Universidad Zaragoza
1542

Master Thesis

Análisis y experimentación de efectos de degradación del rendimiento visual debido a estímulos auditivos en entornos virtuales inmersivos

Analysis and experimentation of visual performance degradation effects due to auditory stimuli in immersive virtual enviroments

Autor

Daniel Jiménez Navarro

Director

Ana Serrano Pacheu

Departamento de Informática e Ingeniería de Sistemas

Master Program in Robotics, Graphics and Computer Vision

Escuela de Ingeniería y Arquitectura
Universidad de Zaragoza
2022

Resumen

En los últimos años, el consumo y el interés de las personas por la Realidad Virtual (RV) están aumentando vertiginosamente. Esta tecnología innovadora ofrece una serie de capacidades vanguardistas a la vez que se ha vuelto más accesible debido su continuo desarrollo. En RV, el usuario es un elemento activo que interactúa de muchas maneras con el entorno virtual, lo cual difiere del su rol pasivo establecido en los medios tradicionales. Esta interacción ocurre de forma natural una vez el usuario se sumerge en el mundo virtual y los sentidos detectan lo que ocurre alrededor. De manera similar a la realidad, la percepción humana puede ser engañada o alterada bajo ciertas condiciones donde nuestros sentidos recompilan información contradictoria. De hecho, un efecto de supresión audiovisual fue presentado por Malpica et al. [1], en el cual se demostró cómo los estímulos auditivos pueden causar la perdida de información visual. El rendimiento visual del usuario se degrada cuando se escuchan sonidos espacialmente incongruentes pero temporalmente consistentes. Nuestro cerebro percibe estímulos tanto visuales como auditivos pero se pierde información visual debido a las interacciones neuronales.

El objetivo principal de este proyecto es analizar y conocer mejor este efecto de supresión audiovisual, concretamente su parte auditiva. Partiendo de la publicación anterior, creamos un entorno virtual en el que estímulos auditivos y visuales serán presentados al usuario. En cuanto a los estímulos auditivos, se investiga cómo pueden influir los sonidos ubicados en los límites del rango auditivo en la aparición de dicho efecto. Para ello, se obtienen los valores de frecuencia asociados a los límites de audición para cada usuario y se utilizan después en los sonidos generados. El participante percibirá tanto estímulos unimodales (solo auditivos o visuales) como estímulos bimodales (audiovisuales). Los estímulos bimodales se generan dinámicamente en ubicaciones fijas manteniendo la consistencia temporal, creando las condiciones adecuadas bajo las cuales se produce el efecto de supresión. Registrando cuando aparecen los estímulos y el desempeño del usuario al detectarlos, es posible comprobar si el usuario ha sufrido dicho efecto.

Los experimentos y el test de frecuencia han sido realizados por un grupo de 20 participantes. Los resultados obtenidos ponen de manifiesto que los ratios de detección y reconocimiento de los estímulos visuales son reducidos por sonidos casi inaudibles. Por lo tanto, el efecto de supresión audiovisual todavía se produce con estímulos auditivos situados en los límites de nuestro rango auditivo. Las encuestas realizadas por los participantes demuestran cómo la mayoría de ellos experimentó una gran sensación de inmersión y presencia en el mundo virtual. Además, no se aprecia que el experimento tenga efectos secundarios significativos ni inconvenientes que empeoren la experiencia virtual. Por último, se sugiere como trabajo futuro el análisis de los datos visuales registrados durante el experimento, con el fin de estudiar cómo se comporta el usuario cuando se perciben sonidos apenas audibles en RV. Con el objetivo de investigar el impacto que pueden tener otros factores como las emociones personales y el estado de ánimo en el efecto de supresión, se proponen también un par de dispositivos pertinentes.

Abstract

In recent years, both consumption and people's interest in Virtual Reality (VR) are increasing dizzily. This innovative technology provides a number of groundbreaking capabilities while has lately become more accessible due to continued hardware development. In VR, the user turns into an active element that can interact in many ways with the virtual environment, which differs from the user passive role settled in traditional media. This interaction occurs naturally once the user is immersed in the virtual world and senses detect what is happening around. Similarly to reality, human perception can be deceived or altered under certain conditions where our senses gather contradictory or too much information. In fact, an audiovisual suppression effect was reported by Malpica et al. [1] in which it was proved how auditory stimuli can cause loss of visual information. User's visual performance degrades when spatially incongruent but temporally consistent sounds are listened at once. Our brain perceives both visual and auditory stimuli although some visual data is lost due to neural interactions.

The main goal of this project is to analyze and get a better insight of this audiovisual suppression effect, more concretely, its auditory part. Using the publication previously mentioned as baseline, we create a virtual environment in which both auditory and visual stimuli will be presented to the user. Regarding auditory stimuli, we research how sounds located at the limits of our hearing range can influence the appearance of this effect. Therefore, frequency values associated to hearing limits are obtained for each user and will be used after in the sounds generated throughout the experiment. The participant will encounter not only unimodal stimuli (auditory or visual only) but bimodal (auditory and visual at the same moment) stimuli as well. Bimodal stimuli are dynamically generated in fixed locations keeping temporally consistency, creating the proper conditions under which the audiovisual suppression effect occurs. By keeping stimuli apparition record as well as user performance regarding the moments when any stimuli was perceived, it is possible to check if the user has suffered the suppression effect.

The experiments and the frequency test have been performed by a group of 20 participants. The achieved results manifest that detection and recognition rates of visual stimuli are indeed decreased by almost inaudible sounds. Thereby, the audiovisual suppression effect still occurs with auditory stimuli located at the limits of our hearing range. Surveys fulfilled by the participants demonstrated how the majority of them experimented a great feeling of immersion and presence in the virtual world. Besides, it is not appreciated that the experiment has significant side effects neither drawbacks that disturb participants while making the virtual experience worse. Lastly, it is suggested as future work the analysis of the eye-tracking data recorded during the experiment, in order to study how users behave when barely audible sounds are perceived in VR. With the aim of researching the impact that other factors, such as personal emotions and state of mind, may have on the suppression effect, a couple of appropriate gadgets are proposed as well.

Índice

List of Figures

List of Acronyms

AR	Augmented Reality
CSV	Comma Separated Values
dB	Decibel. Sound Intensity unit of measure equal to a tenth of a Bel.
ECG	Electrocardiogram
FOV	Field Of View
GPS	Global Position System
GSR	Galvanic Skin Response
GUI	Graphical User Interface
HMD	Head Mounted Device
HTC	High Tech Computer
Hz	Hertz. Unit of frequency in the International System of Units.
IPD	Inter Pupillary Distance
POV	Point Of View
PPI	Pixels Per Inch
RDW	Redirected Walking
SDK	Software Development Kit
UAS	Unity Asset Store
UP	Unity Project
VE	Virtual Environment
VR	Virtual Reality
XR	Extended Reality

1. Introduction

In this first chapter, we introduce the context where the project is located at, as well as its objectives, scope and methodology.

1.1 Virtual Reality

Nowadays, Virtual Reality (VR) is known to be a widespread field, which is getting increasing interest while having a shocking and very fast development. VR is a new tool that allow us to consume content differently from classic devices and traditional media. Unlike films where the user has no control over what is happening on screen, VR surpasses this barrier giving the user the control over the virtual environment and the freedom to interact with it [16].

Originally in last century, VR capabilities were restricted by computing power and hardware availability, whose production costs were very high and dimensions were not completely comfortable. An example of this hardware can be found in Figure 1.1 (left). In recent years, the quick development of microprocessors and graphics cards has triggered the improvement of computers efficiency and performance. This hardware upgrades accompanied by price reductions, were the responsible for general public to become more and more interested in this emerging technology [17]. We can see how different current and first VR headsets are in Figure 1.1, where a Samsung Gear VR headset is also displayed (right).

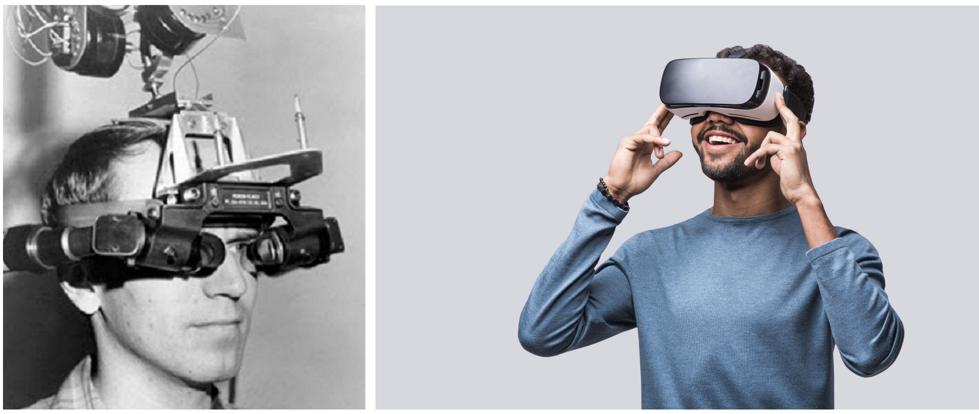


Figure 1.1: VR Headsets: Sword of Damocles (left), considered the first virtual and augmented reality real time viewer or HMD (Head Mounted Display) [2]; Samsung Gear VR headset (right), one of the currently available devices [3].

Consequently, the VR market, investment and consumption has notoriously increased as well, and its growth is expected to continue on rise during this next decade [4], as can be seen in Figure 1.2. Today, just using the required VR equipment, which is still quite expensive, we are able to visualize scenery located miles away without leaving our home, perform almost whatever task only with our own body and a limited area such as our house room, or even immerse in fantastic environments which do not exist in real world, like video games or films.

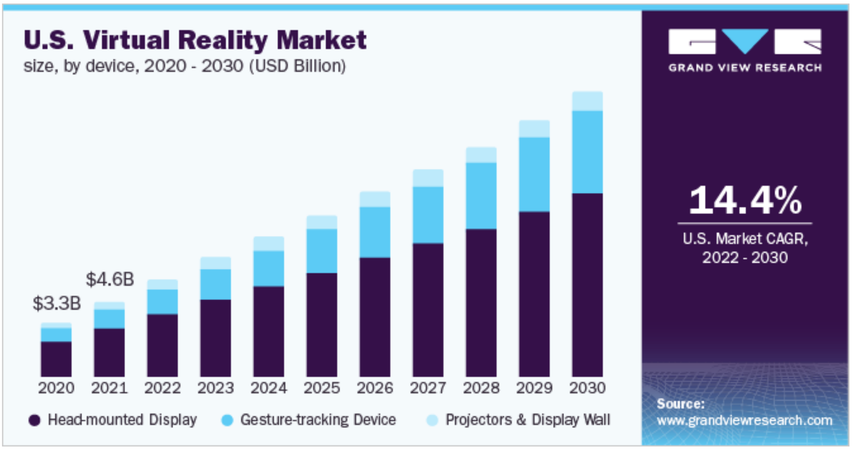


Figure 1.2: Expected Virtual Reality Market in U.S. through this decade [4].

One of the main causes of VR's huge versatility is multimodality support. Basically, stimuli of different modalities (sight, hearing, touch and so on) can be virtually generated such accurately that gives the user the sense of real immersion in the virtual environment, increasing realism and proprioception. These stimuli are detected by our senses and processed by our cognitive system, from whose point of view, there are no differences from the real world stimuli. In the future, we would not be able to distinguish one case from another ideally, which gives VR unprecedented potential [18].

Regarding research, VR advantages have been approached in very different fields: from engineering and architecture, where 3D designs can be virtually evaluated so there is not need to manufacturing them [19]; to education, where engaging environments can be used in learning processes more appropriate than real world [20]. It is also important the use of VR technology on medicine, where some of its applications achieve faster diagnostics and therapeutic treatment of mental diseases, surgical training, prosthesis design or pain management [21]. Other applications could be related to science, leisure, fashion, marketing, logistics, health and psychology among others [18].

1.2 Context and Objective

This project is in the frame of one of the several research lines that the Graphics and Imaging Lab is currently working on. The Graphics and Imaging Lab is a research group from Universidad of Zaragoza, founded by Prof. Diego Gutierrez, which focuses its research projects in hot-topics mainly related to computer graphics and computational imaging.

In general terms, actual projects cover from developing new physically-based rendering and visual appearance methods and models, to virtual reality applications and transient imaging solutions among others. Nowadays, the lab has many people with very different backgrounds working on it, keeps growing its numbers in terms of publications and members, and has achieved international recognition for the work of previous years [5]. The lab's logo can be found in Figure 1.3.

Figure 1.3: Graphics and Imaging Lab Logo [5].

Having introduced VR and the research group where our project is located at, we can already address the main goal of our work and its scope. The purpose of this work is to analyze, explore and evaluate a recently discovered visual degradation effect, which is triggered by auditory stimuli under concrete VR settings.

This visual degradation effect has been reported in *Malpica et al.* [1]. As it is observed, visual performance gets notoriously worse under certain and controlled circumstances, when spatially incongruent auditory stimuli exists. The effect is triggered by using several audiovisual stimuli cases while users' eye behaviour is recorded. It is finally concluded that this visual degradation effect is related to neural and attention working rather than eye motion [1].

Particularly, our contribution will be focused on designing, implementing and performing the necessary user studies and data analysis in order to get a further insight of this phenomenon. Our aim is to explore what factors might have an impact on this effect, such as what type of sounds can trigger it.

Once this suppression effect is better understood, we may be able to use it cleverly as an advantage to improve virtual reality applications. For example, subtle sounds could be used to change the virtual environment without user realizing, in order to overcome physical space restrictions or perform perception tasks.

1.3 Methodology and Schedule

Since this project has a high research load, the first step to take is studying the problem theoretically, involving all factors to consider in our virtual world. Later on, the necessary tools to solve it will be implemented. Therefore, we will firstly study some state-of-the-art publications regarding the virtual variables we need to approach in our project, which is explained in Chapter 2. These factors will allow us to get an idea of how the audiovisual effect may influence the user's performance and experience within the virtual environment.

Additionally, during this procedure, it is required the use of a Head-Mounted Display (HMD) and its associated devices, such as controllers and sensors, to get participants into a total virtual immersion. All this virtual-related gadgets and software are introduced in Chapter 3. Afterwards, we proceed to design and implement the experiments used to test this audiovisual effect. This part is explained throughout Chapters 4 and 5. Once the experiment is completely functional, a significant group of participants will be essential to carry on the tests and draw consistent conclusions from the results, as reported on Chapters 6 and 7.

