cause interferences and disrupt the surrounding magnetic field due to the presence of metal (even small amounts) in the building’s infrastructure.

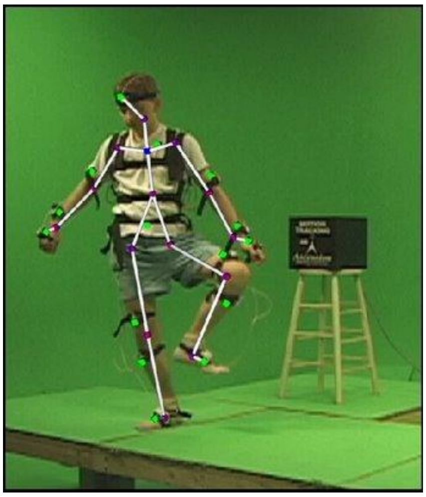


Figure 2-2- Example of a display to use magnetic sensors and resulting skeleton

Source: http://www.sfu.ca/~mma25/iat445/research.html

2.1.3 Inertial sensors

In this category of sensors, users wear suits with accelerometers or gyroscopes to detect movement variations (Roetenberg et al., 2013). To get more accuracy they can be combined with magnetometers which detect the magnetic field of the Earth. The magnetometer helps to define the horizontal component of the sensors. This method also provides six degrees of freedom for each sensor. One disadvantage is the lack of precision relatively to the environment. The principal advantage of this technology is that sensors do not need to be connected to a receiver, which means that the user can move freely. The data is only recorded and processed afterwards.

Figure 2-3- Suit with inertial technology

Source- (Roetenberg et al., 2013)

2.1.4 Optic method

This capture method allows the user to have a higher freedom of movements and space than in previously described methods. On the other hand this method requires a lot of pos-processing, which means that the results can only be observed a few hours after. In addition, this method requires a large number of cameras to allow the capture of a maximum number of points. This kind of technology is subdivided in two categories: passive and active mode.

Passive

In the passive mode, there is a visible or infrared light emitter near the camera. The user has to wear reflective markers on a suit or on the skin. These markers reflect the light which is captured by the cameras. These cameras have a contrast adjustment, which only detects those light reflections. An example is OptiTrack (Natural Point), which is a system based on infrared reflective markers. The user has to be in a 4m x 4m capture space and at least 12 cameras are needed to allow an accurate record.

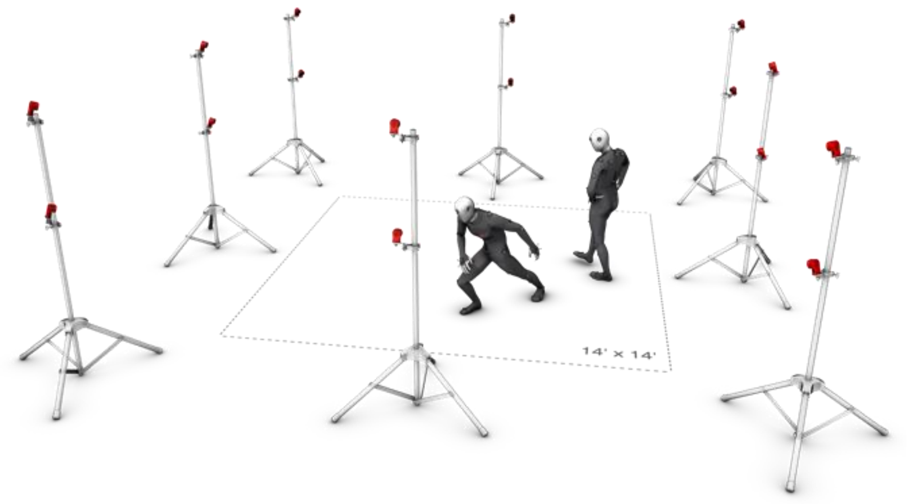


Figure 2-4- The OptiTrack system and the space required to use it
Source- www.naturalpoint.com/optitrack/products/flex-13/indepth.html

Active

In the active mode, the user wears a suit with LEDs (light emitting diode) instead of reflective markers. The LEDs blink one at a time, so the software always knows which points are visible. This kind of technology is mostly used for near activity, and it is capable of tracking objects which require very high precision, such as facial expressions or close hand movements. Examples of systems using this technology are Optotrak and FlashPoint (Li et al. 1999).

2.1.5 Video mode

The motion capture can also be done with video, which is a method that does not require a specific suit. This is an advantage due to the time required to mount the equipment and freedom of movements. Motion capture by video is very complex at a software level because the data is only recorded in 2D by a video camera. For this matter it is necessary at least two cameras for measuring depth. It is possible to perform this motion capture with one or two webcams (Rybarczyk, 2010).

2.1.6 Kinect

The Kinect is a motion sensing input device that was released in 2010 as a game accessory to the Xbox. Ever since its debut, the Kinect's motion sensing potential was recognized and readily adapted by third-party development. This sensor has some similarities with the optic and video methods, but diverges in some aspects. The Kinect has two cameras: a depth camera and a RGB (Red, Green and Blue) camera (see Figure 2.5). The depth camera works with infrared light beams that detect movement. The depth camera has a resolution of 320 x 240 with an encoding of 16 bits at 30 fps. The RGB camera has a resolution of 640x480, 32 bits at 30 fps. Also, the Kinect has a built-in microphone that records at 16 kHz. The Kinect has internal software that detects the users and follows his/her movements. The main advantage of the Kinect is the fact that it is a low cost motion capture device (€150), and enables the user to have a complete freedom of movements without the necessity to wear a specific suit. The main problem with the Kinect is the lack of precision and interference with sunlight. This is typically an indoor sensor.

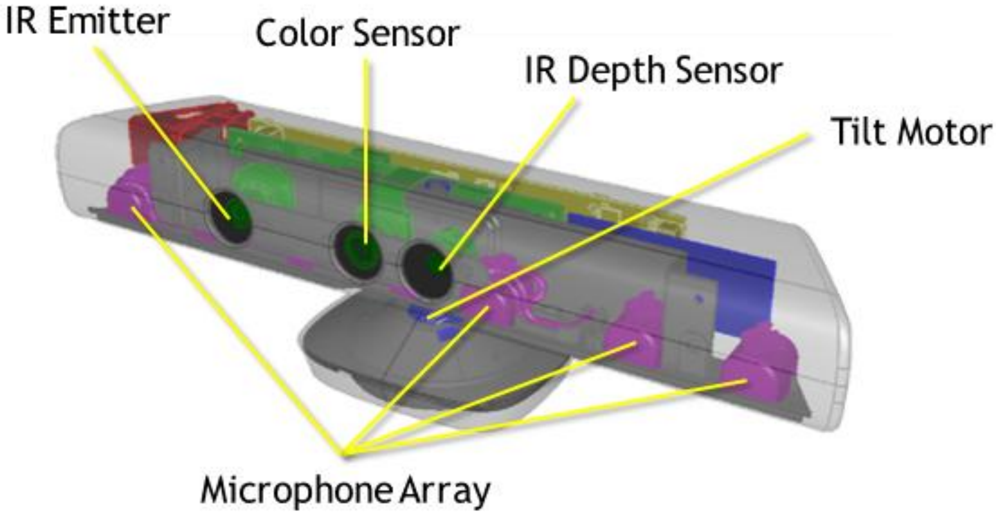


Figure 2-5- The Kinect and its components

Source: http://msdn.microsoft.com/en-us/library/jj131033.aspx

Table 2.1- Main characteristics of motion capture devices

	Technology	Distance to the receiver	Ground reference	Advantages	Disadvantages
Mechanical	Goniometers	By cable	No	No interferences	Hard to keep on the right position
Magnetic	Magnetometers	Max 15 m radius by cable	Yes	Good precision	Interference from magnetic fields
Inertial	Accelerometers ;gyroscopes	Freedom; can be processed later	No	Very accurate	High cost; Pos-processing
Optic	Camera	4m x 4m space	No	High sampling rates for better precision	High cost; Pos-processing
Video	Camera	Camera range	No	No special hardware required	Pos-processing; complex algorithms
Kinect	Infrared depth camera	From 0.6 to 4 meters	Yes	Quick install of setup; no special suit required; low cost	Interference from sunlight

In what matters full body motion capture, the Kinect was chosen to perform the interaction between the user and the virtual environment due to its low price, good precision and real time data processing.

2.2 Development tools

In order to develop a virtual environment, it is necessary to have modeling software and a game engine. The modeling software is used to create tri-dimensional models. After the modeling of the objects, they are imported into a game engine in order to be a part of an environment and to allow interaction. The game engine is a tool for development and creation of simulators and games. Next are the most popular modeling software and game engines.

2.2.1 Autodesk 3ds Max

Autodesk 3ds Max is a 3D modeling software that allows the creation of tri-dimensional models and animations. This software is very popular among professionals who work in modeling and animation, however, it costs around 3900 $(USD). This software is mostly used for games, although it can also be used for film animations. The main advantages of this program is that it is easy to use, and since there is a great community using it, there is also a lot of documentation, tutorials and pre-made models available.

2.2.2 Autodesk Maya

Autodesk Maya is very similar with Autodesk 3ds Max. The main difference is the GUI (graphical user interface) that is more complex, which allows to create more complex objects. This software is more used for film animation than for game creation. Character animation is easy to do with this tool. The cost of this software is the same as Autodesk 3ds max (3900 $). In addition to Autodesk 3ds max, Maya has several simulators, like particles, fluids or hair. It is also possible to add scripts in a scripting language called MEL (Maya Embedded scripting Language).

2.2.3 Blender

Blender is free software, and allows the creation of 3d models to use them in a game engine (inclusive its own). This game engine has simulators for gravity, collisions, fluids and many others, and allows scripting in Python. The GUI of Blender is very different from the others, which makes it not very intuitive and not so easy to learn. However, there is a large community of Blender users that provides tutorials and forums for discussion.

2.2.4 Unreal Engine

The Unreal Engine is one of the most popular game engines used worldwide, and it is used since 1998. It has a large community of users, which means that a lot of information is available. It uses its own language code called UnrealScript. It is free as long as it is for non-commercial purposes. On the game industry there are renowned companies using it, such as Ubisoft or Square-Enix. The platforms it exports to are PC and game consoles such as WII, PS3 and XBOX 360.

2.2.5 CryEngine

This game engine was created in 2007. The latest version of this software is also free for educational and non-commercial purposes. There is also a great community that uses this software, and it has lots of tutorials and information available. The programming language used is LUA or C++. Like the Unreal Engine, it produces for PC, WII, PS3 and XBOX 360.

2.2.6 Unity 3d

Unity 3d is a quite pure game engine, as it only allows modeling of primitive objects and designing virtual worlds. However, models can be easily imported from 3D computer graphics software. Unity is free, although for publishing games for platforms other than PC there is a need to purchase Unity 3d Pro. In terms of graphics, Unity stays behind CryEngine and Unreal Engine. The main advantage Unity has over CryEngine and Unreal Engine is that Unity can produce to a wider range of platforms, such as Mac, browser plugins, iOs and Android. It is used in the game industry by Electronic Arts.

Table 2.2- Comparative table of game engines and modeling software

	Cost	Modeling	Game engine	Platforms to publish
3ds Max	3900$	Yes	No	------------------------
Maya	3900$	Yes	No	------------------------
Blender	Free	Yes	Medium	PC, MAC, Linux
Cryengine	Free/1.2 million$	No	Excellent	PC, WII, PS3 and XBOX 360
Unreal Engine	Free/0.5 million$	No	Excellent	PC, WII, PS3 and XBOX 360
Unity3d	Free/1500$	No	Very good	PC, WII, PS3, XBOX 360, Android, ios, MAC and browser plugin

After these descriptions, it was decided to use Blender as it is good and free software to model the avatars, and Unity 3d to create the virtual environment and the prototype. Unity 3d was used because it is free software and has very good function libraries for use with the Kinect.