Regarding the schedule followed along this assignment, it can be found in the Gantt Diagram shown in Annex A. To sum up, the process is as follows:

- Study previous works related to virtual reality and this audiovisual effect specifically. *Chapter 2.*
- Installation and Configuration of VR regarding hardware and software. *Chapter 3.*
- Design of new user studies considering different scene components and audiovisual stimuli. *Chapter 4.*
- Implementation of the experimental procedure and virtual reality environments necessary for the studies using the Unity framework. *Chapters 4 and 5.*
- Perform user's experiments, recording results and other relevant data (demographics, previous experience...). *Chapter 6*
- Analysis of the results and conclusions reached, proposing future work and applications. *Chapter 7.*

2. Previous Work

Next, we give an overview of the state-of-the-art of some key concepts related to our project.

2.1 Human Senses in VR: Multimodality

Conventional wisdom recognizes sight as the sense on which people most depend daily, considering it more important than other senses. Some studies report that this hierarchy of the senses is nothing more than the result of years and years of culture and life in society [22]. Both us and our ancestors find visual communication (and memory) as the most powerful and efficient media of interacting with the world around us. We just have to think on informative posters on the street, smoke signals used in ancient times or drawings on cave walls, considered forerunners of writing. Hence, it is understandable that the sense we value most is eyesight, about which there are much more research in comparison with other senses [22]. The same happens in VR, where visual media is usually the first contact with the virtual world. Visual displays, such as HMDs, are our door to virtual reality since they act on the dominating sense of human perception, and they are key to get a proper immersion feeling. HMDs are built with better capabilities every time so the gap with how we see real world reduces, but it may never disappear for obvious reasons [23].

On the other hand, hearing sense also plays an important role to the extent that we could match it to vision according to dependence level, as demonstrated by a public survey in United Kingdom [24]. This is because hearing can reach places we cannot see, either because they are very far away or because we do not have a clear line of sight, behind something opaque or ourselves for example. Similarly to vision, audio sense allows us to understand better our surroundings as well as it is vital in our spoken communication. Again, all this features are also applicable to virtual environments. Several works assess how audio improves virtual experience, such as how soundscape sounds affects the feeling of presence [25]. Other studies report the different effects of using mono, stereo, dolby surround and 3D audio in VR, being the latter the one inducing more sense of presence while providing information regarding the audio source location and distance [26]. It is important to include auditory information properly when other information, like visual, is present.

Multimodality plays an important role in VR applications, gathering different sensory stimuli and information to elaborate the general perception of the virtual scene, which improves the user's experience and immersion overall [18].

Since our brain is in charge of collecting the perceived information and respond to the virtual environment, similarly to a real world operation, it is crucial to keep spatiotemporal congruence with synchronized sensory stimuli [18]. In order to understand this idea better, we can imagine a couple of examples. In the case that the user is walking on the street and sees a car accident, the impact sound shall be synchronized with the moment when the two vehicles make contact and crash, neither before nor after. On the other hand, the user could stretch his hand to feel a rough surface. In this case, the feeling of the irregular surface on its own skin shall be synchronized with the moment when the virtual hand reaches the virtual surface.

Common to both cases is that keeping both spatial and temporal congruence when generating stimuli is essential, so they can be perceived as authentic as possible. In the first case, we mix auditory and visual information, while in the second one visual and haptic (tactile) stimuli are combined. An example of this second case can be found in Figure 2.1. There are other kind of stimuli feedback such as olfactory and gustatory, but limited to hardware capabilities [18]. It is capital to maintain spatiotemporal congruence in order to enhance the virtual experience, since multimodality misuse would be counterproductive otherwise. In the case of the car accident, a sound out of phase with the crash or originated elsewhere far away, would greatly reduce the feeling of realism and immersion.

Figure 2.1: Synchronized visual and haptic feedback are used to increase immersion feeling and can even generate body distortion illusions [6].

Regarding realism, another modality that shows up is proprioception. This term refers to the own-body motion sense, to consciousness of associating the real user's body as it belongs to the virtual world, which is associated to the feeling of actually *being* there. This immersion sensation also involves the credibility of the virtual environment so the user's expectations shall be satisfied and cannot be ignored. Therefore, both the virtual scene and the user representation (proprioception) set the realism level and help to improve the virtual experience [18].

Besides, multimodality can be used to affect and influence user's attention in the virtual world. There is a whole field of research whose aim is to analyze which areas attract more user's attention, or in other words, the saliency in regions of the virtual environment. It has been reported that those scene parts where more sensory information is present manage to attract the user's attention better. In this sense, multimodality techniques are starting to being considered to predict saliency in 360º panoramas as well as to guiding users attention through, for example, visual or auditory cues [18]. We can found in Figure 2.2 how multimodal information achieves better results when redirecting human attention than considering only unimodal information [7].

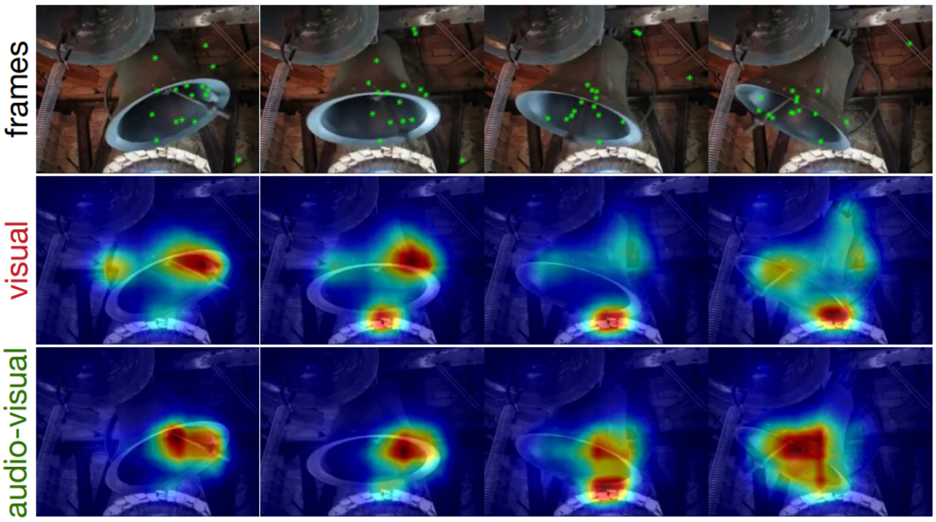


Figure 2.2: Saliency prediction example frames where the third row is the audiovisual (multimodal) case, which concentrates user's attention more than visual-only (unimodal) case printed in second row [7].

It has also been observed how multimodality techniques have an impact on the user's performance. Aside from skills and personal capabilities, a better representation of the real world allows to achieve better results by reducing completion time and improving accuracy in certain tasks [18]. Several works have reported how visual, haptic and auditory information can affect each other in detection tasks. The reason for this is supported on how our cognitive system works at full capacity when different stimuli are perceived by our senses, enabling a richer experience [27].

2.2 Presence Feeling in Virtual Environments

According to the aforementioned, multimodal virtual environments (VE) get a more accomplished immersion sense since they can include, for example, visual, haptic and gustatory information. Consequently, a relation among attention and presence is set, where higher awareness levels induce a higher feeling of presence [27].

However, the lack of full haptic feedback basically means that there is not a 'virtual skin'. As a result, the user cannot feel the virtual ground under his feet while walking or virtual wind gusts along the body. The lack of hardware means it is necessary to leverage other features in order to improve presence. Apart from sensory information, user can interact with the VE through some task so he will become more familiar with the environment, favoring this feeling of presence. For example, some of these interaction techniques are related to manipulation, locating the virtual object and changing its position and orientation using our virtual hands (controllers); or movement, taking into account the correspondence among real and virtual motion while also providing proprioception feedback [28].

Related to the latter, as we can imagine, the virtual world will surely differ from the real place where we are. The VE can be a leafy forest or a cliff on the beach while the real place is a fifteen-squared-meter room. This is where the term Redirected Walking (RDW) appears. As its own name tells, it consists in altering user's walking steps inadvertently. So, the performed path is within space limitations while the user thinks to be doing a different route. There are works that approach this topic presenting different solutions.

For example, it has been reported how to influence the user's motion through subtle alterations in the virtual scene benefiting from visual saccades, which are rapid eye movements associated to fixation points shift. Tracking eyesight, it is possible to take advantage of temporal blindness triggered by saccades and head motion to redirect users via imperceptible scene changes [8]. As we can see in Figure 2.3, altering the virtual scene without the user noticing it allows to redirect its movement subtly as we can see in the comparison between real and virtual paths (c). There are more mechanisms used in multimodal VE in order to adapt the scene to the task to perform, without neglecting the feeling of presence.

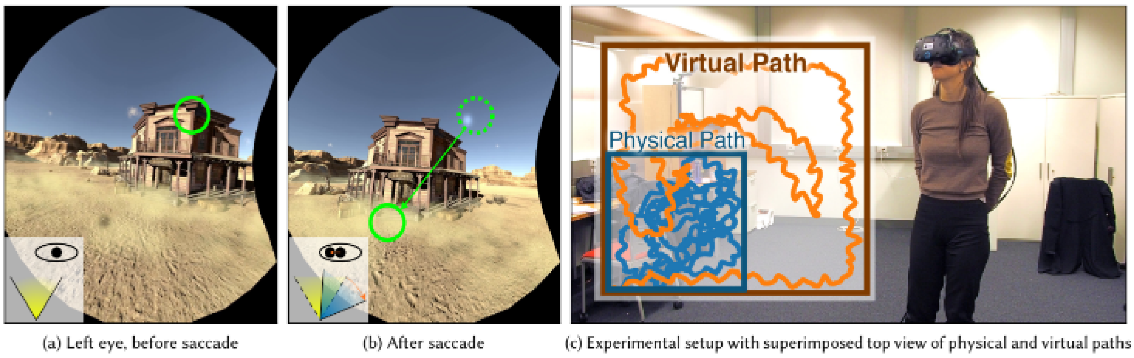


Figure 2.3: *a*) and *b*) Subtle scene alteration when eye gaze change between fixation points (saccades) causing temporal blindness. *c*) Comparison between physical and virtual paths, evidencing user motion redirection [8].

2.3 Illusions: Suppression Effects

Following the previous idea of RDW, multimodality can be used to trick user's perception of the VE through illusions. These illusions are basically facilitatory effects that are originated in our physiological or cognitive systems. In those cases, users perceive new stimuli that does not adjust to real external information [18]. Some illusion-related works are the famous Rubin's vase illusion [29] (Figure 2.4, left), where the intrinsic cognitive process was analysed to understand better how perception alters between two faces or a vase; or the rubber hand illusion [30] (Figure 2.4, right), where it is shown how our brain is tricked to consider the rubber hand as our own body by hiding our real hand and combining visual and haptic stimuli.

Not only did works report illusions or facilitatory effects examples on VR, but suppressive and inhibitory effects as well. The latter consist on losing sensory information due to simultaneous stimuli. We experience this suppression effect on a daily basis. For example, regarding the sense of hearing, louder sounds tend to suppress more quiet audio sources.

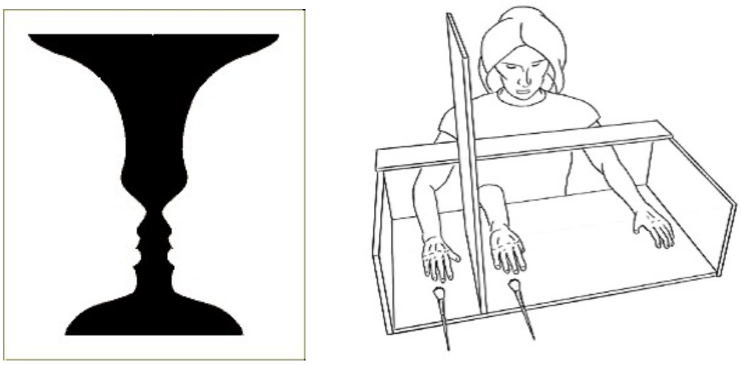


Figure 2.4: Rubin Vase Optical Illusion (left): Two interpretations are valid, perceiving two faces or a vase [9]. Rubber Hand Illusion (right): Touching simultaneously the rubber hand (visible) and the real hand (hidden) makes the brain to gain ownership of the rubber hand, until confusing touches on the rubber hand as real on the own hand [10].

We can also imagine ourselves in a living room with the TV on while having a chat with a relative, how our brain unconsciously suppresses the TV audio and focuses on what our interlocutor is telling us. Then, we may pay attention and listen to our relative while we will not remember what was playing on TV. Another example, in this case regarding the sense of sight, can be blinks and saccades. During these eye movements, our brain suppresses that blurry information without perturbing the perception process [18]. A similar suppression case occurs when our nose is ignored from the visual information captured by our eyes. In all those situations, the brain selects the information from the whole stimuli perceived, losing the less important or relevant content generally.

Until now, we have just only addressed unimodal suppression effects, auditory and visual in the previous cases, but this omission cases exist with crossmodal stimuli as well. Several works reported how perceiving synchronous multimodal stimuli leads to a worse detection of each one, compared to when they occurred standalone. Receiving information from several sources at the same time makes our brain to select which is more important. Consequently, incomplete or incoherent information, such as blurry images obtained from saccades or noise respectively, is suppressed. As a result, several suppression effects can happen even across sensory modalities.

Particularly, it has been demonstrated how visual orientation discrimination performance is degraded when haptic stimuli is applied to the user's hand. This conclusion proves that neural responses to crossmodal stimuli have a close and direct interaction with each other, what causes a visual suppression effect as long as both stimuli are spatially and temporal consistent [11]. An illustration of the setup used in those experiments can be found in Figure 2.5.

Conversely, other works also study how auditory stimuli can suppress visual information when both inputs keep temporal and spatial consistency [31]. As we have already introduced previously, this audiovisual suppression effect is on what this project is mainly focused. It was recently reported in a research group article, more concretely by *Sandra Malpica et al.* [1]. In this case, authors suggest that temporally congruent but spatially incongruent sound degrades visual performance considerably. Besides, they record users gaze behaviour to conclude that this suppression effect is due to close neural interactions or attention redirecting, rather than oculomotor phenomena such as saccades or blinking [1].

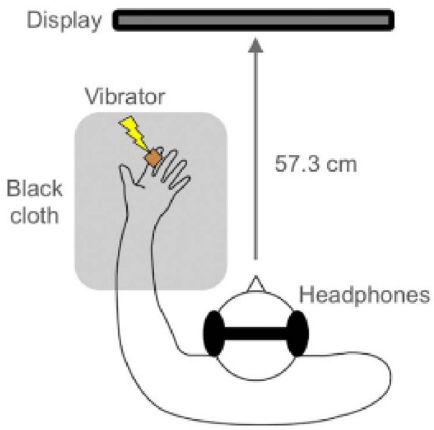


Figure 2.5: Visual-Tactile Suppression Effect: Tactile stimulation is caused by a hidden vibrator in user's hand while visual stimuli is displayed on screen. Headphones are used to conceal the sound of vibrations [11].