2.3 Presence

What is being present? Presence seems to be related to where the body is, but deep down, presence is about the brain representation of the body, also known as body schema. Even when people go to the cinema, play sports or games, they forget where they are physically and “merge” into the task or environment. Mihály Csíkszentmihályi introduced the concept of *flow* (Csíkszentmihályi, 1975). *Flow* is about the immersion someone feels upon a certain task. *Flow* is an emotional state that is achieved when the individual is performing an activity in a high state of concentration, and loses the notion of time and worries. This state allows people to involve themselves more into the task. For Csíkszentmihályi, motivation is a very important factor for reaching *flow*. Other important things are: interest on the task, having goals to reach, seeing results in real time, adjusted difficulty for the individual, losing notion of time and having control over the task. It is not necessary to obtain every point to reach the *flow*. This can also be applied to virtual environments, such as video games, where the individual may become so immerse and present in the virtual world that the game may become an addiction (Park & Hwang, 2009). *Flow* and telepresence are related one to another when the immersion happens in a virtual environment (Takatalo, 2002).

According to Lombard and Ditton (1997) there are two types of presence: social and physical. The social presence refers to the feeling of being with someone. This can happen at distance by means of technology, whether it is by synchronous communication such as telephone or online chat, or by asynchronous communication such as letters and email. Social presence occurs with social interaction. The physical presence refers to the feeling of being present somewhere, whether it is in a physical or virtual environment. When both physical and social presences are achieved simultaneously, the feeling is known as co-presence. Co-presence refers to the feeling of sharing the same space with other person(s), who may be apart. Common ways to achieve co-presence are Shared Virtual Environments (SVE) and teleconference. In Shared Virtual Environments people can interact with each other. One of the most common cases is multiplayer online games, in which various users must, for instance cooperatively solve problems or situations (Bonk & Dennen, 2005). In teleconference people can talk and see each other in real time. Having both audio and video is very good for communicating, because of the importance of non-verbal communication, such as gestures (Clark, 1996). In the case of presence in a remote environment this feeling is known as telepresence.

2.3.1 Telepresence

Telepresence is the feeling of being present somewhere the person is not (Minsky, M. 1980). This feeling can be achieved while an individual is performing a certain task in a virtual environment, such as a game or using a simulator (Slater, 1994). Another way this feeling can occur is in teleoperation, where the user controls a robot at a distance with the help of a camera (Rybarczyk & Mestre, 2013). Telepresence is an important feeling, as it increases the immersion of the individual upon a certain task. Teleoperation and virtual environments are the most common situations in which a feeling of telepresence may occur.

2.3.2 Ownership

One of the most specific cases of telepresence is body ownership, in which the individual is so immerse in the teleoperation task he/she is performing that he/she believes the machine is part of him/her (Rybarczyk et al., 2012). This also happens in virtual reality, in which the individual believes he/she is the avatar. The most famous example of ownership is the Rubber Hand Illusion (Botvinik & Cohen, 1998), in which the participant's hand is hidden and only a rubber hand is visible in its place. A tactile stimulation is applied in simultaneous to both hands. After a while the individual has the feeling that the fake hand is his/her own. This was also experimented with virtual reality (Yuan & Steed, 2010; Tsakiris et al., 2006). Both of the examples (real and virtual environments) show that the individual started perceiving the fake hand as their own. In a similar experiment, the brain activity of the participants was recorded while the experiment was running (Ehrsson et al., 2004). The results showed a significant activation of the parietal cortex only in presence of a synchronous and congruent visuo-tactile stimulation between the rubber and the real hand. In addition, a positive correlation between the physiological and ownership questionnaire data confirms the fact that the participants were considering the rubber hand as their own hand. In another experiment the participants were blindfolded whereas the brain activity has been measured as well (Ehrsson et al., 2005). Since the participants were blindfolded, the illusion could not be explained by simple visual cues but a brain interpretation of environmental stimulations. These experiments showed that ownership occurs in the brain, after integration of multimodal information (vision, touch and proprioception) in order to build a coherent representation of the body.

A similar experiment was made, in which the participants wore a head-mounted display and had a first person view over a body-sized mannequin (Petkova & Ehrsson, 2008). The participant had visual and tactile stimulations over the whole body. The participants had the feeling that the other body was their own. The ownership feeling was measured through skin conductance, which can detect psychological or physiological alteration. Authors stress the fact that a human-like representation of the mannequin and synchronous visuo-tactile stimulation are crucial to trigger the ownership illusion.

Another pertinent example of ownership is when the individual is using artifacts. Studies showed that when an individual is using a tool, he/she is considering it as an extension of his/her arm (Maravita & Iriki, 2004). This previous study was performed with non-human primates and their brain activity was measured. The results show that some specific bimodal neurons coding the monkey's hand fire the same, when a stimulus is applied to the hand or close to the tool manipulated by the animal. This is strong evidence showing that the artifact seems to be integrated into the primate body schema.

This feeling can also happen in a remote environment, where the individual uses a tool at a distance (Sumioka, 2012). In this experiment, participants remotely controlled a human-like machine. They had a first person view over the machine, which replicated every move of the participant. The participants' reactions were measured with skin conductance and the results showed that the participants felt that the machine was their own body.

2.4 Technologies for inducing telepresence

2.4.1 Teleoperation

Teleoperation happens when the user controls a machine over a distance. In fact, the user is using two machines: one is operated by the user (a controller) and the other (the device) receives and executes commands from the controller. Visual control of the distant machine is done by looking at images received from a camera mounted on it. This allows the user to operate machines where it would be harmful or physically impossible for someone to be, such as, to handle toxic residues, to explore the bottom of the sea (Saltaren et al. 2007) or, to perform telesurgery.

However, because the visual supervision of the device is performed with a camera, the visuo-motor control from the user to the device needs to be well learned and calibrated. Peters et al. (2003) showed that six trials on a teleoperated robotic arm are necessary to have an accurate representation of the task.

In a study carried out by Moore, Gomer, Pagano and Moore (2009), the user controls a robot and supervises the environment by means of a camera on top of the robot. The operator's task is to judge whether or not the robot can pass through apertures of various sizes. The results indicate that the participants judged the robot could pass even when it could not. Authors argue that the results may vary regarding the distance between the robot and the aperture and may also be influenced by the height of the camera.

These examples show that, on the contrary of a simple artifact held in hand, the remote control and representation of an electronic device is not straightforward and involves a specific design of the machine to be ergonomically adapted to the human operator.

2.4.2 Virtual environments

A virtual environment can be defined as a three dimensional computer-modeled environment that allows user interaction. One of the most common applications is in simulators. Simulators are popular because they are safe, relatively cheap, and have various applications in entertainment, training, education and even rehabilitation. Notable examples: flight simulators used for pilot training, simulators for military operations, and the gaming industry. One aspect that can increase the immersiveness of virtual reality is the type of interaction between human and machine. If the interaction is done through a classic controller such as a mouse or joystick, the users need to learn the mapping between their own movement and its consequence in the virtual environment (Wise & Murray, 2000). However, the mapping is facilitated if the interaction is done through a natural user interface (NUI). Research has shown that natural user interfaces, in which users can recognize their own movements in the virtual environment, are more immersive (Bruder et al., 2009; Francese et al., 2012).

2.4.3 Avatars

Avatars are very common in virtual environments. It is the alter-ego of the user in the virtual environment. The avatar only exists in a specific virtual environment, and will possess certain intrinsic characteristics, such as its visual appearance, dynamics, etc. Every one of those characteristics may influence the way a person feel towards the avatar (Castronova, 2003).

The avatar is usually seen with a first person perspective or with a third person perspective. Since the control that the user has on the avatar influences the behavior of the user (Meadows 2008), such interaction may be studied on different aspects, such as the avatar’s dynamics or the hardware used to control the avatar.

2.5 Tools for measuring telepresence

2.5.1 Questionnaires

There are several ways to measure telepresence. One way the evaluation can be done is by questionnaire, in which the user answers a few questions in order to express what they felt during the experiment. This is probably the most popular way to evaluate presence. A significant number of experiments on telepresence or ownership have their participants to fill a questionnaire when the experiment is over (Botvinik & Cohen, 1998; Petkova & Ehrsson, 2008; Ehrsson et al., 2004, Maselli & Slater, 2013, etc).

Questionnaires are mostly used because of the simplicity of their implementation and the infinite range of possible questions. They are also a very quick and practical way for people to express their feelings. This way, feelings can be quantified and compared. Although, there are some disadvantages, such as misinterpretation of a question, subjectivity of the answers, the scale level number (odd vs even) or since it happens after the experiment, participants might forget what they felt. Another disadvantage is the number of questions: if there are too many, the participants may start to answer superficially to the questions.

2.5.2 Physiological parameters

Another way to evaluate presence is by physiological parameters such as heart rate, galvanic skin response, electromyography or electroencephalogram. The galvanic skin response (GSR) measure the skin conductance of electricity. The variation of skin conductance occurs by changes in the moisture of the skin. Emotional stimulus triggers the sympathetic nervous system to increase the activity of the sweat glands. Armel & Ramachandran (2003) performed a rubber hand illusion experiment, in which they measured the skin conductance. On the experiment, they threatened to harm the rubber hand, and if the participant thought that the rubber hand was his/her own, the skin conductance results would show signs of arousal.

The electromyography (EMG) measures the electric potential of neurologic activated muscles. The signals can also be used to detect neurologic activity. Slater et al. (2009) performed the rubber hand illusion experiment in the virtual world. When ownership was achieved, the virtual hand was twisted. The EMG form the participants showed that the twist induced motor activity on their real arm.

The electroencephalogram (EEG) directly measures electric activity of the brain in a non-invasive way. González-Franco et al. (2011) proved that EEG can be used to measure presence in a virtual environment. They conducted an experiment in the virtual world in which the arm of the avatar was threatened, and when facing a threat, the participant lowered the motor cortex activity.

2.5.3 Behavioral assessments

Telepresence can be also measured by behavioral assessments. It is based on the behavioral analysis of the individual, during his/her interaction with the surrounding environment.

Fitts' Law

One behavioral assessment very popular in human-computer interaction is the Fitts' Law (Fitts, 1954). This law can predict the time necessary to quickly move to a target area in function of the distance and the size of the target. This law is used to describe the act of pointing, either with the hands and fingers in the real world, or at an object on a computer screen using any sort of pointing device. One of the most common Fitts' Law equations is the Shannon equation (MacKenzie, 1992).

$$T = a + b\ log_2(\frac{D}{W} + 1) \qquad (1)$$

In (1), T is the time taken to complete the task; a and b are constants depending on the device used for human-machine interaction; D is the distance of movement from start to the center of the target; and W is the width of the target. This law has some consequences in terms of design of the human-computer interface. For instance, it suggests that it is better to put icons on the sides, bottom or top boundaries of a screen because of, in this configuration, their width can be considering as infinite, and consequently, faster to reach. Overall, this law implies that big targets at short distance are acquired faster than small targets at long range.

The 2/3 power law

The 2/3 power law (Lacquanity, 1983) is another behavior assessment, which links the kinematics of handwriting with a movement trajectory. This law relates the curvature of a trajectory $c(t)$ with the angular velocity $v_0(t)$ of the tip of the pen. In (2), k represents the slope of each segment.

$$v_0(t) = k\ c^{2/3}(t) \qquad (2)$$

This equation can be related to any drawing movement. Gribble and Ostry (1996) argue that the power law can be altered by neurophysiologic and biomechanical properties. This law was used to evaluate the sensorimotor appropriation or ownership of a teleoperated robot by a human operator (Rybarczyk et al., 2004).

Minimum jerk

Another behavioral assessment is the minimum-jerk. This method was introduced by Hogan (1984) and the jerk refers to the time derivative of the acceleration, which means a third time derivative of the position of the system. It can be applied for moving an arm or an object from one point to another smoothly. If an object has to be moved in a particular trajectory *x(t)*, from start time *ti* to end time *tf*, its jerk cost can be calculated by:

$$\int_{ti}^{tf} {\frac{d^3x(t)}{dt^3}}^2 \, dt \qquad (3)$$

The function x(t), from all the possible functions, is the one that has the least jerk cost is the minimum jerk. This function x(t) will be the function that most smoothly connects the start point to the end point in the specified time. In the scope of the assessment of the Human Machine Interaction (HMI), a high level of telepresence is expected to exhibit a smooth dynamic control of the artifact.

2.5.4 Affordances

In this project a behavioral assessment known as affordance will be used to measure telepresence in a virtual environment.

Affordances are a concept first suggested in the literature by Gibson (1979). An affordance is an action possibility whereby people perceive their environment and the objects within it as possibilities of doing certain actions and not doing other actions. Affordances exist where the characteristics of the object and the characteristics of the person match in a particular way. For instance, most chairs will afford sitting to most adults, but will not afford sitting to a 6-months baby, and might afford standing to someone making a speech. This can be applied even among other animal species, like for example a tree can afford nourishment to a giraffe but for a bird it can afford nesting. An affordance is a combination of the physical characteristics of the object and the person, the knowledge about the object, and the needs to the person at a particular time. In some cases, the action possibility may be harmful, in which case the person may choose not to perform the action. For example a knife affords cutting into various surfaces because it has a blade. If someone grasps it by the handle it affords cutting into bread or cut through paper but it also affords injury if grasped by the blade. Another example is about apertures. An aperture will only afford passage if it is wider than the individual. If it is narrower it will only afford passage if the individual performs a rotation upon himself (Warren & Whang, 1987). Affordances are based on experience, in the sense that people learn to perceive the relevant characteristics of the environment and objects within it. This means they will be common to many individuals (e.g., passing through apertures which are large enough) but different from one individual to another (e.g., a rugby player, a gymnast, or a child will fit through different apertures).

After the initial study by Warren and Whang (1987) testing affordances, other studies have followed which explore and test the notion of affordances (e.g., Mark, 1987; Esteves, de Oliveira, & Araújo, 2011). One crucial finding was that the possibilities for action available to an individual are scaled to the individual's body. This scaling factor is important because it links object properties and individual's dimensions through an invariant value; this means there is a lawful relation underpinning (at least some) affordances. Such lawful relations have been found in various animals. In human participants, this was found in stair climbing where participants deem a stair climbable (without the aid of hands) if the raiser is smaller than 0.88 their leg length (Warren, 1984). This was also found to be the case in passing through apertures where participants rotate their shoulders over their longitudinal axis if the aperture is smaller than 1.4 the width of their shoulders (Warren & Whang, 1987). In terms of HMI, Rybarczyk et al. (2012) have shown that affordances are a concept that can be applied to assess the ownership of a robotic arm by a teleoperator.

3 Proposal

This chapter explains the details of the proposal, based on the research in the previous chapter, intended to be used in the development of a Multi-Modal system. This proposal consists of a system architecture and a prototype. The system to be developed is intended to make the proof of concept in what concerns the telepresence the user feels within a game environment through the comparison of characteristics of the avatars that represent the user in such environment.

The main purpose of this study is to evaluate the effect of an avatar on the feeling of telepresence on a virtual environment. Several aspects could be studied regarding the avatar, such as its dynamics, morphology or physical appearance. In addition, aspects such as the camera perspective of the avatar or a visuomotor feedback from hardware might change the feelings of telepresence. The aspects that were studied on the avatar were its morphology and the dynamics of its movements.

The morphology of the avatar is an important aspect for achieving body ownership. For instance, Tsakiris & Haggard (2005) carried out the experiment of the rubber hand illusion with a fake hand and with a stick instead of the hand. Their results have shown that with the rubber hand, ownership was easier to achieve than with the wooden stick. If these results show ownership with a hand, it can also be applied in a larger scale to a scaled-to-user avatar. This idea was previously applied by Petkova and Ehrsson (2008) in the real world with a camera on a mannequin.

The dynamics of the avatar may also help inducing the feelings of telepresence and ownership on its user. If there is real-time congruence between the movements of the user and the movements of the avatar, then the feelings of telepresence and ownership should be greater than with incongruent movements. This fact is perfectly described in Kalckert and Ehrsson (2008). In this experiment, authors have shown that the rubber hand illusion can be induced through a simple visuomotor correlation, without the necessity of a tactile stimulation as it was used in the original study of Botvinik and Cohen (1998).

There are two experimental conditions. In one of the conditions, the avatar is morphologically proportional to the corresponding participant. It is possible to have a dynamic avatar fully proportional to the user thanks to a full body natural user interface. In the other condition, the avatar resembles the first one in how it looks, but it is always the same (standard) for every participant in this condition. In addition, this standard avatar did not have a kinematic movement that exactly matched the participant’s movements, as it only moved sideways and rotated upon itself. Aside from the avatar condition, there was also a speed condition: fast and slow.

3.1 System architecture

The MATTA (Model for Avatar Telepresence Testing with Affordances) is the proposed model. MATTA is a concept of a virtual environment in which the user controls an avatar, and some of its parameters are tested with affordances. For MATTA to be materialized it needs a modeling software, a game engine and a natural user interface (NUI).

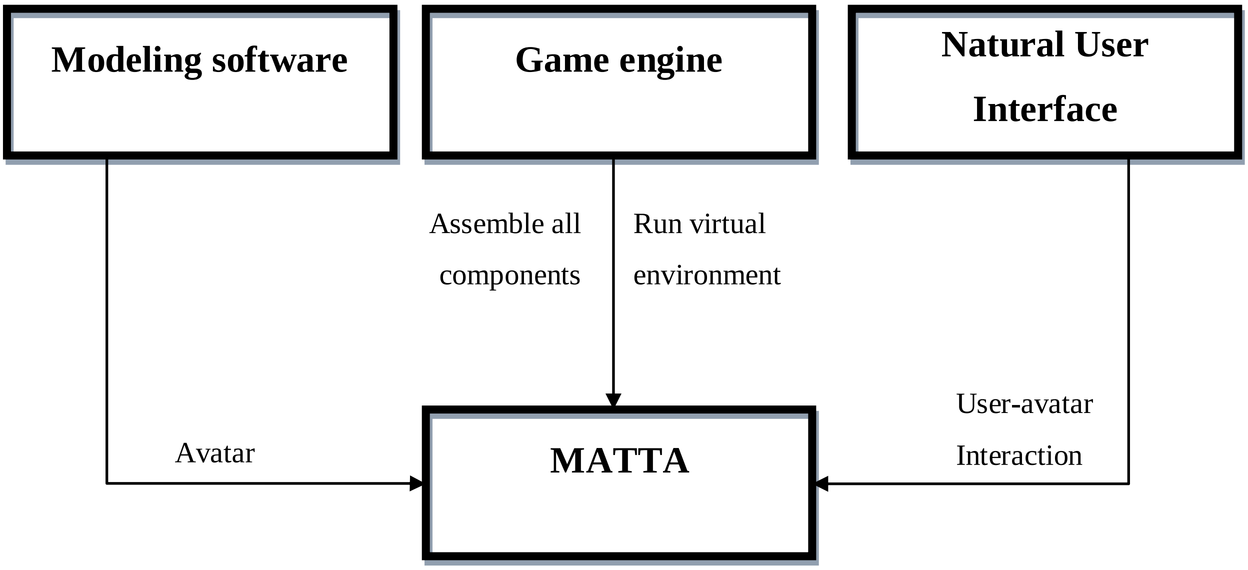


Figure 3-1- System architecture of MATTA

To design the avatars, modeling software has to be used. With this kind of software, objects such as the avatars are modeled and created. Only with modeling software the avatar can have a skeleton, which is important to allow movement and create animations.

To capture the movements of the participants, a full body motion capture has to be used. With a full body motion capture it is easy to capture the accurate position of every joint of the participant. This technology allows tracking the movements of the participant and process the data in real time. This means the avatar can replicate the participant’s movements. In this project it is important to have an interface that does not have delay and should allow the participants to have a visual response of their actions in real time.

A game engine is an essential tool for virtual environments. In this project, the game engine was used to assemble all the parts of the application, such as the avatars and the interface. The avatar can easily be imported from modeling software. The game engine is also important for modeling the virtual environment and run the application.

Modeling Software

- Model the avatar
- Build a skeleton on the avatar

Game Engine

- Model the environment
- Make the integration of all components

Natural User Interface

- Allow the participants to control the avatar with their own body

Figure 3-2- Components of MATTA

In order to test telepresence in a virtual environment, a good parameter is one that the participant is not aware of. These parameters are known as behavioral assessment. Using them as an evaluation parameter is an advantage because people perform actions that they are not aware of them, such as affordances, or movements that can be described with equations. As these behavioral assessments are present in the real world, telepresence in a virtual world can be related if the same behavioral assessments are also present there.

The options regarding technology and behavior assessment were chosen to be balanced as a very good solution in both aspects. Several aspects needed to be taken in consideration when studying avatars, such as the virtual environment, the interface, and the method to perform the study.

3.2 Prototype

The ATTAVE (Avatar Telepresence Testing: Affordances in Virtual Environments) is the name of the prototype developed based on MATTA. The ATTAVE is a virtual environment where the avatars exist and where experiments using affordances are performed. The virtual environment (ATTAVE) is the same for all the conditions of the avatar. This way, the only parameters that change along the experiment are the avatars’. The design of the prototype is based on a study performed by Warren & Whang (1987). In this study the authors evaluate how people pass through apertures considering the shoulder width of the participant and the ratio between the aperture and the shoulder width. The participants passed through several apertures of various sizes and the angle made by the shoulders was recorded. The apertures started by being narrow and got larger, and then got narrow again. Consequently, a same aperture width is tested twice during an experimental trial. There were two speed conditions: a slow and a fast walking speed. Results have shown that the participants only walked frontally through the aperture when the ratio of the aperture with the shoulders was smaller than 1.4. The present study is similar to the one performed by Warren & Whang (1987), but, this time, it is performed in a virtual environment. Relying on a study performed on the real world, telepresence can be verified if the same behaviors that occur in the real world are reproduced in the virtual environment.

The display of ATTAVE consists in a virtual scenario showing a long treadmill that moves towards a visible avatar (and also towards the participant). The avatar resembles a wooden mannequin and was visible from head to knees as the viewpoint of the participant was 2 m behind the avatar. On the treadmill, there are frontal green walls with an aperture on the left, centre or right side of the wall. Only the apertures located on the center are used for data collection. These center apertures varied in size proportionally to the shoulder width of the participant from 0.7 to 2.2. The apertures located on the side are dummy apertures, which are always twice the shoulder width of the participant. The treadmill is enclosed on the side by tall walls. All surfaces have texture (see Figure 3.3 and 3.4). The participants could control the translation and rotation of the avatars by moving side to side and rotating their shoulders. Shoulder rotation proportionally decreased the speed of the treadmill. The task for participants is to avoid collisions and pass through all doors as fast as possible.