This report is the starting point of our project, since we are going to analyze and test how sound can affect our visual perception, using their configuration as baseline (Figure 2.6) and then varying several stimuli factors, aiming to find out its behaviour and scope.

In Figure 2.6, we can find the incongruent spatial configuration *(Image A)* chosen to distribute the audiovisual stimuli as well as the consistently-temporal profile *(Image B)* followed to generate them. According to both spatial and temporal layout, auditory and visual stimuli are presented to the user inducing the suppression effect. Performing several experiments, they infer that this suppression effect is still present in high cognitive load tasks while visual recognition and detection performance decreases no matter where sound and target are located at (spatial incongruent stimuli) [1].

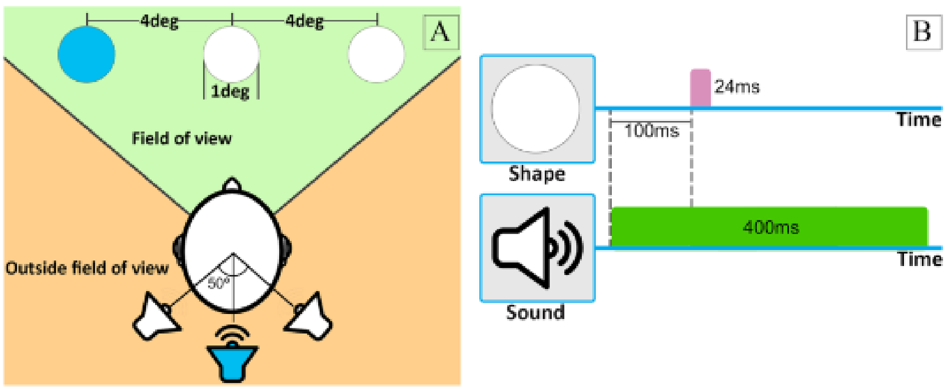


Figure 2.6: Audiovisual Suppression Effect. (A) Spatial setup: Three possible locations for visual targets inside the user's Field of View (FOV) and fixed locations for audio sources outside the user's FOV. A feasible case is highlighted in blue. (B) Temporal profile: Bimodal stimuli are made up of a sound of 400 ms and a target that spawns 100 ms after audio, disappearing when 24 ms elapse. [1]

Our contribution to this audiovisual suppression effect is related to human hearing range and will be addressed throughout this report. As we all know, humans can listen sounds only if they are within the audible range of frequencies. This range goes from 20 hertz (Hz) to 20.000 Hz, which are the physiological absolute limits. However, both limits varies for each person, setting real upper and lower hearing limits among which our auditory sensitivity increases. Besides, apart from sound tone, audible range is also set by sound intensity. Humans can usually hear from 0 decibel (dB) to 85 dB, from where a long exposure could be dangerous [32]. Therefore, both audio sources tone (frequency) and intensity (volume) are qualities we can modify and study its impact on the audiovisual suppression effect. It could be interesting to find out what happen if a sound located at the limits of the hearing range is able to trigger this suppression effect or not, as well as observing users behaviour. In essence, this idea is explored during our work so that our findings and the results achieved will be discussed in the last sections of this document.

3. Virtual Reality Equipment

Previously to the section where the experiment will be widely explained, we first approach both software and hardware tools used in this project. Regarding software, we have developed our experiments using Unity, which is a game engine with which we build our virtual scene and application. Besides, in order to introduce our setup, we address the essential hardware equipment used in the final steps of this project, more specifically, when testing the experiments that the participants carried out.

3.1 Unity

Unity is a real-time content creation platform known worldwide, which is used in very different fields and industries. Regarding videogames market, Unity arises as the world's leading platform, with over 50% multi-platform games being developed with Unity [33]. Unity offers an unmatched versatility to game developers: from indie games and small studios to enterprise studios and huge teams, while freely choosing among multiplayer and single-player, or 2D and VR to create a new game. Besides, Unity offers powerful tools, with which designers and developers optimize workflows overall, and gaming services like cloud data storage, security systems and profiling tools [33].

On the other hand, Unity also provides solutions to optimize and visualize products on automotive and manufacturing markets, speed up rendering pipelines used in animation and media producers, and create real-time 3D experiences in engineering or government industries. This coupled with the fact that Unity supports all major platforms like consoles, desktop and VR/AR (augmented relality), allows to reach even more customers and users [33].

In our case, what mostly interests us is the possibilities it offers to create our interactive 3D experience to test suppression effects. There are several products available according to user's needs, which include different tools and services. In our case, we just need to install Unity Hub and download the newest Unity Version [34]. In order to create a new Unity Project, we just need to open Unity Hub, select the latest Unity version available and the folder where it will be stored. Finally, it is worth to comment on the existence of the Asset Store. The Asset Store (UAS) offers a great variety of animations, environments, textures, materials, graphical user interfaces (GUIs), add-ons and tools, among others. The community itself is responsible for creating and sharing all assets while compatibility issues, updates and announcements can be discussed in massive public forums [35].

3.2 HTC VIVE Pro Eye 2

Regarding hardware, we need a virtual reality setup (HMD, controllers, sensors...) to perform our experiments in the virtual environments as well as interact with them. Counting on the devices available in the research group (Graphics and Imaging Lab), we are provided with a HTC VIVE Pro Eye 2 developed by High Tech Computer (HTC) Corporation. This headset is at the forefront regarding capabilities. It includes precision eye tracking with professional-level sound and graphics, offering a high quality immersive experience [12].

Among all his features, we can highlight that the eye tracking system achieves lifelike face movements and expressions while optimizes GPU workload by using foveated rendering [36]. Since it relies on *SteamVR*, an application developed by VALVE [37], wide-area experiences can be accomplished with deep immersion and comfort. The headset itself includes two dual-OLED displays with a combined resolution of 2880 x 1600 pixels and 615 pixels per inch (PPI) [36]. Additionally, it is designed to fit everyone's comfort preferences, being easy to adjust and comfortable to wear for long periods. Due to its design, adverse side effects are minimised and can be worn even with glasses [36].

In Figure 3.1, we can see the appearance of the HTC VIVE Pro Eye 2 headset as well as the two controllers and base stations that are included.

Figure 3.1: HTC VIVE Eye Pro 2. Headset, controllers and base stations. [12]

In order to use it to properly play our virtual experiences, we first need to complete the setup process. We are required to install VIVE and SteamVR software, link and configure VIVE hardware, and last define our play area. Once this process is accomplished correctly, we already have the HMD, the base stations and the controllers all connected and working.

3.3 Eye-Tracking: Tobii

As an improvement compared to other HMD, the HTC VIVE Eye Pro 2 headset includes high-precision real-time eye-tracking via software powered by Tobii.

Tobii is a global leader company in eye-tracking solutions whose methodology is based on capturing head motion and eye movements, transforming gestures into gaze data and generating insights on people's behaviour [38].

Eye-tracking is a cutting edge technology which consists of following in real time where the detected user is looking at. User's pupils motion and attention shifts are recorded by cameras and those data stream are computed using machine learning and image processing algorithms. Hence, visual data insights like pupil's location, gaze vector, convergence distance and blinking are obtained. Figure 3.2 shows a sketch from which we can extract the essence of how eye-tracking setup works.

Figure 3.2: Eye-Tracking system sketch. Cameras record pupil's motion and eye movements to get insights of user's behaviour [13]

Similarly to what we mentioned about VR, eye-tracking technology has many application areas. It can be used in surgical robots to enhance care and optimize performance, gaming by increasing audience engagement through visual attention and in-game awareness, clinical research by supporting early diagnosis of medical conditions, and marketing by noting consumers behaviour from their perspective [38]. Specially, its usage in virtual reality applications gives them a larger dimension by enabling new gaze-based interactions and hand-eye coordination tasks, improving performance assessment, easing VE exploration and giving virtual avatars a more human look through expressive gestures [39] .

In our case, eye-tracking cameras are incorporated in our VR headset so its integration with our virtual experience is straightforward. Some relevant parameters of our eye-tracking hardware are a refresh rate of 90 Hz and a FOV of 110º. The necessary software to install eye-tracking support is called *VIVE Pro Eye* [40]. Following the *VIVE Pro Eye* installation process, eye-tracking will be available and working. First of all, it is mandatory to perform eye-tracking calibration with each participant. This process must be started from *SteamVR* application, by selecting *VIVE Pro Eye* option. After placing the headset properly in front of the eyes, it adapts to the user by modifying inter pupillary distance (IPD). The process ends with the user performing the required gaze movements and checking that eye calibration is correct [41]. It is worth commenting that gaze data is recorder to further study user's visual behaviour but it is not necessary at all to trigger the audiovisual suppression effect.

4. Pilot Experiment

In this chapter, we explain everything related to the experiment performed, from the valued proposals and design, to its implementation and validation. We focus on how the experiment is composed and developed while justifying all decisions made. We also approach the surveys presented to the participants, the eye-tracking data collected and the output logs among other topics.

4.1 Study Proposals

As we mentioned previously at the end of Chapter 2, our starting idea is to research how audio sources can affect the audiovisual suppression effect, how sound qualities impact on the visual information loss. Hence, we first collect different aspects present in Malpica et al. work [1] and value how they could be modified to influence the audiovisual suppression effect. So, initially, we classify the existing variables into: audio, changing audio sources (auditory stimuli); targets, varying visual targets (visual stimuli); scene, where the experiment is carried out; functionality, including different conditions and cognition levels; and metrics.

In the first case, we valued to act on auditory information in order to find out what would happen when the user encounters specific sounds and how it would affect the suppression effect. For example, it was proposed to include dynamic (moves along with the user) and static (fixed location) sounds, modify sound volume and frequency, and incorporate lateral silences, color noise and ambient sounds. These options arise from the most distinctive parameters of the auditory modality. Something similar happened with visual information by proposing to include new simple shapes, scene objects or flashes, and perform subtle changes on scene. Additionally, it was suggested to use different types of virtual scenes with enough information to affect the suppression effect such as a building roof and a museum room. Different metrics were also valued to get a deeper insight into the effect as well as other type of features present on the scene like randomness, cognitive tasks, exploration and emotions.

All these ideas are summarized in the table displayed in Figure 4.1. The shaded boxes are those variables that are included in our work and will be taking into account when performing the experiments. In addition, those highlighted boxes are the chosen sound qualities to study in our work. Since including all variables remains unfeasible, we need to select those parameters we are most interested in. The goal is to observe how our auditory range influences the suppression effects while varying audio intensity (volume) and tone (frequency).

We will infer if those sounds near to upper and lower hearing limits have a significant impact on the perceived visual information, considering some questions to solve like: *Do we lose more or less visual data with those near-limits sounds?*, *How these type of sounds affect the sense of immersion? Do they reduce or increase it?...*

Audio	Targets	Scenes	Metrics	Functionality
Dynamic-Static	Dynamic-Static	Simple Room	Eye Tracking	Variability
Volume	Simple Shapes	House Room	Saliency Maps	Simultaneity
Frequency	Scene Objects	Museum Room	User Path	Randomness
Color Noise	Subtle Changes	Outdoor Space	Reaction Time	Exploration
Environmental	Flashes	Rooftop	Saccades	Cognitive Task
Fetching	Videos	Stormy Forrest	Blinking	Redirected Walking
Semantic	Blind Spots	Leisure Activity	Task Performance	Emotions
ASMR			Surveys	Personalization
Lateral Silence				

Figure 4.1: Summary Table of the candidate variables to be included in our experiments. For each component present in our VE, different case studies are proposed in order to define and choose which ones are included according to our goal (those highlighted in blue).

4.2 Design

As we have already introduced in previous sections, our project takes as baseline the setup used in Malpica et al. [1]. So, in order to design our experiments, we will initially try to replicate their configuration as much as possible and then include our novelties. We consider how to generate the audiovisual stimuli necessary for triggering the suppression effect as well as what data we want to record.

4.2.1 Visual Targets

In the first instance, the visual stimuli that will be presented to the user are similar from the ones used in our base paper [1]. The idea of these visual targets is to spawn in front of the user and disappear shortly after. As we saw in Figure 2.6, visual targets are presented in three fixed locations inside user's field of view (FOV). The first location, which is the central one, must take only one degree of the user's FOV, specifying a very small size for this visual targets consequently. The other two locations, right and left ones, are placed at a distance of four degrees on each side of the central position respectively. All these three locations must be fixed, which makes the dynamic targets move according to the user's position and orientation. At this point, we already know the specifications regarding the spawning positions.

Besides, visual targets shapes are restricted to three simple 2D shapes only which will be circle, square and rhombus. No more shapes are included since Malpica et al. [1] demonstrated that there is not significant difference among them and they are enough for each sound. The idea is that these visual targets will spawn in a random time within a range so they cannot be predicted by the user, and its lifetime is only 24 ms after which they will disappear. Ultimately, target shape and location must be random among the available possibilities, so each possible target can appear at any of the three possible locations arbitrarily.

4.2.2 Audio Sources

Audio stimuli follows similar conditions to meet, which can also be seen in Figure 2.6. Audio stimuli behaves similarly to visual targets, but in this case they spawn behind the user and disappear after a while as well. Audio spawning locations are outside the user's FOV and there are also three fixed locations, situating side positions with an angle of fifty degrees with respect to the central location. Other than that, there are no more requirements related to audio spawning locations.

In this case, the audio sources that will be presented to the user adjust to the purpose of this work. Since our idea is focused on researching on the effective human hearing range, the generated audio sources will be located at the limits of this range. Hence, auditory stimuli are designed selecting high and low values for the sound intensity (volume) and tone (frequency). So, there will be four audio sources combining those two sound qualities: low volume and low frequency *(vol1_freq1)*; high volume and low frequency *(vol2_freq1)*; low volume and high frequency *(vol1_freq2)*; and high volume and high frequency *(vol2_freq2)*. These frequencies are set for each user to bring the sounds closer to the limits of the audible threshold, as explained in Chapter 5. Besides, pink noise is also included making a total of five audio sources.

As it happens with visual targets, audio sources spawning time is also randomized within a range and they will sound during 400 ms before fading. It is important to remember that both auditory and visual stimuli have to be temporally consistent, this requirement is key to trigger the audiovisual suppression effect so the temporal profile displayed in Figure 2.6 (B image) has to be fulfilled.

4.2.3 Data Logs

Once we have commented the input stimuli that the user will encounter in the experiment, we can address what output data we are interested in recording. These output logs are classified in three different categories: stimuli logs, which are the records of which stimuli was present at each moment, as a kind of ground truth; performance logs, where it is shown each time the user has seen a visual target or listened a audio source (or both); and eye tracking logs, which exhibit user's gaze behaviour throughout the experiment. Therefore, analyzing these output logs we may be able to find out how many stimuli the user has detected of those that have been presented.

Thereby, user's performance logs in contrast with stimuli logs will evidence the appearance of the audiovisual suppression effect, in those cases where the visual data has been lost when auditory stimuli were present. In addition, eye tracking data will not only serve to corroborate that this visual information loss is not due to eye movements like saccades or blinking, as reported by Malpica et al. [1], but also to study user's response to spatially incongruent audiovisual stimuli, among others.

4.3 Implementation and Validation

Once we have reported the features we want our experiment to have, we can introduce its operativity while explaining the performed implementation process.

The main goal of our experiment is that the user has to explore the virtual scene where he is located at, while looking around and moving within a limited area. Meanwhile, audiovisual stimuli will show up, so that the user will notify when listening audio sources or seeing visual targets, or both at once. The total sum of the presented cases will be 45: 15 visual stimuli, 15 auditory stimuli and 15 bimodal cases (audiovisual stimuli), with which we want to trigger the audiovisual suppression effect. Next, we explain in depth all the experiment components.

In this point, we can see the UP Hierarchy in Figure 4.2. As we can see, our VE is composed by four components (blue boxes): TobiiXR Initializer, which is in charge of starting eye-tracking process; Scene, formed by the living room and environmental sounds; Pilot Experiment, which is in charge of managing the experiment itself; and Player, which includes everything related to the user embodiment. If not stated otherwise, each box represents a Unity Gameobject while the scripts associated (if any) appear inside the box. Unity Gameobject is the base class used for all entities in Unity Scenes, like a 3D item (a chair, a speaker or a square) or the camera, and scripts define their behaviour changing items appearance and location for example. As we can see, Pilot Experiment box includes an script so it appears in black and will be addressed in section 4.3.6.

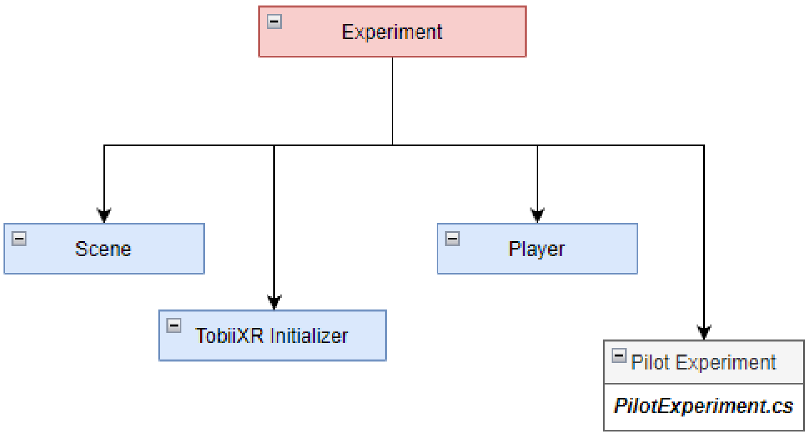


Figure 4.2: Experiment's Virtual Environment Hierarchy.

Next, we explain more in detail all these experiment components.

4.3.1 Scene

The used VE is the same as the one used in our base paper [1]. The scene consists of a living room where the user can explore its surroundings and move within a limited area. The living room is made up of paintings, shelves, a couch, a stereo, a TV, lamps, tables and so on. The stimuli that the user has to perceive will spawn surrounded by this place and these objects. This virtual scene is available in Unity Asset Store [42] so it is only needed to download and import it into our Unity project. Once this is done, some scene scaling changes have to be done to adapt it to our VR setup. Hence, what we need to do is to open Unity Asset Store, online or from our Unity project (UP), and search for the 3D Living Room [42]. As soon as we download and import it into our project, the virtual living room components and the scene will appear in it.

Besides, apart from auditory stimuli, there are a couple of environmental sounds. These two background sounds are a podcast that is being played in the stereo and bird sounds coming from the windows.

Next, we can see a couple of images of the virtual scene in Figure 4.3. In those images we can observe the very real appearance of the items that compose the virtual living room. We can also comment that in the 360° image (left), the couch has been moved to get a better visualization. The real location of that couch used in the experiments is the one shown in the 3D image (right).

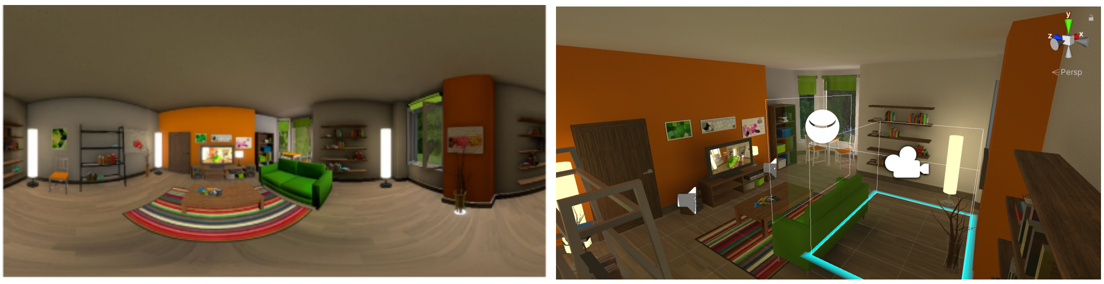


Figure 4.3: Living room virtual scene used in the experiments. Left: 360° panorama rendered from the central point of view [1]. Right: 3D Virtual scene seen from a top corner. It can be appreciated the podcast speakers in the stereo, the camera symbol associated to the player and the player's motion zone highlighted in blue. This players zone corresponds to a physical space of 3.5 x 2 meters where the user can move freely.

4.3.2 Eye-Tracking

As we have already commented previously in this report, integrating the eye-tracking software powered by Tobii with our VIVE HMD is very straightforward, since all necessary hardware is already included in the headset and we have already installed VIVE Pro Eye software to enable eye-tracking [40]. However, we need to do some previous stuff to use headset eye-tracking in our UP. We first need to download VIVE SRainpal SDK (software development kit) and import the associated Unity package into our UP, where we already have the living room scene. Next, we need to enable and set up VR support which is achieved installing the XR (extended reality) Plugin Management. Last, it is necessary to download and import the Tobii XR SDK package into our UP [43].

At the end of this process, Tobii XR SDK package includes sample scenes to check that eye-tracking is indeed enabled and works correctly. One of these sample scenes consists of a empty virtual room with three white cubes in front of the player, which light up red when the player looks at them. In order to see better how it works and where the user is looking at, we have drawn the gaze ray using the eye-tracking data (gaze ray origin and direction) and a line renderer. In order to plot this line renderer dynamically, we need to specify start and end 3D points. The start point is straightforward since it is directly the gaze ray origin, while the end point will be the gaze ray direction multiplied by some enough distance value and added to the gaze ray origin. So, the points that define line renderer are computed as follows:

$$\begin{aligned} P1 &= GazeRayOrigin; \\ P2 &= P1 + GazeRayDirection \cdot distance; \end{aligned} \tag{4.1}$$

We can see in Figure 4.4 (left) how the gaze ray comes out of the user's eyes and reaches the cube, which changes to red consequently when the user looks at it. Thereby, it is verified that the eye-tracking system runs smoothly and the gaze data recorded is correct as well. So, after installing these plugins and software, Tobii eye-tracking is working and linked to our UP, so we are already able to develop using gaze data.

In our project hierarchy, the component that enables and starts eye-tracking is the TobiiXR Initializer gameobject, which appears in Figure 4.2. At the same time, its own structure can be visualized in Figure 4.4 (right). As we can notice, it consists of two scripts to record Eye-Tracking Logs and Stimuli Logs, which appear in black and will be addressed in section 4.3.6.

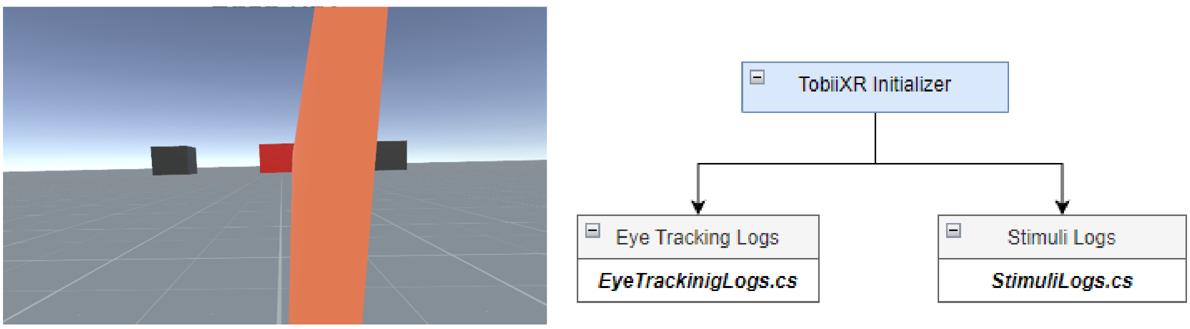


Figure 4.4: Eye Tracking validation using a sample scene (left) and hierarchy (right)

4.3.3 Player

In order to include the player in our UP, we need to combine the eye-tracking components and plugins included in the previous section with SteamVR plugin. This new SteamVR plugin can be downloaded from UAS and offers a smoothly interface between Unity and SteamVR. It also manages 3D models for VR controllers, handles input from those controllers and estimates what user's hand looks like while using those controllers [44]. So, we just need to install this SteamVR plugin in our UP in order to include the mentioned features. The result of adding this new capabilities to our VE can be seen in Figure 4.5 (left). Including the controller models in our previous sample scene, we can check how the controller also appears and follows user's hand while eye-tracking still works properly.

If we include these controller management into our living-room scene, it looks like Figure 4.5 (right) shows, with the controller working as the virtual user's hand.

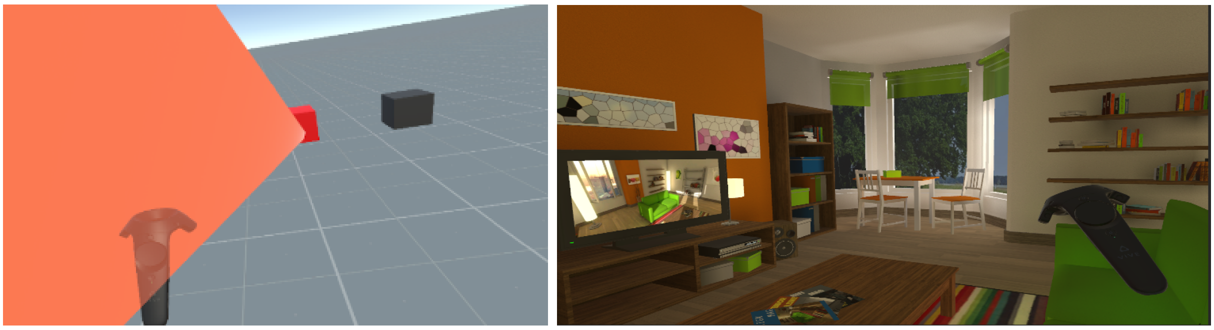


Figure 4.5: Player's controller features included in the sample scene (left) and our virtual living room (right).

Additionally, a player zone has to be delimited from SteamVR application when configuring the VR setup. This setup consists first of connecting the HMD and controllers as well as placing base stations properly so they see the headset and the controllers. Then, the play zone has to be delimited in order to ensure player safety relative to real world. Using the controllers, the user limits the play zone to the real area where he can move freely without colliding with anything. This play zone also appears in the VE so the user is aware of its limits even wearing the headset. In Figure 4.6, this play zone can be found highlighted in blue.

Regarding its hierarchy, the player's structure is shown in Figure 4.7. It includes the camera rig which represents the play zone, limiting the space where the camera can move. We can notice how it then consists of the main camera corresponding to the headset and the left and right controllers, although we will use only one of them. Last, the main camera also includes the canvas where the pause messages and visual targets will appear, and the auditory stimuli to spawn. This stimuli generation will be explained in section 5.1.1.

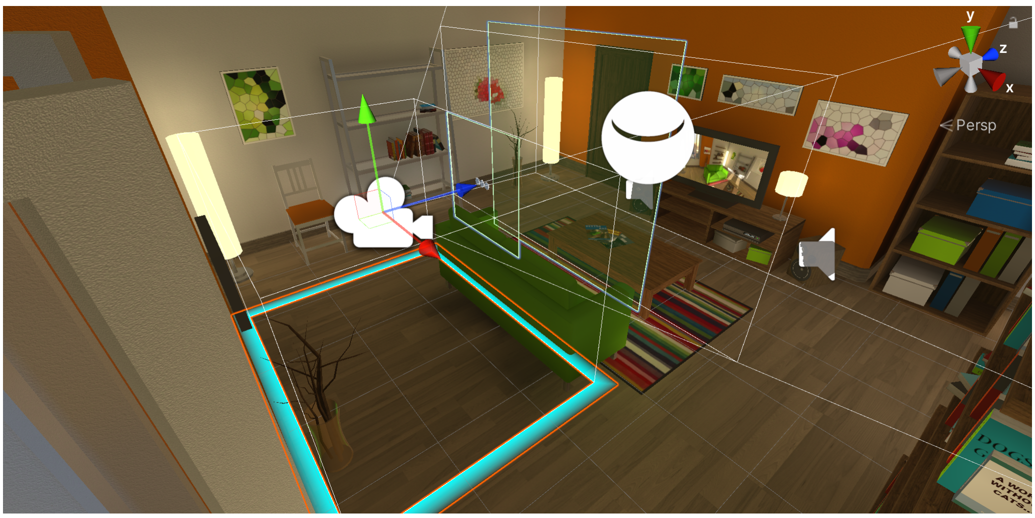


Figure 4.6: Player playing zone highlighted in blue. Player's camera and stereo's audio sources that play the podcast are also visible.