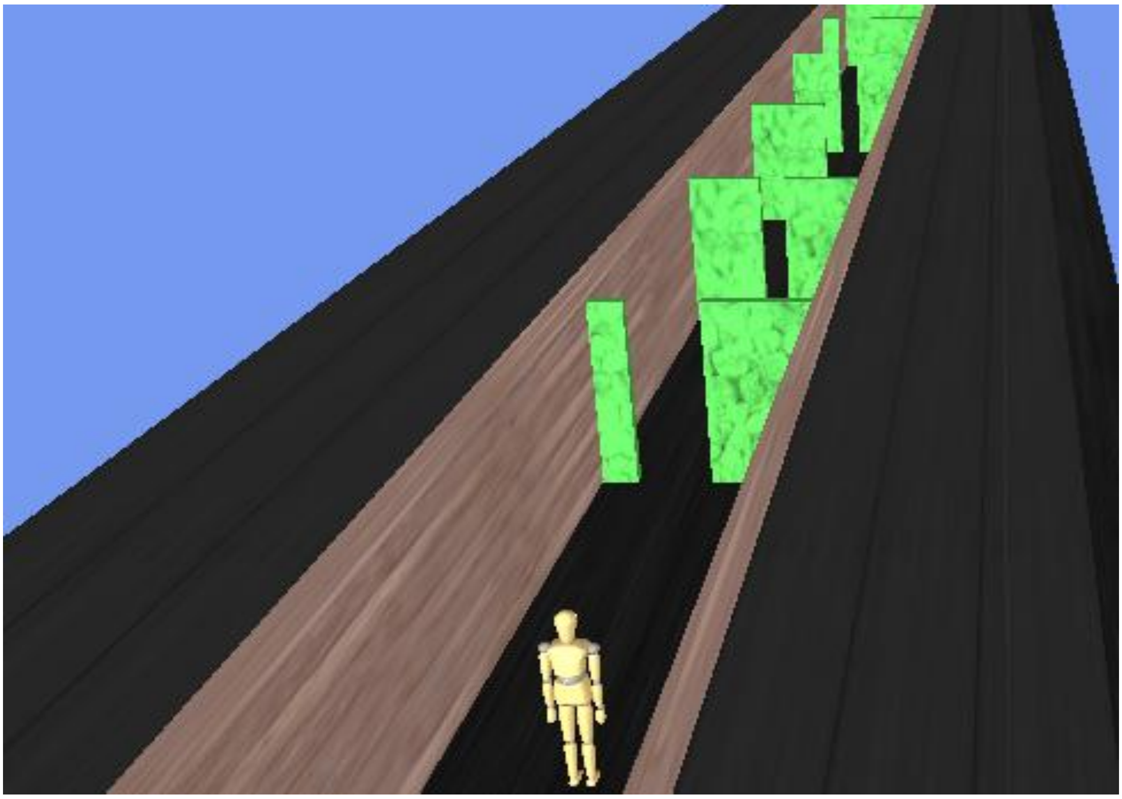


Figure 3-3- ATTAVE seen from above (this was not the view used in the study)

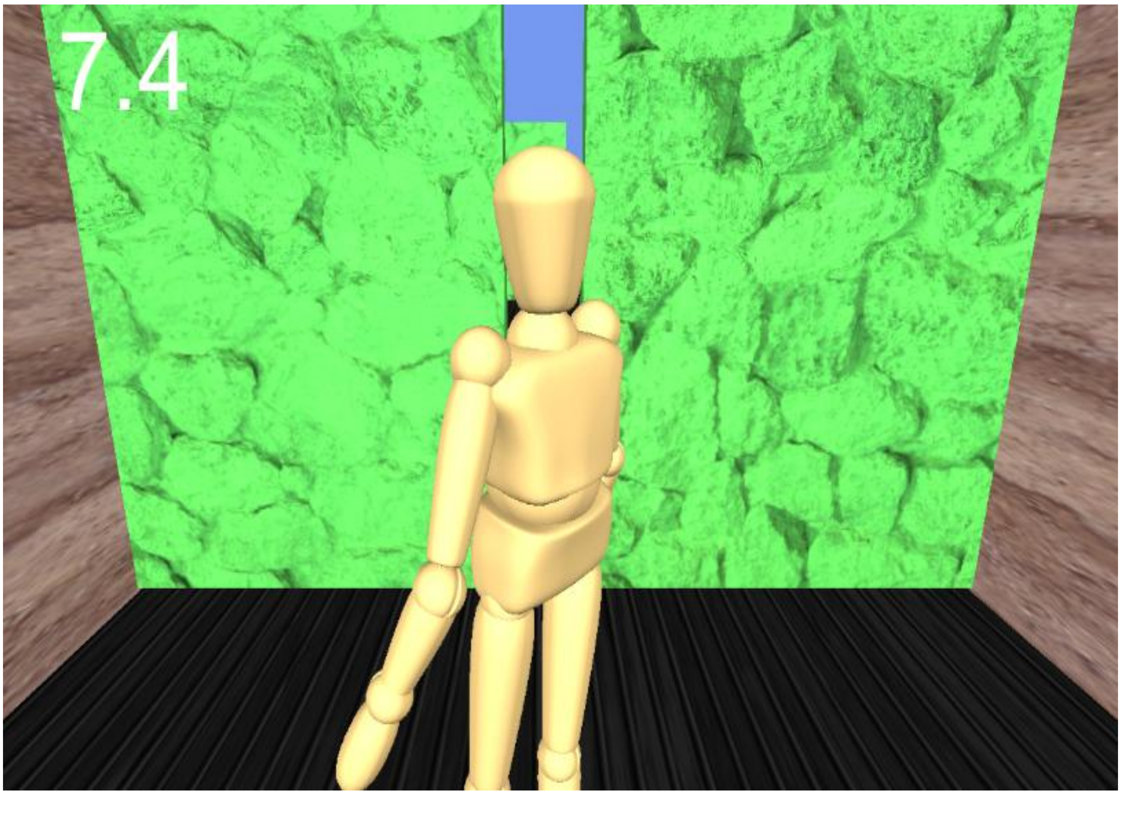
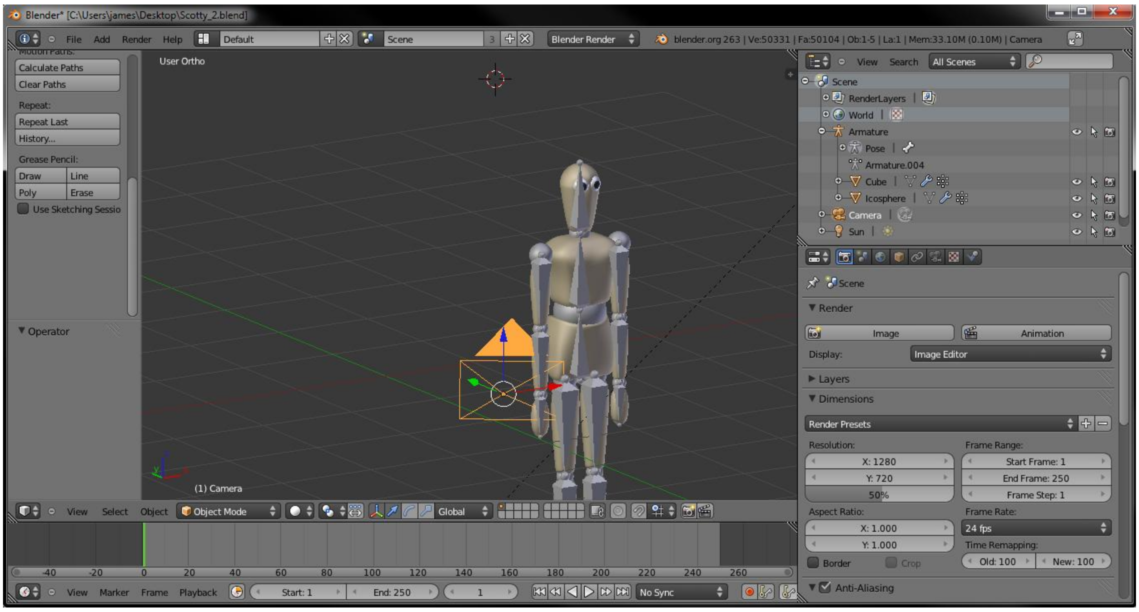


Figure 3-4- Virtual avatar and environment from the participants view

For modeling the avatars, Blender was the chosen software to perform the task. The avatars that were modeled were quite simple, as they resemble a wooden mannequin. As the avatars were simple models, Blender adjusted perfectly, considering the fact that it also a free software. To create the animation the software used was iPi Recorder and iPi Mocap Studio (iPi Soft). This software was used to create animation on avatars because it can record the movements captured by the Kinect and associate those movements to the avatar. With this method, it was created an animation of a sidestep with a human model performing the action, so the animation results of a natural human movement.



Resolution: X: 1280, Y: 720, 50%

Frame Range: Start Frame: 1, End Frame: 250, Frame Step: 1

Aspect Ratio: X: 1.000, Y: 1.000

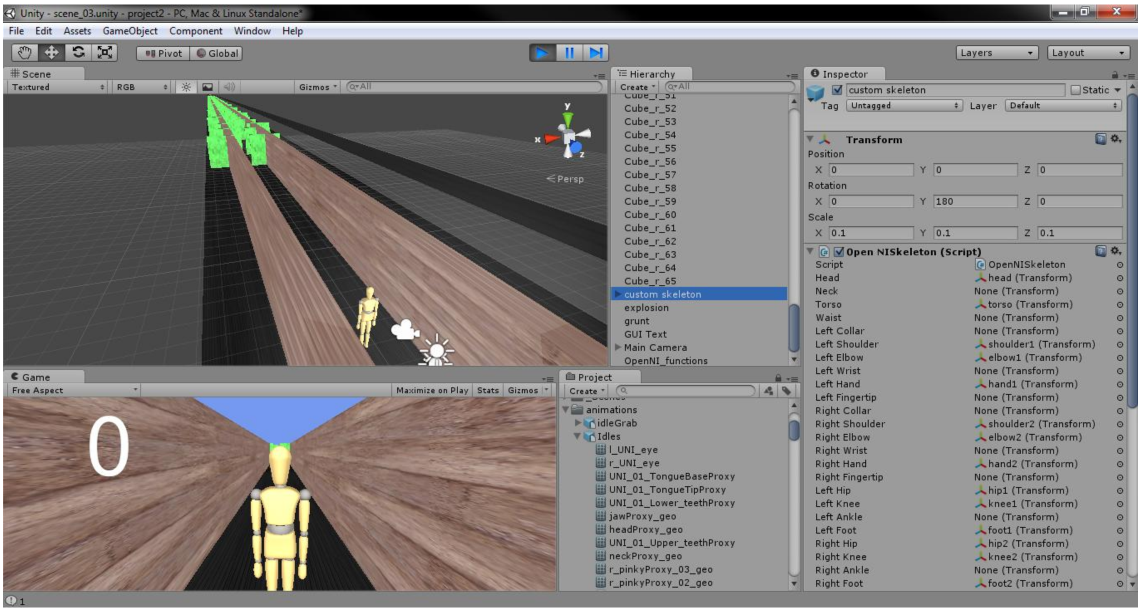
Frame Rate: 24 fps

Time Remapping: Old: 100, New: 100

Start: 1, End: 250, 1

Figure 3-5- Blender development environment

As for the game engine, the software chosen was Unity 3d for designing ATTAVE and running the application. MonoDevelloper was used as an IDE (Integrated Development Environment) to facilitate the scripting part of the game. The scripting was made entirely in a Java scripting language for Unity. One of the main reasons that lead to the use of Unity 3d in the project was its easy integration with the Kinect, which happens due to a framework called OpenNI(Open source Natural Interaction) provided by Primesense. The free version of Unity 3d was used because the needed features were all included in this version.