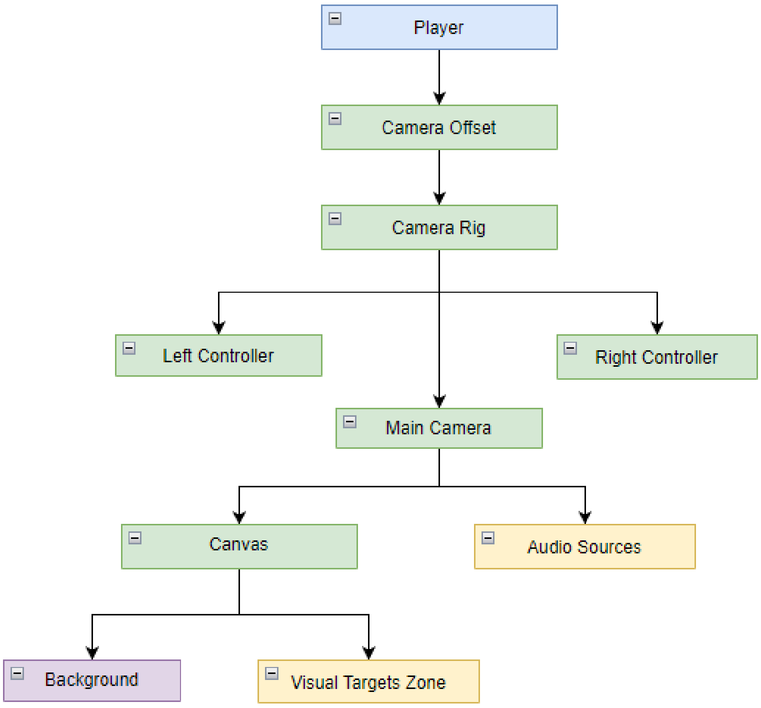


Figure 4.7: Player hierarchy

As we have already mentioned, the user's (player) work throughout the experiment is to explore the virtual scene, looking around and taking a few steps. When the user perceives some audiovisual stimuli, he only has to press the controller's trigger to stop the experiment and start the pause routine, which uses the canvas mentioned before and will be explained in section 4.3.5.

4.3.4 Stimuli Generation

In this section, we are going to explain how audiovisual stimuli are generated and how the experiment works. As we can see in Figure 4.2, the Pilot Experiment gameobject has a script associated. This script is the one who manages the experiment and stimuli generation. All audiovisual stimuli presented to the user are in fixed positions relative to the player, so they are generated taking user's position and orientation into account. In this way, the spatial setup showed in Figure 2.6 is completely maintained in our case. Therefore, both visual and auditory stimuli are generated from the player's location as Figure 4.8 shows. Visual stimuli appear in the targets zone associated to the canvas, whose location is fixed in front of the camera location. Similarly, audio stimuli are always generated behind the user's location so they are associated to the camera location as well. All stimuli-related boxes appear in yellow in all hierarchy drawings.

Firstly, we are going to explain what kind of visual stimuli do we generate. As we have already mentioned, the visual targets the user is going to face are a square, a circle or a rhombus, three simple shapes. Besides, according to spatial setup represented in Figure 2.6 (A), there are three possible locations for this visual targets to spawn.

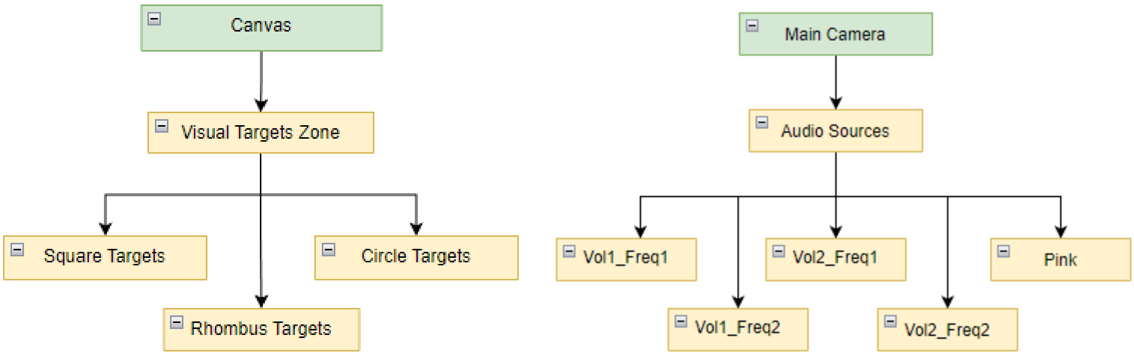


Figure 4.8: Visual (left) and Auditory (right) stimuli hierarchy

Therefore, combining the three shapes (square, circle, rhombus) and the three locations (left, centre, right), there are a total of 9 cases for the visual targets. All these cases are present in front of the user all along but the targets are initially transparent so they are not seen. When a visual target has to spawn, we just make that target opaque so it seems to have been created. When it has to disappear, we set its appearance transparent so it cannot be seen again. So, we manage to appear and disappear visual targets just by modifying its material transparency. We can find in Figure 4.9 how the cases of circle on the left, square on the center and rhombus on the right are seen from user's POV (point of view).

Figure 4.9: Three possible visual stimuli cases: circle on the left pose, square on the center pose and rhombus on the right pose. Its scale has been increased for visualization reasons.

On the other hand, we have also commented that auditory stimuli were classified according to sound intensity (volume) and tone (frequency). Setting high and low values for both qualities makes it 4 possible audio sources. Adding the pink noise for comparison reasons, makes a total of 5 audio sources available: Low volume and frequency ($Vol1 - Freq1$), high volume and low frequency ($Vol2 - Freq1$), low volume and high frequency ($Vol1 - Freq2$), high volume and frequency ($Vol2 - Freq2$) and pink noise. For each user, its lower *(Freq1)* and upper *(Freq2)* frequency values are obtained from the Frequency test, which we will explain in Chapter 5. Being these the two frequency values, the volume values are 0.2 *(Vol1)* and 0.8 *(Vol2)*. These values are associated to the Unity speakers volume, which ranges from 0 (minimum, no sound) to 1 (maximum).

Similarly to visual stimuli, the spatial setup represented in Figure 2.6 (A) is maintained for auditory stimuli as well. Hence, combining in this case the five audio sources ($Vol1 - Freq1$, $Vol1 - Freq2$, $Vol2 - Freq1$, $Vol2 - Freq2$, Pink) and the three locations behind the player (left, centre, right), there are a total of 15 cases for the audio stimuli. In this case, when audio stimuli has to be generated, an audio speaker playing the correspondent sound is created in the specified location at the right time and is destroyed a moment later. We can see in Figure 4.10 two examples of how visual (left) and auditory (right) stimuli are generated from user's POV.

Figure 4.10: Visual (left) and Auditory (right) stimuli examples appearance

At this point, we already know how individual visual or auditory stimuli are generated as appropriated. In order to generate bimodal stimuli which includes visual and auditory information, we have to follow the temporal profile shown in Figure 2.6 (B). As it looks, audio source starts playing and after 100 ms the visual target appears. 24 ms after the visual target spawn, it disappears and when the audio source reaches 400 ms it stops playing. Thus, we keep temporal congruence among stimuli although spatially they may be inconsistent, since visual target appears on front and audio source plays behind.

Knowing how unimodal or bimodal stimuli are generated, we can comment that the experiment consist of a total of 45 stimuli cases. These cases are divided in 15 visual stimuli, 15 audio stimuli and 15 bimodal stimuli. The interval between stimuli cases is randomly chosen between 5 and 10 seconds each time a stimuli has to be generated. Also, all 45 cases are shuffled at the beginning. In this way, we try to avoid the user to be able to predict when a audio source will be played or a visual target will spawn (or both). The visual-only stimuli are compounded by 15 random selection of the target shape and location, which define one of the 9 possible existing cases. The auditory-only stimuli are dense sampling of the 15 odds (5 audio sources x 3 locations). Finally, bimodal stimuli are a mixture of both unimodal cases: visual information is randomly selected while auditory information is densely sampled.

4.3.5 Pause Subroutine

Once the audiovisual stimuli (unimodal or bimodal) are being generated throughout the experiment, the user has to pause the experiment whenever any stimuli is detected. We need to remember that both the podcast being played on the stereo and the birds sounds does not count as stimuli but as environmental sounds belonging to the virtual scene.

In order to pause the experiment, the user only needs to press the trigger button on the controller. Once the trigger is pressed, the experiment stops and the background canvas appears. If we remember Figure 4.6, this canvas depends on the camera location so it can be properly seen by the user always. We can see in Figure 4.11 how the canvas is composed by a background where the different texts (pause, options and end) are displayed. As we may observe, canvas-related boxes appear in purple in all hierarchy drawings.

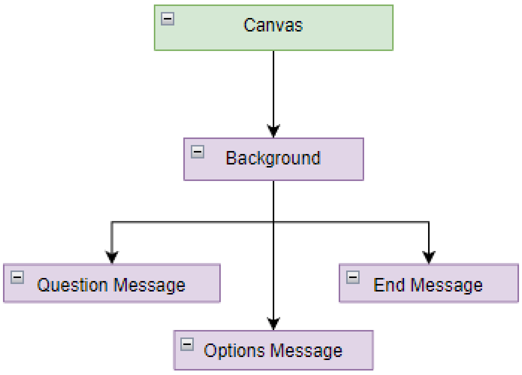


Figure 4.11: Canvas hierarchy: Background and messages.

Similarly to how visual targets spawn works, the canvas is always placed in front of the user but it is transparent so it cannot be seen. Once the user presses the trigger button, its background becomes translucent by modifying its transparency. Then, it displays a question where the user is asked for what stimuli has been detected. The answer options are audio, target, bimodal or error, if the user has pressed the trigger unintentionally. This first message in the middle of the translucent green background can be seen in Figure 4.12 (left). If the user answers with the options of target or bimodal, a second message appears which asks for the shape of the visual stimuli. The available options of this case are circle, square, rhombus and do not know, if the user has not been able to distinguish it. This second message can also be seen in Figure 4.12 (right). The user's answers are introduced via keyboard and they are recorded all times the user stops the application. All these data are also recorded and it will be also addressed in section 4.3.6.

Figure 4.12: Pause Messages. When the user presses the trigger button, the first message and its options appear (left). If a visual stimuli has been perceived, the second message with its respective options appear (right).

Finally, once the 45 stimuli cases are presented, the experiment ends displaying the end message that can be found in Figure 4.13.

The background appears again displaying a message which informs the user that the experiment has already finished. Then, the user has to press the trigger button one last time to terminate the experiment and stop the application. After that, the user can remove the HMD and the controller, ending the virtual experience.

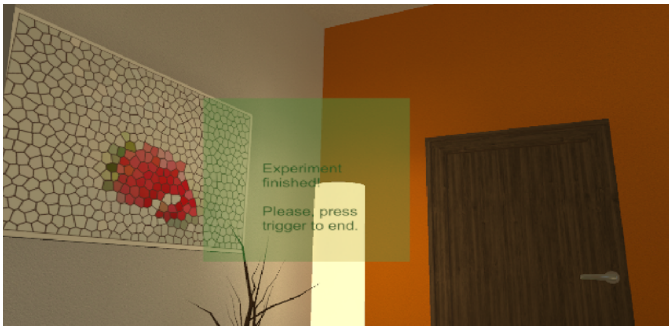


Figure 4.13: End message displayed once the experiment has ended.

4.3.6 Output Logs

Through the experiment, we record data of the stimuli presented to the user, user performance and eye-tracking data. The Unity gameobjects that have the scripts which control this data register appear in grey in Figures 4.2 and 4.4 (right). All data are recorded in three comma-separated values (CSV) files, one file per data type: Eye-tracking, Stimuli and Performance.

The first one stores user's gaze information while the second one keeps the moments when any stimuli was present. Last, performance data shows user's performance throughout the experiment. The user's provided answers are gathered and stored in the correspondent file after restarting the experiment each time the user pauses it. Eye-Tracking data is stored as the experiment progresses and this data is also dumped into a file only when user detects any stimuli and pauses the experiment. So, time is optimized since the output file is opened, written and closed from time to time. The same process is followed when recording stimuli data, using user's pauses to dump data into another output file as well.

In the case of eye-tracking data, the gaze information stored are the timestamp, camera location and rotation, convergence distance, gaze ray origin and direction, and eye blinking. An example of eye-tracking data can be found in Figure 4.14. For the stimuli data, the collected data are the timestamp, if there is a visual target and which one is, and if there is an audio source and which one is. An example of stimuli data can be found in Figure 4.15 (left). Lastly, the recorded performance logs are made up of the timestamp, the stimuli detected and its shape if any. An example of stimuli data can also be found in Figure 4.15 (right).

	Timestamp	CameraLocation	CameraRotation	ConvergenceDistanceIsValid	ConvergenceDistance	GazeRayIsValid	GazeRayOrigin	GazeRayDirection	IsLeftEyeBlinking	IsRightEyeBlinking
0	12,7777	(-10.0, 21.5, -21.2)	(0.0, 0.0, 0.0, 1.0)	True	0,4017847	True	(0.0, 0.0, 0.0)	(0.1, 0.0, 1.0)	True	True
1	12,82159	(-10.0, 21.5, -21.2)	(0.0, 0.0, 0.0, 1.0)	True	0,4039817	True	(0.0, 0.0, 0.0)	(0.0, 0.0, 1.0)	True	True
2	13,49077	(-10.3, 21.3, -21.8)	(-0.1, -0.1, 0.0, -1.0)	True	0,359728	True	(0.0, 0.0, 0.0)	(-0.1, 0.0, 1.0)	False	False
3	13,53835	(-10.2, 21.2, -21.8)	(-0.1, -0.1, 0.0, -1.0)	True	0,3493676	True	(0.0, 0.0, 0.0)	(-0.1, 0.1, 1.0)	False	False
4	13,54925	(-10.2, 21.2, -21.8)	(-0.1, -0.1, 0.0, -1.0)	True	0,3292957	True	(0.0, 0.0, 0.0)	(-0.1, 0.1, 1.0)	False	False

Figure 4.14: Examples of Eye-Tracking output logs

	Timestamp	ExistsVisualTarget	VisualTargetShape	ExistsAudioSource	AudioSource
1225	27.52877	True	Circle_Center	False	None
1226	27.53996	True	Circle_Center	False	None
1227	27.55124	True	Circle_Center	False	None
5008	70.85227	True	Square_Center	False	None
5009	70.8636	True	Square_Center	False	None

	Timestamp	Stimuli	Shape
0	21.58115	Audio	None
1	30.01569	Target	Square
2	38.62561	Error	None
3	40.58607	Audio	None
4	53.59577	Bimodal	DontKnow
5	63.77962	Bimodal	Square

Figure 4.15: Examples of Stimuli (left) and Performance (right) output logs

4.4 User Surveys

Additionally, users need to fulfill some surveys in order to collect more data. The previous survey has to be done before performing the experiment while the post survey has to be completed once the user has finished the virtual experience. Also, a discomfort survey needs to be carried out before and after the session. All surveys have been developed using Google Forms [45].