☑ custom skeleton ☐ Static

Tag: Untagged Layer: Default

Transform

Position X 0 Y 0 Z 0

Rotation X 0 Y 180 Z 0

Scale X 0.1 Y 0.1 Z 0.1

☑ **Open NISkeleton (Script)**

Script	OpenNISkeleton
Head	head (Transform)
Neck	None (Transform)
Torso	torso (Transform)
Waist	None (Transform)
Left Collar	None (Transform)
Left Shoulder	shoulder1 (Transform)
Left Elbow	elbow1 (Transform)
Left Wrist	None (Transform)
Left Hand	hand1 (Transform)
Left Fingertip	None (Transform)
Right Collar	None (Transform)
Right Shoulder	shoulder2 (Transform)
Right Elbow	elbow2 (Transform)
Right Wrist	None (Transform)
Right Hand	hand2 (Transform)
Right Fingertip	None (Transform)
Left Hip	hip1 (Transform)
Left Knee	knee1 (Transform)
Left Ankle	None (Transform)
Left Foot	foot1 (Transform)
Right Hip	hip2 (Transform)
Right Knee	knee2 (Transform)
Right Ankle	None (Transform)
Right Foot	foot2 (Transform)

Figure 3-6- Unity 3d environment

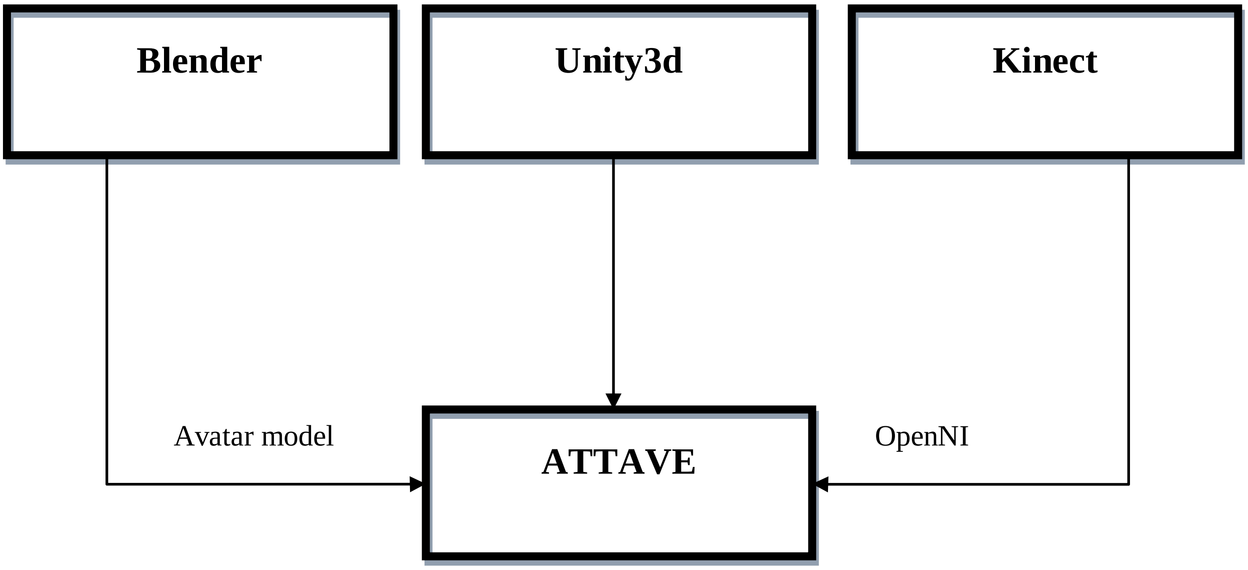


Figure 3-7- Diagram with the tools used to create ATTAVE

As a full body motion capture, the Kinect was chosen. This choice was due to the easy way to set up the equipment, which can be ready to use in less than five minutes, and also to its low cost compared to other equipment of the same category. The freedom of movements was also a positive aspect that influenced the choice of the equipment. The Kinect retrieves a skeleton model of the user, in which every joint position of the user can be accessed. Since the

Kinect only detects the center point of the shoulder the maximal shoulder width of the avatar was calculated based on anatomical data. In order to make the participant aware of the virtual space, a sound and a small graphic on the shoulders was introduced whenever there was a collision with the walls (that create the aperture).

When the participant rotated the shoulders, the speed of the treadmill decreased in proportion to the cosine of the angle between the shoulders and the moving walls, which increased the runtime of the task as per the formula below.

$$\text{current speed} = \text{set speed} * (1 - 0.4 * \sin(\text{angle})) \qquad (4)$$

In (4), the value of the angle is taken as 0 if the individual is in a frontal position towards the door and 90º if the individual is facing the side walls. The angles are in absolute values from 0º to 90º. This decreasing of speed with rotation was introduced because it also happens in a real environment. The decreasing value of 0.4 used to calculate the current speed based on pilot trials.

On the screen, the participant could see on the top left corner a number corresponding to the current speed. This way, the participant could have a visual feedback regarding his/her motion velocity, in order to keep the velocity as high as possible (see figure 3.4).

4 Experimental study

On this chapter there will be a description of all the methods and materials used to achieve the experimental results. Later in this section, the results will all be presented and then discussed.

4.1 Materials and Methods

4.1.1 Participants

Participants were 24 university students (18 male and 6 female, aged between 20 and 28), with normal or corrected-to-normal vision and varied experience in playing video games. The participants were quite distinct from each other as they differed in background, ethnicity and citizenship. This was possible to accomplish since some of the participants were participating in the eNTERFACE’13 conference, in which this project was recruiting volunteers. With this, the used sample represents a multicultural population.

Half of the participants performed in two conditions (similar fast, similar slow) and the other half performed in two conditions (standard fast, standard slow). This was done to enable the study of eventual learning effects under each speed condition. Table 4.1 shows the morphological data of participants on the first condition (similar avatar) and Table 4.2 shows the data for the participants in the second condition (standard avatar).

Table 4.1 -Morphological data of the participants in the first condition

Participant	Height(cm)	Shoulder width(cm)	Age	Gender
1	155	35	21	F
2	179	46	26	M
3	173	45,5	26	M
4	169	39	28	F
5	164	42	27	M
6	162	42	20	F
7	182	44	27	M
8	171	45	26	M
9	202	51	28	M
10	169	36	23	F
11	189	43	27	M
12	179	45	26	M

Table 4.2- Morphological data of the participants in the second condition

Participant	Height(cm)	Shoulder width(cm)	Age	Gender
13	170	42,5	23	M
14	173	47	24	M
15	173	41	25	M
16	182	47,5	24	M
17	170	41	25	M
18	185	46	29	M
19	176	43	27	M
20	163	42	27	F
21	188	44	30	M
22	175	44	26	M
23	185	44	29	M
24	168	39	27	F

4.1.2 Setup

The experiment was conducted in a 3 × 3 m area. Participants stood 3 m away from a 75 cm height table. On the table was mounted an off-the-shelf Kinect sensor (Microsoft, for Xbox

360) and an 18" computer screen (1440 × 900 pixel resolution), both connected to a PC. Three small marks on the floor indicated the positions aligned with the three apertures on the display (see Figure 4.1).

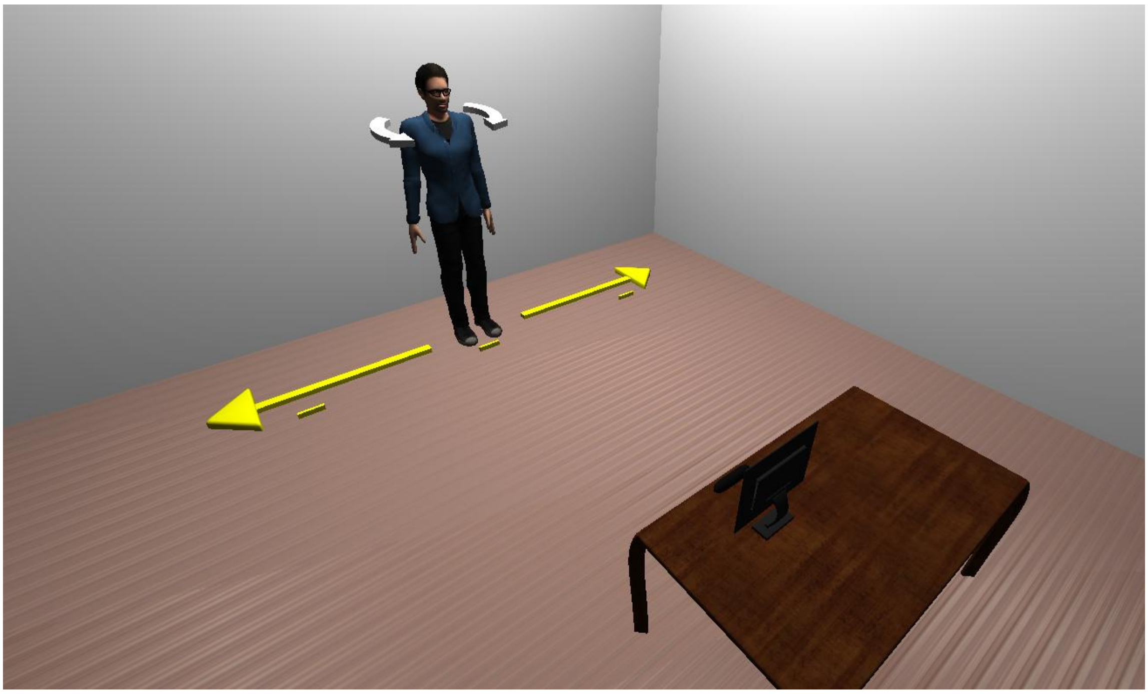


Figure 4-1- Physical setup of the experiment

4.1.3 Experimental design

Participants performed 32 trials for each of 2 speed conditions and for each of 4 sessions. Also, there were 2 avatar conditions; each used in a group of participants. The 32 trials consisted of apertures that showed in the central position with widths gradually increasing relative to the avatar’s shoulders from 0.7 to 2.2 and then gradually decreased from 2.2 to 0.7 (in steps of 0.1). When the avatar passes through each of these apertures, the value corresponding to the angle between the shoulders is recorded. These trials were alternated with 32 dummy trials with apertures of constant size shown in the right and left side positions. These side apertures were twice the shoulder width of the avatar. These side openings were not used for data collection. Every aperture is 10 meters away from the next aperture. The two speed conditions were slow and fast (respectively 5 and 10 Km/h) and were chosen following the walking speeds reported by Warren and Whang (1987) and pilot testing. The avatar condition consisted of manipulating the morphology and movements of the avatar. In the similar avatar condition the avatar was anatomically proportional to the dimensions of the participant and all segments were animated to mimic the natural movements of the participant. In the standard condition the dimensions of the avatar were standard for all participants and the avatar only had two degrees of freedom: translation sideways and rotation upon itself. An animation of a sidestep was implemented on the avatar when the participant performed a sidestep. This animation was recorded with a natural user interface and results from a sidestep performed by a human being.

The ratio between each virtual door and the avatar's shoulders width was the independent variable manipulated. The dependent variable was the angle between the shoulders upon the passage of each aperture.

4.1.4 Procedure

The experiment started with participants reading and signing the consent form. Then, the Kinect was calibrated to the participants' movements. Participants were instructed to avoid collisions and complete the test in the shortest possible time, and were informed that shoulder rotation proportionally decreased the speed of the treadmill. In each session, participants completed the increasing-decreasing series in the slow condition followed by the fast condition. Participants were asked to fill in a questionnaire adapted from Witmer et al. (2005). Finally, the measure of participants' height and shoulder width was taken. In total, each session lasted about 20 minutes.

4.2 Results

4.2.1 Data analysis

The main dependent variable was the critical ratio after which the participant passed without rotation. Following Warren and Whang (1987), the two values of the critical ratio of shoulder rotation from the increasing-decreasing series were averaged. The critical ratio was that with an angle smaller than 16º and after which all angles were smaller than 16º (one exception was permitted provided the angle was smaller than 40º and the average angles remained smaller than 16º). A critical ratio was calculated for each participant, condition, and session and these were used in the data analysis.

The statistic methods used to perform this study were MANOVA, ANOVA and the Pearson's r. An ANOVA is an analysis of variance. The ANOVA is the method used to compare two or more conditions simultaneously and can only have one dependent variable. This method tests if there is any statistical difference between independent variables. The MANOVA is a multivariate analysis of variance. This method is similar to the ANOVA but more complex.

It can test multiple dependent variables which may contribute to the same independent variable. Pearson's r is a statistic method that defines if two variables are connected through a linear relationship. This method retrieves a value between 0 and 1, being 1 the result of a very strong relationship and 0 a very weak one. Pearson's r specifies if the relationship is positive or negative. In a positive relationship both variables increase or decrease together whereas in a negative relationship the increase in one variable is accompanied by the decrease in the other variable.