4.4.1 Previous Survey

The first survey to be completed by the user is the Pre-Experiment Survey. The first thing that appears in this survey is the data protection laws of University of Zaragoza and Google, which the user has to read and agree. The user has also to agree that the session will be recorded, among other things, and understand that the participation on the experiment is voluntary, being able to stop whenever the user desires. In order to keep the recorded data and the surveys results anonymous, each user is granted with a personal ID which does not change throughout the process.

After this, a series of demographic questions are presented as well as questions related to visual and auditory conditions, since this information is relevant in our case and it could affect the results. Finally, the user is asked regarding his experience in VR and video games.

The survey itself can be found in the Annex B.

4.4.2 Post Survey

Once the user has ended the experiment and has the HMD and controller removed, he has to fulfill the Post-Experiment Survey. The first thing that this survey requires is the user's ID, the same that was introduced in the Pre-Experiment Survey.

The user is asked questions regarding the experiment such as if he had any problem wearing the HMD or performing the experiment, if he was able to distinguish visual targets easily or predict when a stimuli was going to be generated, and if he has any comment regarding the experiment. Lastly, some questions related to immersion in the VE are added with the aim of measuring the user's sense of presence and realism in the virtual scene.

The survey itself can be found in the Annex C.

4.4.3 Discomfort Survey

In order to consider and measure possible side effects, the user has to complete the Discomfort Survey.

The idea is that the user completes this survey twice, first before performing the experiment and again after finishing it. So, the user's condition before and after the experiment can be compared, inferring what effects can have the experiment on users' state. The survey consists of some symptoms that the user has to rank among 'None' if the symptom does not exist, and 'Severe' if the symptom is significant.

The survey itself can be found in the Annex D.

5. Limit Frequency Test

In parallel with the main experiment broadly explained in the previous section, we also carry out a frequency test to all the participants. The goal of this experiment is to obtain the hearing frequency limits for each participant. Once this hearing interval is known, which is determined by the upper and lower boundaries, these frequency limits are used in the audio sources played in our experiment, as we have commented previously.

5.1 Design, Implementation and Validation

In order to perform this frequency test, we have to create a new UP. In this fresh virtual environment, the user will be in an empty room since no scene is needed in this case. Including eye-tracking is also unnecessary because our aim is to obtain user's frequency boundaries only. Therefore, this Frequency Test components hierarchy is much more simple than the experiment's structure. This project will be only composed, as shown in 4.7, by the user (player) as well as the canvas and audio sources attached to it. We do not need to include everything commented in 4.3.3, since the player's functionality in this new VE is simpler and only needs the camera linked to the HMD. We can see the structure of this Frequency Test project in Figure 5.1.

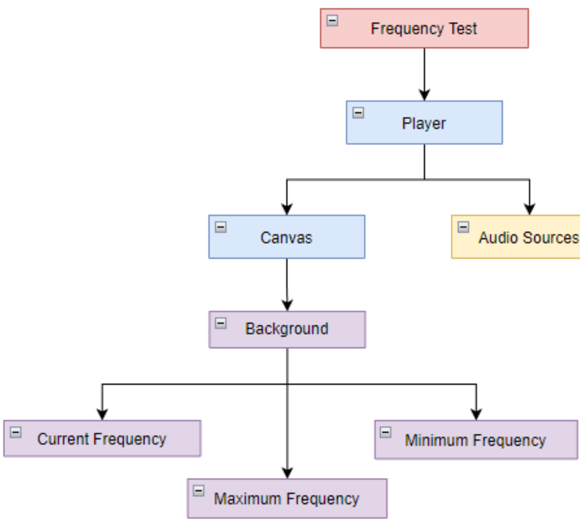


Figure 5.1: Frequency test hierarchy

Therefore, this VE has the canvas attached to the player so its location is fixed in front of the user, similarly to the canvas used in the experiment but now it is always visible. And the two audio sources, one on each side to get stereo sound and both also attached to the player. These two audio sources produce a sound of a certain frequency, which appears on the canvas. The idea is to modify this frequency until reaching the upper and lower limits of the hearing range, which will be stored and written on the canvas as well.

5.1.1 Sound Generation

Regarding the two audio sources present in our virtual scene, remember that each one is located at one side of the player, they are generated using speakers gameobjects available in Unity. Each audio source has the same script attached, which emits the generated wave sound through the output audio channel. This wave sound is basically a sinusoidal waveform whose temporal equation is shown below.

$$w(t) = K * sin(2\pi * f * t/sample\_rate) \tag{5.1}$$

The sinusoidal output data *w(t)* is calculated using the sine function over a period while taking into account the timestamp t and its frequency **f**. This last parameter is the one we have to modify in order to find out which frequencies delimit users hearing range. The product of these variables is divided by a *sample rate*, which is used to discretize the theoretical analog wavelength into digital samples values which can be indeed used in the output audio channel to generate sounds. This digital value can be finally multiplied by a gain factor K. In our case, this gain factor is not used since we set the amplitude of the wavelength (sound intensity) using the volume parameter of the speaker, which turns out to be the same actually.

Once the Frequency Test starts, both audio sources start playing the sound whose frequency is modified to find out user's hearing range limits. This way of generating sound is the same one used in the experiment to generate auditory stimuli, with the difference of those audio sources have a fixed frequency which will correspond to the upper or lower heard limit.

We can see the appearance of this VE in Figure 5.2. As we can observe, it consists on the player, audio sources which play the sound on each side and the canvas where all test information appears.

5.1.2 Interface

During this frequency test, the user has to wear the HMD and he will see, as we have mentioned, the canvas while listening the audio sources. This canvas displays the current frequency that is being played and the frequency values stored as the maximum and minimum values with which the user can still heard the sound.

Starting from a low frequency value (0 Hz) which the user cannot hear, it will be increased until the user reports that the sound becomes audible, setting the minimum frequency value (lower limit).

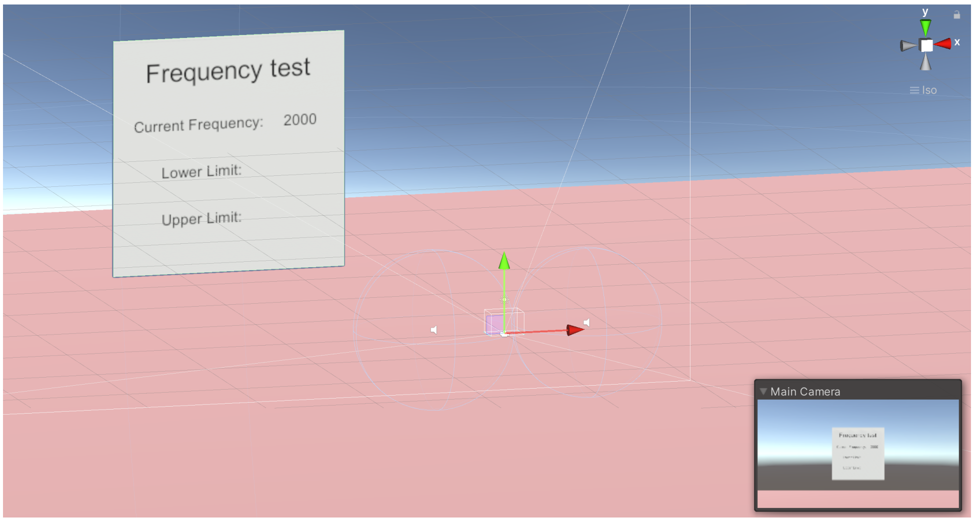


Figure 5.2: Frequency Test VE: Player's camera, audio sources and canvas.

Then, the frequency keeps increasing while the user realizes that sound becomes higher pitched until it stops being heard. At that moment, the frequency value is saved as the maximum value (upper limit). Both the variation (increase or decrease) of the frequency value and limit values saving are performed via keyboard.

The canvas shows the test information in real time. The current frequency is being updated at the moment it changes while the lower and upper limit values are displayed when they are saved. Once the upper value has been reached and stored, a message appears on the canvas reporting that the test has finished. After that, the test ends and the application closes.

In Figure 5.3, we can see what the user sees when performing the test. We can notice how the canvas information varies, from the early stages of the test (left) to when it is completed (right).

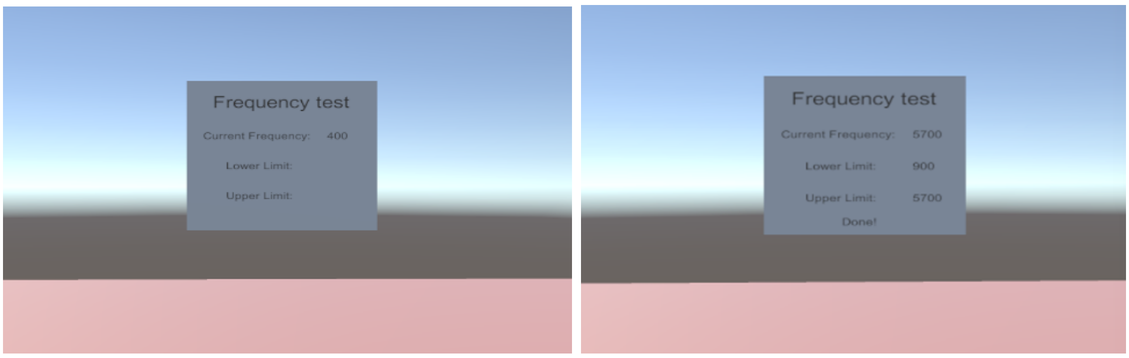


Figure 5.3: Frequency Test under user's POV. The test starts only showing the current frequency (left) while the upper and lower limit frequency values are also displayed when ending (right).

6. Results

Both the pilot experiment and the frequency test, as well as the pre-test, post-test and discomfort surveys, are done by all the voluntary participants. The participants group is composed by 20 people, of which 12 are men (60%) and 8 are women (40%). The mean age of the participants is 23.47 ± 1.43 years. More information regarding the participants will be provided later on. The experiments have been carried out in University of Zaragoza facilities and the achieved results are presented next.

6.1 Limit Frequencies

Firstly, we are going to comment the results obtained when performing the Frequency Test to all the 20 participants. We can see in Annex E, a table where the user ID, the experiment date, and both lower and upper frequency values, which delimit user hearing range, are all displayed.

As we can observe, the lower limit ranges between 300 and 150 Hz, being the mean value 191.00 ± 110.03 Hz. On the other hand, upper limit range is wider, between 17500 and 8300, being the mean value 13087.5 ± 1824.96 Hz. It could be mentioned that these values are within the psychological hearing range commented at the end of Chapter 2, which was from 20 to 20000 Hz. However, they differ significantly from each other since they depend exclusively on the hearing capabilities of each user. We can also notice that the values are generally rounded to the hundreds in order to get a valid approximation.

6.2 Surveys

We now move to report some data extracted from the results collected when performing the Pre-Experiment, Post-Experiment and both Discomfort surveys.

6.2.1 Previous Survey

Regarding the Pre-Experiment Survey, the participants mean age is 23.47 ± 1.43 years while the highest education level completed is tied between University Degree and Master Degree with a 38.17% each one. More than half (57.1%) reported to have any vision problem, being myopia and astigmatism the most common ones.

For those who reported having a vision problem, answers when they were asked if they had it fixed were diverse: 38.5% said to have it fixed by wearing contact lenses and 30.8% by wearing glasses instead, but 30.8% admitted to not have it corrected. On the other hand, no one reported to have any auditory problem and one participant informed to have dyslexia.

Changing topic, 61.9% of participants said to play video games, of which 58.3% clarified not to play more than an hour but 8.3% admitted to play more than three hours on a daily basis. Almost everyone (95.2%) said to have heard about VR while the 76.2% declared to have ever used a VR device. From those who have ever used any VR device, more than a half (56.3%) have used it occasionally (less than 5 times in total) and mostly report to have used some computed-type device like HTC VIVE or Oculus headsets (87.5%).

6.2.2 Post Survey

In the Post-Experiment Survey, 81.8% of the participants reported to not have any problem wearing the HMD but others commented that headset size was slightly small. Also, almost all participants (95.5%) found the experiment easy to perform.

Regarding stimuli presented, 81.8% admitted not to be able to distinguish the visual targets easily, reasoning that they disappeared very quickly and their size was very small. More than the half (63.6%) suggested that they would include scene objects as a new type of visual targets, also making suggestions such as changing target's color depending on the background color or including moving targets. On the other hand, only 22.7% said to not be able to listen audio sources easily, commenting that birds and podcast ambient sounds made it more difficult. The majority of participants (81.2%) said to not feel any kind of discomfort throughout the experiment while a 40.9% admitted to be able to predict when a stimuli would be generated.

With respect to the sense of immersion and presence, we use a scale from 1 to 5 where 5 represents your normal experience of being in a real place. Around 68.2% of the participants awarded a 4 or more to have the feeling of "being there" and 77.3% admitted to remember the virtual experience as a place they visited rather than images they saw. At last, users were asked to report any suggestion or comment to improve the experiment. Some answers were that there were some distracting flickers in lighting and metallic materials, and that a mandatory secondary task to perform could be included to make participants more entertained and distracted from the experiment's main goal.

6.2.3 Discomfort Surveys

The last survey that participants had to fulfill was Discomfort Survey. We can see in Figure 6.1 the comparison between the answers given before (left) and after (right) the experiment. The possibilities given for each symptom ranged from None (light blue), if the user does not suffer that symptom, to Severe (purple), when the user suffered the symptom noticeably. As we can observe, more discomfort answers appear with darker colors.

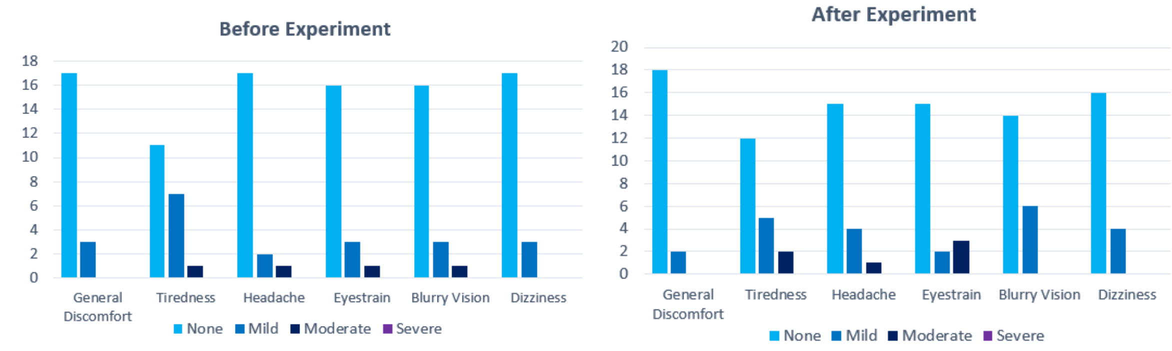


Figure 6.1: Discomfort Survey answers: Comparison among the answers given before (left) and after (right) performing the experiment.