To examine learning effects from session to session, the critical ratios were submitted to a multivariate repeated measures analysis of variance (MANOVA) with the factor session (4 levels), and using the 4 conditions as measures (slow-similar, fast-similar, slow-standard, fast-similar). Based on the results of this analysis, the averages of the last 3 sessions were used in the remaining analysis.

To examine the effect of conditions on the critical ratios, the individual critical ratios from the last 3 sessions were averaged and submitted to a repeated measures analysis of variance (ANOVA) with factors speed (2 levels: slow and fast) and avatar (2 levels: similar and standard). The same analysis was conducted for the total durations, that is, the time that the participants took to complete the test by going through all the apertures. The same analysis was also used to test the effects on collisions.

To examine how participants felt regarding the experienced environment, the scores for each dimension of the questionnaires were averaged and submitted to a multivariate repeated measures analysis of variance (MANOVA) with factors avatar (2 levels: similar and standard) and session (4 levels) and using the 5 dimensions as measures (realism, possibility, quality, ownership, and self-evaluation).

Finally, it was used Pearson's r to test the correlation between critical ratios and the dimension of ownership as measured by the questionnaire.

4.2.2 Learning of critical apertures

Overall, there was a significant learning effect on the critical apertures, $F(12, 86) = 6.85$, $p < .001$, $\eta^2 = .49$. This was reflected in the four conditions: slow-similar, $F(3, 33) = 17.74$, $p < .001$, $\eta^2 = .62$; fast-similar, $F(3, 33) = 3.08$, $p < .05$, $\eta^2 = .22$; slow-standard, $F(3, 33) = 8.74$, $p < .001$, $\eta^2 = .44$; and fast-standard, $F(3, 33) = 3.07$, $p = .05$, $\eta^2 = .22$.

Pair wise comparisons showed significant differences between the first and the last three sessions which did not differ between them (see Figure 4.2).

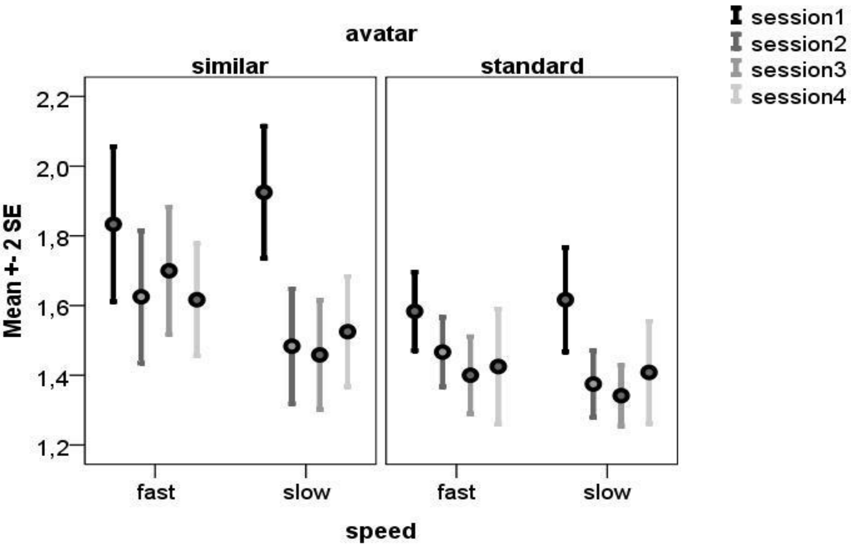


Figure 4-2- Average ratio for the four conditions over the four sessions

4.2.3 Critical apertures

There was a strong tendency for an effect of critical aperture on avatar, F(1, 11) = 4.62, p = .055, η^2 = .30, which was caused by participants rotating their shoulders at smaller critical ratios when the avatar was standard than when the avatar was similar (standard M = 1.40, se = 0.05, similar M = 1.57, se = 0.07).

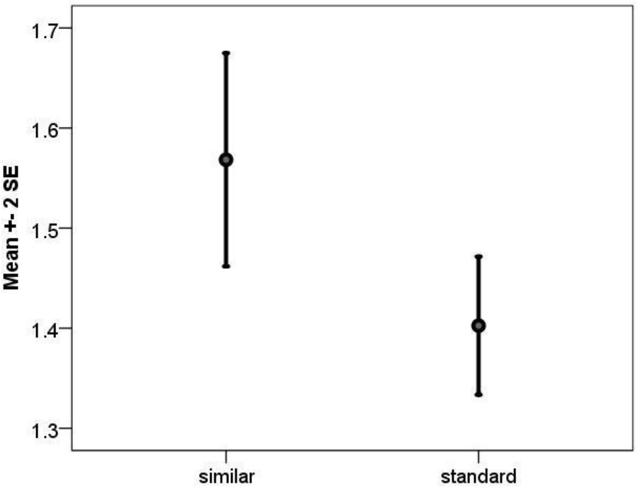


Figure 4-3- Effect of the avatar on the critical aperture

On Figure 4.4 it can be seen a comparison between the turning angles and the correspondent ratios for both similar and standard avatars. The red dots are data from the similar avatar, and the blue dots are from the standard avatar.

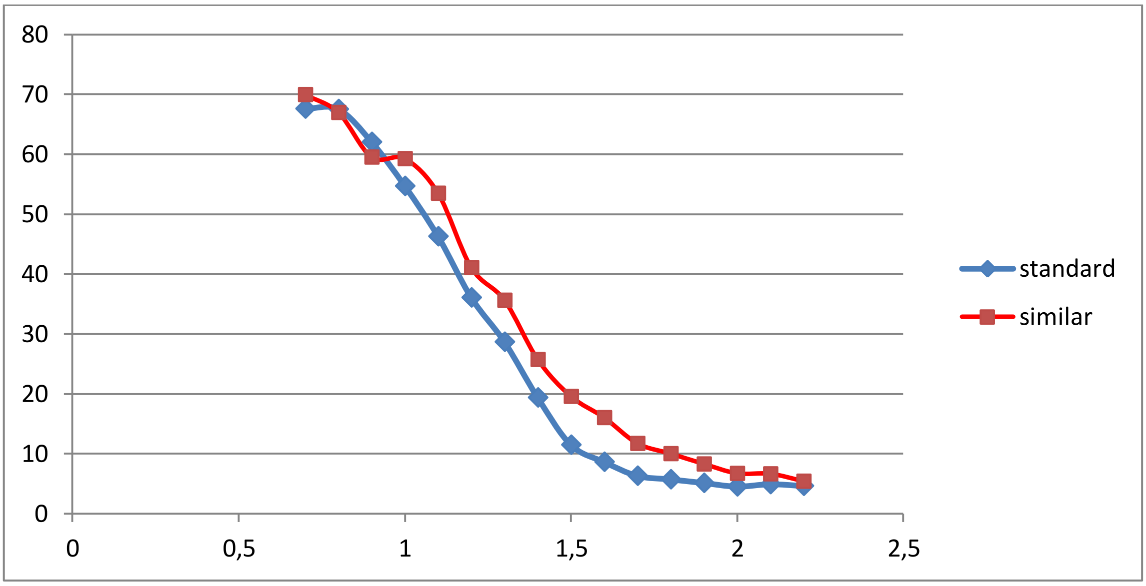


Figure 4-4 –Effect of the avatar on critical aperture (all participants)

4.2.4 Speed

There was a significant main effect of critical aperture on speed, $F(1, 11) = 5.13$, $p < .05$, $\eta^2 = .57$. This was caused by participants rotating their shoulders at smaller critical ratios in the slow condition compared to the fast condition (slow $M = 1.43$, $se = 0.04$; fast $M = 1.54$, $se = 0.05$).

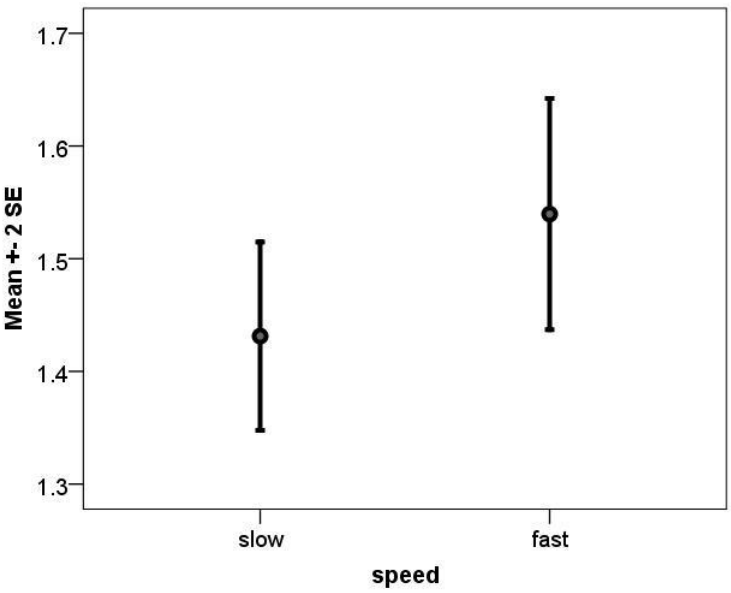


Figure 4-5- Effect of speed on critical aperture

There was no Speed × Avatar interaction, F(1, 11) = 2.51, however it is noteworthy that the effect of speed was more marked when avatars were similar than when avatars were standard(Figure 4.7 shows speed on x-axis and ratio on the y-axis, lines represent avatar condition).

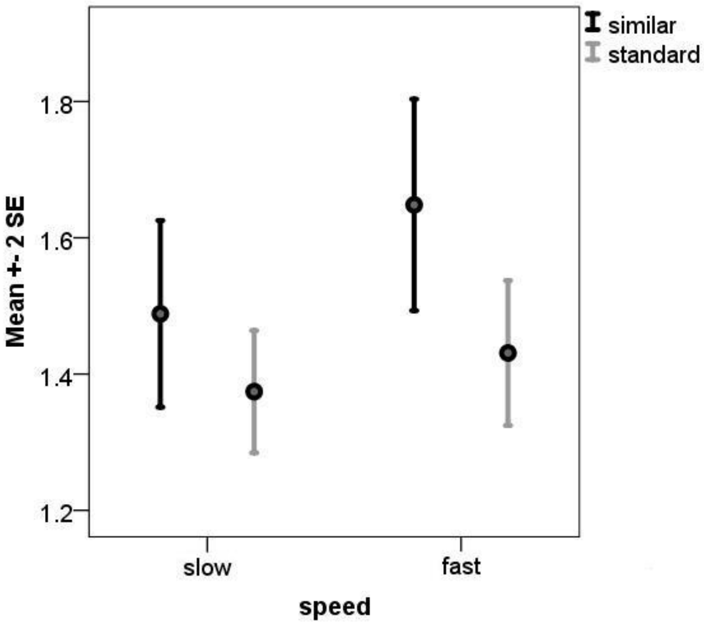


Figure 4-6- Effect of speed and avatar on the critical ratio

On figure 4.7 it can be seen a comparison between the turning angles at different ratio apertures and the two test speeds. The ratios obtained with the fast speed are in red and the slow speed is represented in blue.

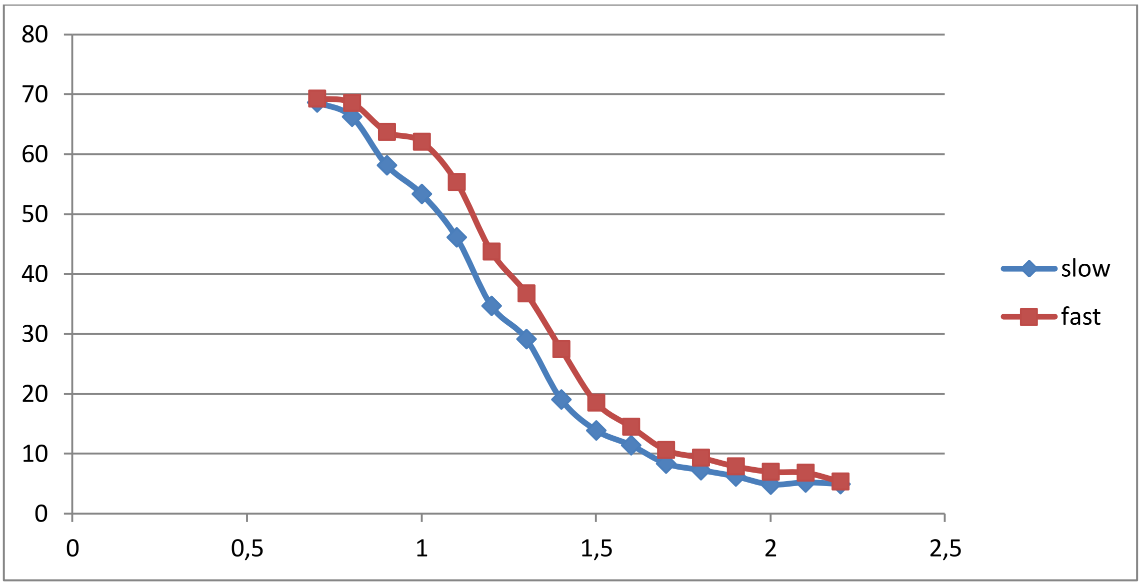


Figure 4-7- Effect of speed on critical ratio (all participants)

4.2.5 Duration

The main effect of duration on speed, $F(1, 11) = 7.07$, $p < .001$, $\eta^2 = .99$, was caused by the condition itself (slow $M = 476.5$, $se = 2.7$; fast $M = 248.3$, $se = 2.2$). The main effect of avatar was not statistically significant, $F(1, 11) = 1.93$, although participants took slightly shorter in the standard than in the similar condition (standard $M = 359.3$, $se = 4.06$, similar $M = 365.6$, $se = 1.6$). There was a significant Speed × Avatar interaction, $F(1, 11) = 19.09$, $p < .001$, $\eta^2 = .63$. This was because in the slow condition, participants performed slower with similar avatars (slow standard $M = 468.3$, $se = 5.03$, slow similar $M = 484.7$, $se = 2.05$), whereas in the fast condition participants performed faster with similar avatars (fast standard $M = 250.2$, $se = 4.42$, fast similar $M = 246.4$, $se = 1.32$).

4.2.6 Collisions

There were no significant main effects of collisions on speed, $F(1, 11) = 0.0$, avatar, $F(1,11) = 2.19$, and no significant interaction effect, $F(1, 11) = 0.40$. On average there were 2 collisions on each session across conditions.

4.2.7 Questionnaire

Overall, there was no significant main effect of avatar, F(2, 7) = 1.83 or session, F(2, 7) = 0.93. However, there was a significant Avatar × Session interaction, F(15, 80) = 1.87, $p < .05$, $\eta^2 = .24$. This significant interaction was reflected in three dimensions: realism, F(3, 33) = 4.00, $p < .05$, $\eta^2 = .27$; ownership, F(3, 33) = 3.93, $p < .05$, $\eta^2 = .26$ and self-evaluation, F(3, 33) = 3.17, $p < .05$, $\eta^2 = .22$. This interaction occurred because feelings of realism, ownership and self-evaluation increased in the similar condition and decreased in the standard condition.

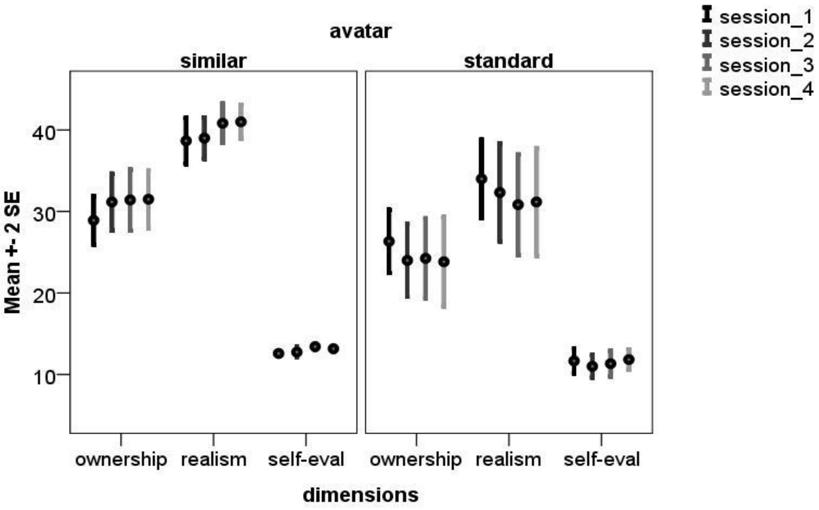


Figure 4-8- Questionnaires results for the three dimensions over sessions

4.2.8 Ownership and critical ratios

Overall, there was a positive correlation between feelings of ownership and critical ratios, r = 0.36, n = 96, $p < 0.05$. A scatter plot summarizes the results (Figure 4.7). There was also a small, positive correlation between these two variables indicating that increases in one were accompanied by increases in the other.

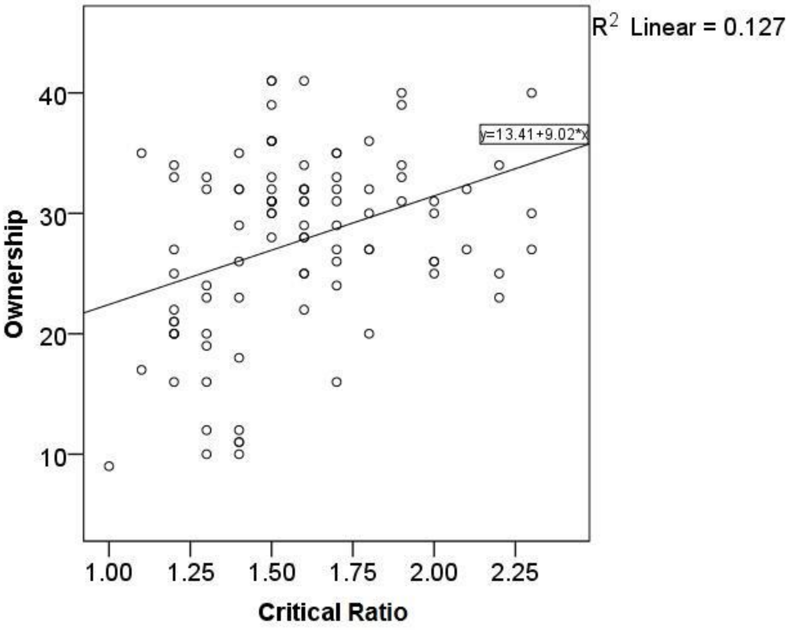


Figure 4-9- Scatter plot from Pearson's r

4.3 Discussion

The main objective of this research was to study whether the dynamics and the morphology of an avatar would reflect on the feelings of telepresence and ownership on the participant. A secondary purpose was to know whether affordances were used in a virtual environment the same way they are in a real environment. In order to perform the experimentation, a prototype named ATTAVE was developed. ATTAVE possess a virtual environment where the participant's avatars were performing the experimental tasks. The user controlled the avatar through a natural motion capture carried out by a Kinect NUI. There were two avatars conditions to test: one morphologically proportional to the user and that replicated his/her movement dynamic; and another with an avatar that was identical for every participant, and exhibited a limited mobility, as it could only rotate upon himself and move sideways.

The results show that participants adjusted to the virtual environment after taking their first session. This learning effect is only significant in the first session, and the other three sessions were similar to each other in the same condition. The learning effect that occurred in the first session indicates that participants are learning to use appropriate information, provided by the virtual environment, to solve the problem of passing through apertures. After learning, the critical ratios obtained in the virtual environment (1.4 and 1.57) were very similar to those in the real environment (1.4 reported by Warren and Whang, 1987), which demonstrates that people perceive similar body-scaled affordances in the virtual environment as in real environments.

The effect of different avatars was significant. Participants with the standard avatar had a smaller critical ratio than the participants with the similar avatar (1.4 vs 1.57). This may have happened because individuals who participated in the trials with the similar avatar were more immersed into the virtual environment and the avatar, which lead to leaving a greater safety margin. This can be supported by the results of the correlation between ownership and critical ratios. That relationship between ownership and the critical ratios exists and the feelings of ownership increase when the ratios also increase.

The higher ratio in the similar condition comparing to the standard condition could be also explained by the fact that the participants got distracted with their own movements, which were exactly identical to the avatar's movements, this is, the amount of freedom in movements might have lead the subjects to concentrate in something other than the shoulders and the apertures. The movements of the standard avatar were much more restricted (moving sideways and rotating), which decreased the amount of information needed to control the avatar. In the standard condition, the avatar was controlled similarly to a teleoperated machine, this is, the avatar could be controlled with the same complexity of a joystick. This scenario can be understood as similar to teleoperation on a virtual environment. On a study performed by Moore et al. (2009), in teleoperated robots, there was a miscalculation of the apertures. Participants in this teleoperation experiment believed that the aperture was bigger than it really was. The same effect could be happening in the virtual world, since the avatar was controlled similarly to a machine. The subjects might have unconsciously thought they were controlling a machine and not an avatar.

Analyzing the effect of speed on the critical ratio, it is observable that in both avatar conditions the critical ratio is larger when the speed is higher. This happens because at higher speeds people leave a larger safety margins. In the real world, when an individual is confronted with an aperture, he will reduce his speed in order to fit through without hitting. In this project, the only way to decrease the speed was to rotate the shoulders, which resulted in a higher critical ratio. The relation between speed and accuracy is well-known in the area of motor control and was described for the first time as Fitts' law (Fitts, 1954). If participants in this study used the same motor control principle, this indicates that they were immersed in the virtual environment.

The questionnaires showed that the similar avatar elicited an increasing feeling of ownership and realism over the four sessions. In contrast, the standard avatar elicited a decreasing feeling of ownership and realism from session to session. This could mean that with a similar avatar, the participants feel every time more at ease with the avatar. With the standard avatar people could become more bored due to the avatar’s movement not being so dynamic. Also, the lack of dynamism could lead the participants to act as if they were controlling a machine instead of a virtual representation of themselves.

Considering the questionnaires results regarding realism and ownership, the avatar with the same natural movements as the user is more immersive than a standard avatar.

5 Conclusions and future work

5.1 Conclusions

The main purpose of this thesis was to understand how the avatar could influence people's behavior in a virtual environment. The participants interacted with the environment through a Kinect. It has been studied how the avatar's morphology and movements affect the completion of the task. The task of the participants was to evaluate whether they could fit on apertures in a moving virtual environment or not and turning their shoulders if they thought they could not.

The results have shown that the morphology and natural movements of the avatar have an importance on the way people feel towards the avatar. The data from the virtual environment demonstrated that people pass through apertures as similar as in a real environment. The questionnaires have shown that the sense of ownership and the realism of the avatar are greater when it has the same movements and morphology as the human user.

This project is one of the first to test the morphology and movements of an avatar using a Kinect (NUI). With this study, it can be said that using avatars with a natural user interface in which the avatar's movements and morphology are similar to the user is important in getting him/her more immerse and present in the virtual environment.

5.2 Future work

In this project, it has been studied the morphology and the natural movements of the avatar. Many other aspects can be approached such as the effect of the user's perspective of the avatar on the feelings of telepresence. It can be studied whether the user feels more immerse when he/she has a first person view over the avatar than with a third person view. Some recent studies have already suggested that a first person view is crucial to trigger a strong feeling of ownership in the real world (Petkova &Ehrsson, 2008) or in virtual reality, as well (Maselli & Slater, 2013). The influence of the height or the distance from the camera to the avatar can also be a matter of study. In teleoperation, Moore et al. (2009) suggest that the camera height may alter the perception of passability of a teleoperated robot through apertures. This can be also tested on avatars in a virtual environment. Maselli & Slater (2013) also argue that an incongruent perspective over the avatar might alter the feelings of ownership.