6.3 Experiment Performance

For each participant that has performed the experiment, we have obtained graphs of the audiovisual stimuli presented, which we can use as ground truth. We can see this graphs, which correspond to one of the participants, in Figures 6.2, 6.3 and 6.4.

Visual-only stimuli ground truth is displayed in the first image (6.2), where the y-axis shows the shape of the target spawned while the spawning time is in the x-axis. Similarly, auditory-only stimuli ground truth can be found in the next image (6.3), where x-axis keeps being the timestamp but y-axis represents the type of the spawned audio source. Finally, the last image (6.4) shows the multimodal stimuli cases ground truth, which include visual and auditory cues. X-axis is also the timestamp while y-axis displays the shape of the visual part of the bimodal stimuli, like Figure 6.2. In this last case, dots color does matter since it exhibits the audio source being played, which is the auditory part of the bimodal stimuli.

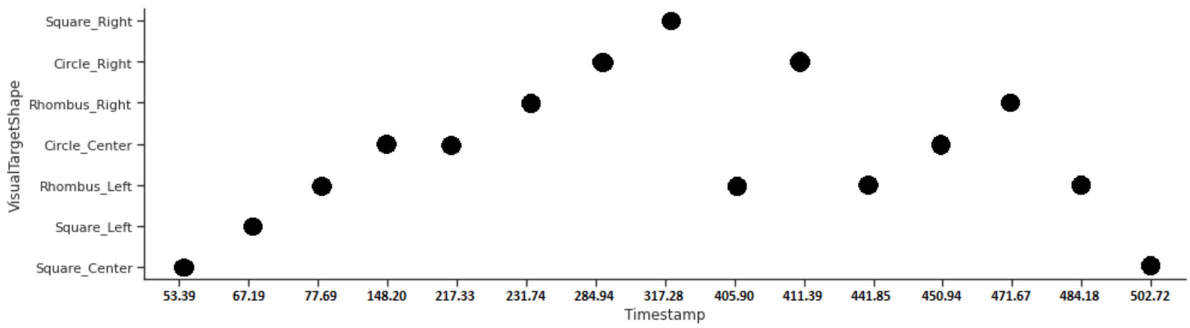


Figure 6.2: Visual stimuli ground truth graph. Each dot corresponds to a visual-only stimuli generated throughout the experiment. The timestamp is displayed on the X-axis while the shape of the visual stimuli is displayed on the Y-axis.

Additionally, another graph is computed from each participant performance data. This chart has on its x-axis the timestamp when the user paused the experiment for having perceived a stimuli. Then, the user communicates the type of the stimuli perceived (visual, auditory or both) which is represented by the dot color. If the user pressed the pause button by error, it also appears in the graph (error case).

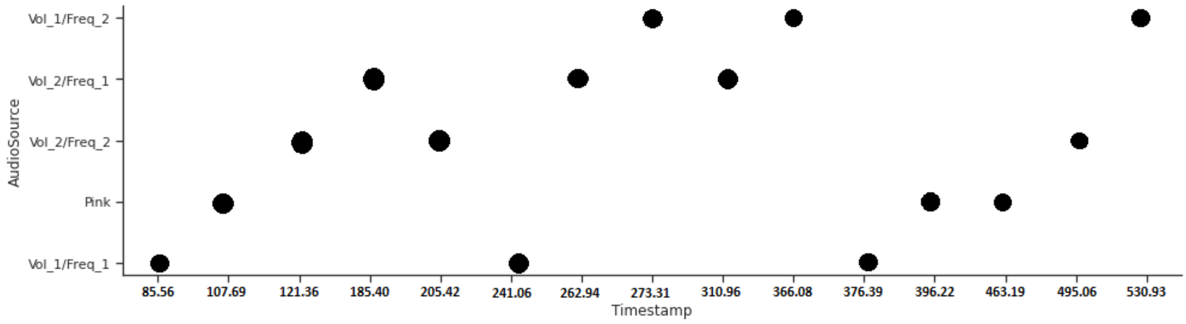


Figure 6.3: Auditory stimuli ground truth graph. Each dot corresponds to a auditory-only stimuli generated throughout the experiment. The timestamp is displayed on the X-axis while the audio source type of the auditory stimuli is displayed on the Y-axis.

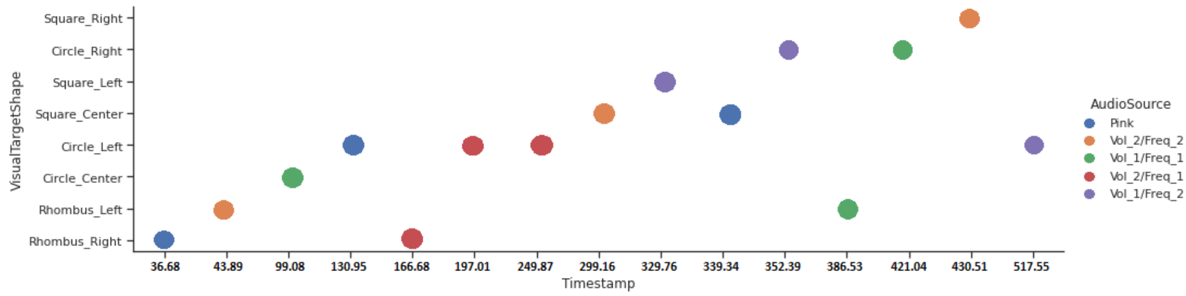


Figure 6.4: Bimodal stimuli ground truth graph. Each dot corresponds again to a multimodal (audiovisual) stimuli generated throughout the experiment. The timestamp is displayed on the X-axis while the shape of the visual stimuli is displayed on the Y-axis. In addition, dots color represents the audio source type that was being played. Therefore, each stimuli keeps both visual and auditory information.

After having classified the stimuli, the visual shape (circle, square or rhombus) seen by the user, if any, appears on its y-axis. As we may notice, auditory-only stimuli do not have visual information neither shape, so they all appear as *None* shape. If the user did not distinguish the shape of the visual targets for the visual-only or bimodal stimuli cases, it appears as *DontKnow*. The performance graph of a participant can be found in Figure 6.5.

Contrasting both data, ground truth and performance graphs, we obtain a comparison graph in which the detection and recognition ratios (y-axis) per each type of audio source (x-axis) for bimodal stimuli cases appear. Combining all participants comparison charts, a total comparison chart which shows the overall result of the experiment is created. This experiment performance chart is presented in Figure 6.6.

The previous graphs show how participants are practically able to detect all the stimuli presented throughout the experiment. It is noticeable how the stimuli cases are generated randomly as desired, with different time intervals between them. Besides, bimodal and unimodal cases are mixed while both visual targets shapes and audio source types are also randomly presented. Consequently, the user is prevented from predicting when a stimuli is going to appear. The performance plot probes how our method is able to trigger the auditory suppression effect. We can observe that there is a clear difference between bimodal and visual-only cases regarding visual performance. Both detection and recognition ratios are significantly lower when an audio source type was ringing compared to visual-only stimuli case, whose ratios are higher.

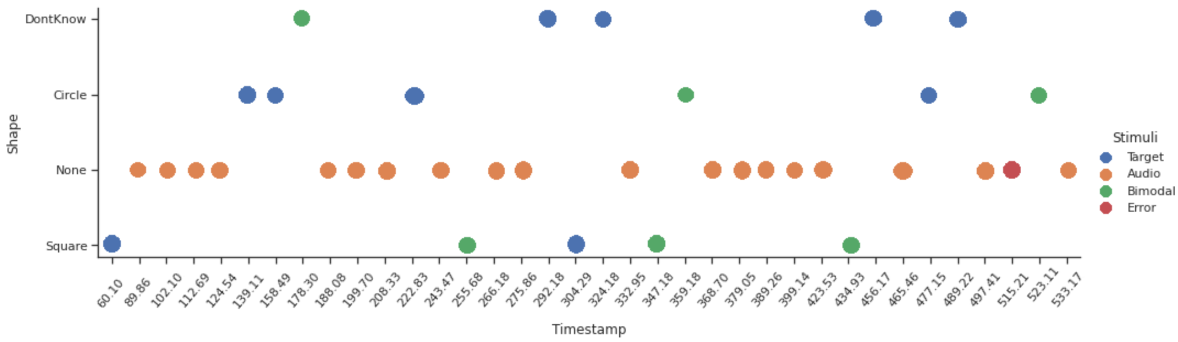


Figure 6.5: Participant performance graph. In this case, each dot corresponds at the moment when the user detected any stimuli. The color of each dot corresponds to what the user declared to having detected, classifying the perceived stimuli in unimodal (audio or target) or bimodal (both). It also appears if the user stopped the experiment by error. Additionally, the visual information of the stimuli (if any) assorted by the user appears on Y-Axis. The options that there was not visual information or that it was not distinguished are also contemplated. Lastly, the detection timestamp is displayed on the X-axis.

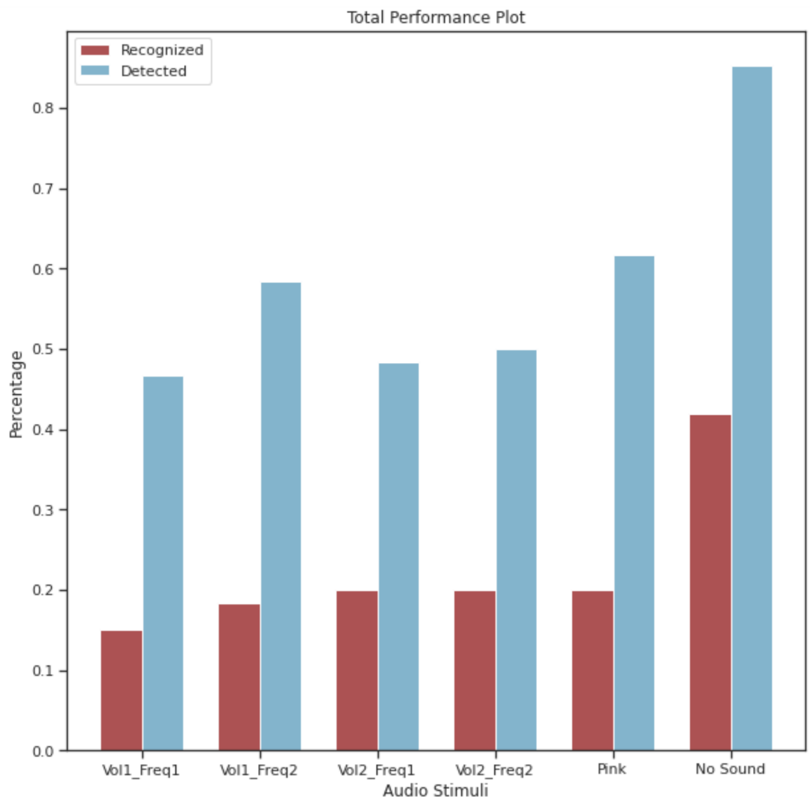


Figure 6.6: Experiment performance plot. Visual information detection and recognition ratios are represented for bimodal and unimodal stimuli. Both detection and recognition ratios values are displayed on the Y-axis, while auditory cases are displayed on the X-axis. These cases include the audio source types related to bimodal stimuli and the visual-only case, without any sound being played.

7. Conclusion

In this last chapter, we discuss the achieved results presented in the previous chapter and comment on the possibilities of future works while proposing new goals. Consequently, it is proposed to acquire new devices with which the user's response to multimodal stimuli could be better measured.

7.1 Discussion and Future Work

As we could see in the final image (Figure 6.6) of the previous Chapter 6, there is a clear difference regarding both detection and recognition rates for visual information in bimodal or unimodal stimuli. Visual only stimuli performance is displayed on the *No Sound* column, achieving a detection ratio of 85.3% and a recognition ratio of 45.0%. These values are significantly higher than the top results achieved with bimodal stimuli. This top-bimodal case corresponds to when the Pink audio source was played, with an achieved detection ratio of a 61.6% and a recognition ratio of 20.0%. This may be because the pink sound used in the experiment was a bit loud and very different to the other auditory stimuli, so it could attract more user attention.

The rest of the bimodal cases achieve even lower detection and recognition ratios, so we can conclude that participants experienced the audiovisual suppression effect when bimodal stimuli was presented. For the cases when $Vol2 - Freq2$, $Vol2 - Freq1$ and $Vol1 - Freq1$ were played at the same time that visual information appeared, detection rates were around 48.3%, which is less than the half, while recognition rates decrease up to less than that 20.0% and becoming 15.0% for $Vol1 - Freq1$ case. Although we can observe how the lowest ratios correspond to the lower frequency limit, upper-frequency limit associated results are also satisfactory in the sense that user's performance has worsened enough in comparison with visual-only (unimodal) stimuli cases.

Therefore, it has been demonstrated how the audiovisual suppression effect still emerges with auditory information almost imperceptible to the participants. It has been proved that near-hearing-limits sounds maintain the capacity of suppressing visual information so the user misses it. We can also mention that the experiment works properly by observing Figures 6.2, 6.3, 6.4 and 6.5. It can be checked how all the desired 45 stimuli cases are presented to the user, who achieves to perceive all of them one way or another. Lastly, commenting that no type of problem or failure has been observed in the experiments performed and that all participants have managed to finish it comfortably.

On the other hand, regarding the surveys fulfilled by users, we can comment that most of them found the virtual experience authentic enough to perceive it as reality. So, we can confirm that strong feelings of presence and immersion were induced in the participants throughout the experiment. We could also relate that the general low recognition rates could be due to the very small visual targets size and its quick disappearing, as reported by the participants. According to Figure 6.1, our experiment does not have significant side effects to worry about. As expected, there were minimum increments in symptoms like tiredness and eyestrain, consequence of performing the virtual experiment by participants who never use VR devices or use them very occasionally in many cases.

It is necessary to mention that the stored eye-tracking data could be used in order to study further what happens with the user's behavior and its reactions when the stimuli are detected and the suppression effect takes place. As reported in Malpica et al. [1], this effect is not related to eye movements like saccades, attention shifts or blinking but to neural phenomena instead. So, knowing that the origin of this effect is not based on eye-tracking data, we decided to skip gaze data analysis due to time restrictions, leaving it as future work consequently.