Another interesting study is to test whether or not the visual resemblance between the avatar and the user may influence telepresence. In this project, it has been only tested the body proportion of the avatar regarding the individual, but not the avatar physiognomy. For example, would it change the feelings of telepresence and ownership if the avatar is the same genre/race, has a similar face/hair or wears clothes with the same color as the individual?

Regarding hardware, it can also be analyzed whether more immerse equipment, such as a virtual reality goggles or a haptic suit may enhance the feeling of avatar ownership. For example, if the individual wore a head-mounted display and had a first person view over the avatar, would it be easier to deceive the individual into accepting the avatar as his/her own body? Or, would the ownership increase if there were a haptic suit that provided force feedback on the shoulder upon impact? There are emerging affordable technologies in the area of head mounted displays, such as Oculus Rift (Oculus VR) and Wrap1200VR (Vuzix).

If all previous test possibilities were to be done with ATTAVE, a lot of cross-references could be studied and, since the virtual environment is kept the same, it will be possible to compare different avatars' parameters.

From an entertainment perspective, since the virtual environment simulated a treadmill, an application can be created for use with a real treadmill. For example in gyms, when people use a treadmill, which can be a monotonous activity, they can play a game with a natural user interface, in which they have to avoid obstacles and/or reach for objects in the virtual world while running.

Bibliography

Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1523), 1499-1506.

Bonk, C., & Dennen, V. (2005). Massive multiplayer online gaming: A research framework for military training and education (Technical Report No. 2005-1). Washington, DC: US Department of Defense (DUSD/R): Advanced Distributed Learning (ADL) Initiative.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756-756.

Bruder, G., Steinicke, F., & Hinrichs, K. H. (2009, March). Arch-explore: A natural user interface for immersive architectural walkthroughs. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on* (pp. 75-82). IEEE.

Castronova, E. (2003). *Theory of the Avatar*. Cesifo.

Clark, Herbert H. "Using language. (1996)." *Cambridge University Press: Cambridge)* 952 (1996): 274-296.

Csikszentmihalyi, M. (1975). Beyond boredom and anxiety. San Francisco, CA: Jossey-Bass.

Ehrsson, H. H., Holmes, N. P., & Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *The Journal of Neuroscience*, *25*(45), 10564-10573.

Ehrsson, H. H., Spence, C., & Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science*, *305*(5685), 875-877.

Esteves, P.T., de Oliveira, & Araújo, D. (2011). Posture-related affordances guide attack in basketball. *Psychology of Sport and Exercise*, *12*(6), 639-644.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*(6), 381.

Francese, R., Passero, I., & Tortora, G. (2012, May). Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 116-123). ACM.

Gibson, J.J. (1979). The ecological approach to visual perception. Houghton Mifflin: Boston.

González-Franco, M., Peck, T. C., & Slater, M. (2011) Virtual Embodiment Elicits a Mu Rhythm ERD When the Virtual Hand is Threatened.

Gribble, P. L., Ostry, D. J., Sanguineti, V., & Laboissière, R. (1998). Are complex control signals required for human arm movement?. *Journal of Neurophysiology*, *79*(3), 1409-1424.

Hoffman A. H., Glynn R. C., & Ault H. K. (2006). The Effect of Restricted Ankle Motion on Stair Ambulation. In Proc. of 29th RESNA Conference, 2006.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *The Journal of Neuroscience*, *4*(11), 2745-2754.

Kalckert, A., & Ehrsson, H. H. (2012). Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Frontiers in human neuroscience*, *6*.

Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta psychologica*, *54*(1), 115-130.

Li Q, Zamorano L, Jiang Z, Gong JX, Pandya A, Perez R & Diaz F.(1999). Effect of optical digitizer selection on the application accuracy of a surgical localization system - a quantitative comparison between the OPTOTRAK and FlashPoint tracking systems. Comp Aid Surg 1999*;* 4*:* 314–321.

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

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, *7*(1), 91-139.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in cognitive sciences*, *8*(2), 79-86.

Mark, L.S. (1987). Eyeheight-scaled information about affordances: a study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 361-370.

Maselli, A., & Slater, M. (2013). The building blocks of the full body ownership illusion. *Frontiers in human neuroscience*, *7*.

Meadows, M.S. (2008). *I, avatar*. Berkley, CA: New Riders Press.

Minsky, M. (1980). Telepresence. *Omni*, *2*(9), 45-51.

Moore, K. S., Gomer, J. A., Pagano, C. C., & Moore, D. D. (2009). Perception of robot passability with direct line of sight and teleoperation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(4), 557-570.

Park, S. & Hwang, H. (2009) Understanding online game addiction: Connection between presence and flow. In Human-Computer Interaction. Interacting in Various Application Domains, volume 5613 of Lecture Notes in Computer Science, pages 378–386. Springer Berlin /Heidelberg, 2009.

Peters, R. A., Campbell, C. L., Bluethmann, W. J., & Huber, E. (2003, September). Robonaut task learning through teleoperation. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on* (Vol. 2, pp. 2806-2811). IEEE.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: perceptual illusion of body swapping. *PloS one*, *3*(12), e3832.

Roetenberg D., Luinge H., & Slycke P., (2013) “Xsens mvn: Full 6dof human motion tracking using miniature inertial sensors,” Xsens Technologies.

Rybarczyk, Y., Mestre, D., Hoppenot, P., & Colle E. (2004). "A biological model for the evaluation of human-machine adaptation." *AMSE Modelling, Measurement and Control. 65,* 23-33.

Rybarczyk, Y. (2010). "A low cost optical tool for a 3D markerless motion capture." 33th European Conference on Visual Perception. Lausanne, Switzerland.

Rybarczyk, Y., Hoppenot P., Colle E., & Mestre D(2012). "Sensori-motor appropriation of an artefact: a neuroscientific approach." *Human Machine Interaction - Getting Closer.* Ed. Maurtua Inaki. Rijeka: InTech. 187-212.

Rybarczyk, Y., & Mestre D. (2013). "Effect of visuo-manual configuration on a telerobot integration into the body schema." *Le Travail Humain, 76(3),* pp. 181-204.

Saltaren, R., Aracil, R., Alvarez, C., Yime, E., & Sabater, J. M. (2007). Field and service applications-exploring deep sea by teleoperated robot - An underwater parallel robot with high navigation capabilities. *Robotics & Automation Magazine, IEEE*, *14*(3), 65-75.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence*, *3*(2), 130-144.

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience*, *3*(2), 214.

Sumioka, H., Nishio, S., & Ishiguro, H. (2012). Teleoperated android for mediated commmunication: body ownership, personality distortion, and minimal human design. *Social Robotic Telepresence*, 32.

Takatalo, J. (2002). Presence and flow in virtual environments: An explorative study. Master's Thesis, University of Helsinki, Helsinki, Finland.

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance, 31*(1), 80.

Tsakiris, M., Prabhu, G., & Haggard, P. (2006). Having a body *versus* moving your body: How agency structures body-ownership. *Consciousness and Cognition*, *15*(2), 423-432.

Warren Jr, W.H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 683-703.

Warren Jr, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 371.

Wise, S. P., & Murray, E. A. (2000). Arbitrary associations between antecedents and actions. *Trends in neurosciences*, *23*(6), 271-276.

Witmer, BJ, Jerome, CJ, & Singer, MJ (2005). The factor structure of the Presence questionnaire. *Presence, 14(3),* pp. 298-312.

Yoxall, A., Heller, B. & Chamberlain, P.(2011). Oom Bop Bop Good Vibrations: the use of sensory feedback to create motion inhibition. In: Include 2011, Royal College of Art, London, 18-20 April 2011.

Yuan, Y., & Steed, A. (2010). Is the rubber hand illusion induced by immersive virtual reality?. In *Virtual Reality Conference (VR), 2010 IEEE* (pp. 95-102). IEEE.

Appendix

Questionnaire

Name:
Date: __/__/2013
Session: / 4
Sex: M / F
Height:
Shoulder width:

Characterize your experience in the environment, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

WITH REGARD TO THE EXPERIENCED ENVIRONMENT

1. How much were you able to control your avatar?

|________|________|________|________|________|________|________|

NOT AT ALL	SOMEWHAT	COMPLETELY

2. How responsive was the environment to actions that you initiated (or performed)?

|________|________|________|________|________|________|________|

NOT	MODERATELY	COMPLETELY
RESPONSIVE	RESPONSIVE	RESPONSIVE

3. How natural did your interactions with the environment seem?

|________|________|________|________|________|________|________|

EXTREMELY	BORDERLINE	COMPLETELY
ARTIFICIAL		NATURAL

4. How much did the visual/auditory aspects of the environment involve you?

|________|________|________|________|________|________|________|

NOT AT ALL	SOMEWHAT	COMPLETELY

5. How natural was the mechanism which controlled movement through the environment?

|________|________|________|________|________|________|________|

EXTREMELY ARTIFICIAL	BORDERLINE	COMPLETELY NATURAL

6. How realistic was your sense of objects moving through space?

|________|________|________|________|________|________|________|

NOT AT ALL REALISTIC	MODERATELY	VERY REALISTIC

7. How much did your experiences in the virtual environment seem consistent with your real world experiences?

|________|________|________|________|________|________|________|

NOT CONSISTENT	MODERATELY CONSISTENT	VERY CONSISTENT

8. Were you able to anticipate what would happen next in response to the actions that you performed?

|________|________|________|________|________|________|________|

NOT AT ALL	SOMEWHAT	COMPLETELY

9. How completely were you able to actively survey or search the environment using vision?

|________|________|________|________|________|________|________|

NOT AT ALL	SOMEWHAT	COMPLETELY

10. How realistic was your sense of moving around inside the virtual environment?

|________|________|________|________|________|________|________|

NOT REALISTIC	MODERATELY REALISTIC	VERY REALISTIC

11. How closely did you feel that the avatar’s proportions fit yours?

|________|________|________|________|________|________|________|

NOT AT ALL	PRETTY CLOSELY	VERY CLOSELY

12. How well were you able to estimate distances?

|________|________|________|________|________|________|________|
NOT AT ALL SOMEWHAT PERFECTLY

13. How involved were you in the virtual environment experience (gauge your abstraction level regarding the surrounding real environment)?

|________|________|________|________|________|________|________|
NOT MILDLY COMPLETELY
INVOLVED INVOLVED IMMERSED

14. How much delay did you experience between your actions and expected outcomes? **(CAUTION: inverted scale)**

|________|________|________|________|________|________|________|
NO DELAYS MODERATE LONG
DELAYS DELAYS

15. How quickly did you adjust to the virtual environment experience?

|________|________|________|________|________|________|________|
NOT AT ALL SLOWLY LESS THAN
ONE MINUTE

16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

|________|________|________|________|________|________|________|
NOT REASONABLY VERY
PROFICIENT PROFICIENT PROFICIENT

17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities? **(CAUTION: inverted scale)**

|________|________|________|________|________|________|________|
NOT AT ALL INTERFERED PREVENTED
SOMEWHAT TASK PERFORMANCE

18. How much did the control devices interfere with the performance of assigned tasks or with other activities? **(CAUTION: inverted scale)**

|________|________|________|________|________|________|________|
NOT AT ALL INTERFERED INTERFERED
SOMEWHAT GREATLY

19. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities (= transparency)?

|________|________|________|________|________|________|________|
NOT AT ALL SOMEWHAT COMPLETELY

20. During the experiment there were moments in which you felt as if the virtual avatar was your own body.

|________|________|________|________|________|________|________|
STRONGLY NEUTRAL STRONGLY
DISAGREE AGREE

21. When the avatar hit the wall, how much did you feel that your own body hit the wall?

|________|________|________|________|________|________|________|
NOT AT ALL SOMEWHAT COMPLETELY

22. During the experiment there were moments in which you had the sensation of having more than one body.

|________|________|________|________|________|________|________|
STRONGLY NEUTRAL STRONGLY
DISAGREE AGREE