In the view of the presented results, we corroborate the achievement of our goal which were to study and analyze the relationship between the audiovisual suppression effect and barely audible sounds located at the limits of the hearing range. Next steps could be related to using inaudible sounds to suppress visual information without the user noticing it. This technique could be used in different applications to modify the virtual scene without the user realizing, what would allow to overcome real space limitations or increase recognition task capabilities. In fact, our solution is more versatile than the RDW approach mentioned in Chapter 2. RDW needs powerful enough hardware to perform constantly scene changes when saccades occur while our solution does not have those requirements. In our case, there is not need of such a powerful hardware since subtle scene changes may be performed when the audiovisual suppression effect is triggered. Actually, both solutions could be implemented together to increase suppression performance since their origins are utterly different.

7.2 Physiological Measures

Another possible continuation of our work would be to research what impact have human feelings or emotions in the audiovisual suppression effect. Similarly to how auditory information can cause the loss of visual data, different states of mind or excitement levels could also influence detection and recognition tasks. The idea would be to stimulate various types of emotions, both positives and negatives, and find out if they reduce sensory perception performance.

In this sense, more personalized experiments may be designed, in which participants fears, phobias, hopes or passions could be taken into account in the VE. For example, a participant with fear of heights or the sea could be asked to perform some cognitive task, which requires minimum levels of concentration, on a building roof or in the middle of the ocean respectively. Similarly, receiving an award in front of a cheering crowd or a situation that makes the user laugh are positive feelings instead that can also influence participants focus and performance during the task. In these cases, a similar sensory suppression effect could be triggered but now also considering participants mindset.

In order to make this possible, it is necessary to record different types of user's data other than gaze behaviour. Since our goal now is to analyze users intangible responses, new physiological data have to be obtained from the participant behaviour. This biological measures are, among others, heart rate, stress, temperature and the production of sweat on the skin. There are two kind of devices that offer these kind of measurements and could be used in the experiment, which are an electrocardiogram (ECG) and sensors to measure the galvanic skin response (GSR).

Searching for different options of these devices that are available on the market currently [46], we have selected a couple of them whose capabilities meet our requirements while have good value for money. It is also important to emphasize that the user will have to wear or use this devices while performing the experiment, wearing the HMD and grabbing the controller as well. Consequently, we sought devices whose size was as small as possible, easy to use and comfortable.

Regarding the ECG, the most attractive solution was the smartwatch Fitbit Sense. It includes a wide variety of services such as Global Positioning System (GPS) and phone call management. However, what we are interested on is that it provides an ECG application, electrodermal activity feedback, blood oxygen level, heart and breathing rate monitoring, stress management levels and a mood log [14]. On the other hand, Bitbrain Ring biosensor controls accurately electrodermal skin activity or GSR, which is basically the sweating level on user's skin, and cardiac activity in real time [15].

Both devices, the Fitbit Sense (left) and Bitbrain Ring (right), can be found in Figure 7.1. The first is used to record heart electrical signal and rate while the other supplies GSR measurements. We can also notice how both devices are compatible with the HMD and the controller, since they are placed on two phalanges of the fingers and the wrist, so all of them could be used simultaneously without problem. Therefore, these gadgets might be used to estimate users state of mind and find out how it may influence the audiovisual suppression effect.

Figure 7.1: Fitbit Sense smartwatch used to measure ECG [14] (left) and Bitbrain Ring biosensor used to measure GSR [15] (right).

Bibliography

[1] Sandra Malpica, Ana Serrano, Diego Gutiérrez, and Belen Masia. Auditory stimuli degrade visual performance in virtual reality. *Scientific Reports*, 10:12363, 07 2020.

[2] Proyecto IDIS. Espada de damocles, 1968. `https://proyectoidis.org/espada-de-damocles/`, accessed: 31.05.2022.

[3] Riplee. Ar, mr, vr, the rise of the 3rd screen, 2021. `https://riplee.tv/ar-mr-vr-3rdscreen/?lang=en`, accessed: 19.06.2022.

[4] Grand View Research. Virtual reality market size, share trends analysis report by technology (semi fully immersive, non-immersive), by device (hmd, gtd, pdw), by component (hardware, software), by application, by region, and segment forecasts, 2022 - 2030, 2022. `https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market`, accessed: 31.05.2022.

[5] Graphics and Imaging Lab. Graphics and imaging lab webpage, university of zaragoza, 1999. `https://graphics.unizar.es/`, accessed: 17.05.2022.

[6] Jean-Marie Normand, Elias Giannopoulos, Bernhard Spanlang, and Mel Slater. Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PLOS ONE*, 6(1):1–11, 01 2011.

[7] Antigoni Tsiami, Petros Koutras, and Petros Maragos. Stavis: Spatio-temporal audiovisual saliency network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[8] Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan Mcguire, David Luebke, and Arie Kaufman. Towards virtual reality infinite walking: Dynamic saccadic redirection. *ACM Trans. Graph.*, 37(4), jul 2018.

[9] Google Sites. Rubin vase optical illusion, 2009. `https://sites.google.com/site/illusionsphere/rubin-vase-optical-illusion`, accessed: 04.06.2022.

[10] Jim Eakins. Hand cooling from illusion not linked to change in body ownership, 2017. `https://improbable.com/2017/12/20/hand-cooling-from-illusion-not-linked-to-change-in-body-ownership/`, accessed: 04.06.2022.

[11] Masakazu Ide and Souta Hidaka. Tactile stimulation can suppress visual perception. *Scientific Reports*, 3, 2013.

[12] HTC Corporation. Htc vive pro eye, 2011. `https://www.vive.com/us/product/vive-pro-eye/overview/`, accessed: 11.06.2022.

[13] Tobii. Tobii eye tracking, 2022. `https://www.tobii.com/group/about/this-is-eye-tracking/`, accessed: 12.06.2022.

[14] Fitbit. Fitbit sense smartwatch, 2020. `https://www.fitbit.com/global/es/products/smartwatches/sense`, accessed: 19.06.2022.

[15] Bitbrain Technologies. Bitbrain ring, 2018. `https://www.bitbrain.com/es/productos-neurotecnologia/amplificador-de-biosenales/ring`, accessed: 19.06.2022.

[16] Francisco Javier Pérez Martínez. Presente y futuro de la tecnología de la realidad virtual. 2011.

[17] Eve Museos. Breve historia de la realidad virtual, 2018. `https://evemuseografia.com/2018/03/30/breve-historia-de-la-realidad-virtual/`, accessed: 31.05.2022.

[18] Daniel Martin, Sandra Malpica, Diego Gutierrez, Belen Masia, and Ana Serrano. Multimodality in vr: A survey. *ACM Comput. Surv.*, dec 2021. Just Accepted.

[19] Juan Manuel Davila Delgado, Lukumon Oyedele, Peter Demian, and Thomas Beach. A research agenda for augmented and virtual reality in architecture, engineering and construction. *Advanced Engineering Informatics*, 45:101122, 2020.

[20] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers Education*, 147:103778, 2020.

[21] Lan Li, Fei Yu, Dongquan Shi, Jianping Shi, Zongjun Tian, Jiquan Yang, Xingsong Wang, and Phoenix Zhang. Application of virtual reality technology in clinical medicine. *American Journal of Translational Research*, 9:3867–3880, 09 2017.

[22] Fabian Hutmacher. Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology*, 10, 2019.

[23] Eli Peli. Real vision virtual reality. *Optics Photonics News - OPT PHOTONICS NEWS*, 6:28–34, 07 1995.

[24] Jamie Enoch, Leanne McDonald, Lee Jones, Pete Jones, and David Crabb. Evaluating whether sight is the most valued sense. *JAMA Ophthalmology*, 137, 10 2019.

[25] Angelika C. Kern and Wolfgang Ellermeier. Audio in vr: Effects of a soundscape and movement-triggered step sounds on presence. 2020.

[26] Willem-Paul Brinkman, Allart Hoekstra, and Rene Vanegmond. The effect of 3d audio and other audio techniques on virtual reality experience. volume 219, 01 2015.

[27] David Hecht, Miriam Reiner, and Gad Halevy. Multimodal virtual environments: Response times, attention, and presence. *Presence*, 15:515–523, 10 2006.

[28] Mark R. Mine. Virtual environment interaction techniques. Technical report, USA, 1995.

[29] Nicholas Peatfield, Nadia Müller, Phillipp Ruhnau, and Nathan Weisz. Rubin-vase illusion perception is predicted by prestimulus activity and connectivity. *Journal of vision*, 15:429, 09 2015.

[30] Matthew M. Botvinick and Jonathan D. Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391:756–756, 1998.

[31] Souta Hidaka and Masakazu Ide. Sound can suppress visual perception. *Scientific Reports*, 5, 05 2015.

[32] Widex. El rango auditivo humano, 2022. `https://www.widex.cl/es-cl/blog/rango-auditivo-humano`, accessed: 12.06.2022.

[33] Unity Technologies. Unity solutions, 2022. `https://unity.com/`, accessed: 11.06.2022).

[34] Unity Technologies. Unity download, 2022. `https://unity.com/es/download`, accessed: 11.06.2022.

[35] Unity Technologies. Unity asset store, 2022. `https://assetstore.unity.com/`, accessed: 11.06.2022.

[36] HTC Corporation. Htc vive pro eye, 2011. `https://www.vive.com/us/product/vive-pro-eye/features/`, accessed: 11.06.2022.

[37] VALVE. Steam vr, 2022. `https://www.steamvr.com/es/`, accessed: 11.06.2022.

[38] Tobii. Tobii webpage home, 2022. `https://www.tobii.com/`, accessed: 12.06.2022.

[39] Tobii. Tobii vr, 2018. `https://vr.tobii.com/`, accessed: 12.06.2022.

[40] VIVE Business. Vive pro eye, 2011. `https://business.vive.com/eu/setup/vive-pro/`, accessed: 13.06.2022.

[41] VIVE Developers. Eye tracking calibration, 2011. `https://developer.vive.com/us/support/sdk/category_howto/how-to-calibrate-eye-tracking.html`, accessed: 12.06.2022.

[42] Unity Technologies. 3d living room, 2019. `https://assetstore.unity.com/packages/3d/environments/urban/3d-living-room-62120`, accessed: 13.06.2022.

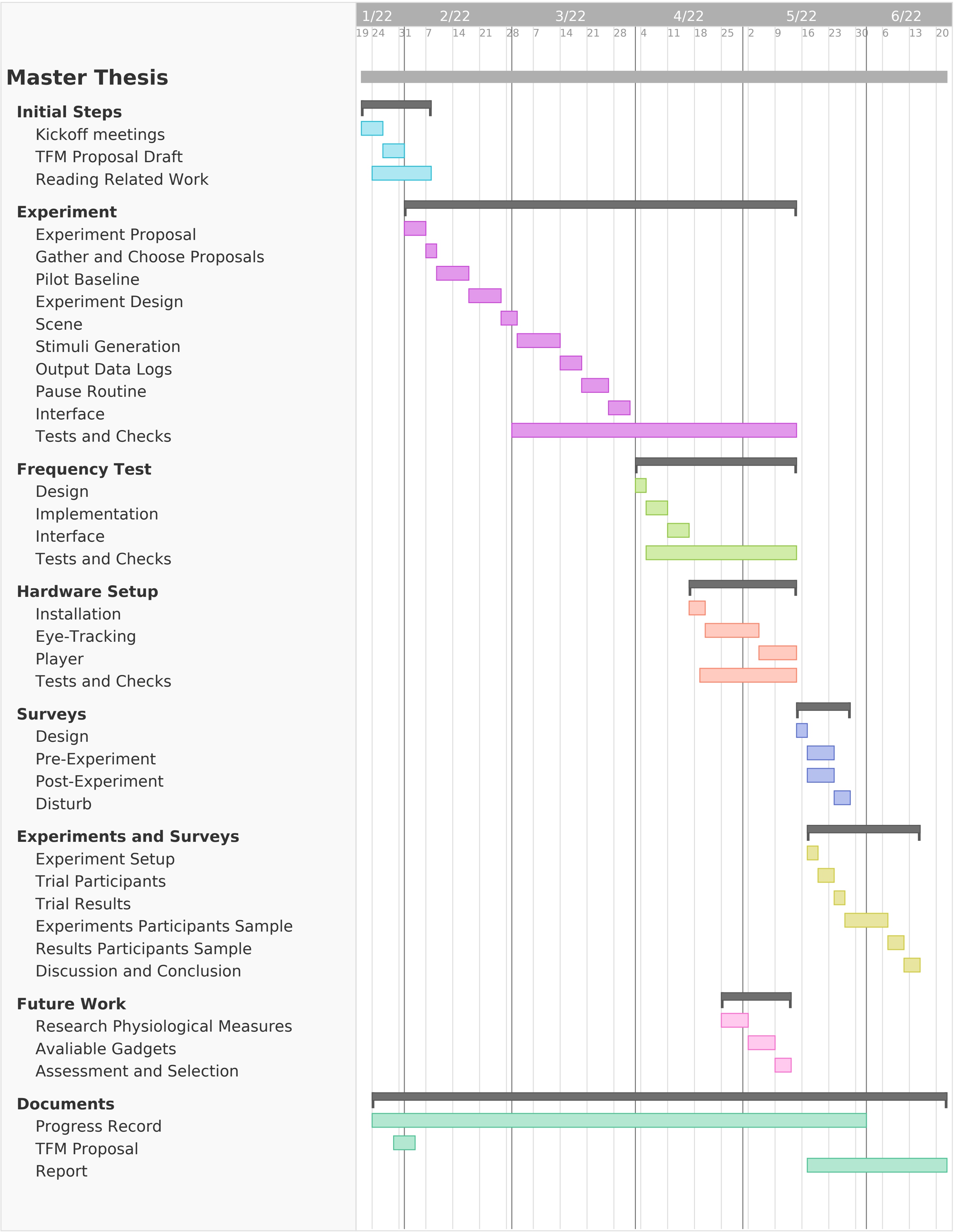
[43] Tobii XR Devzone. Htc vive pro eye development guide, 2022. `https://vr.tobii.com/sdk/develop/unity/getting-started/vive-pro-eye/`, accessed: 14.06.2022.

[44] Valve Corporation. Steamvr plugin, 2021. `https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647`, accessed: 14.06.2022.

[45] Google. Google forms, 2022. `https://www.google.es/intl/es/forms/about/`, accessed: 15.06.2022.

[46] Emily Cronkleton. 9 of the best heart rate monitors of 2022, 2022. `https://www.healthline.com/health/fitness/heart-rate-monitor`, accessed: 19.06.2022.

Anexo A. Gantt Diagram


1/22
2/22
3/22
4/22
5/22
6/22
19 24 31 7 14 21 28 7 14 21 28 4 11 18 25 2 9 16 23 30 6 13 20
Master Thesis
Initial Steps
Kickoff meetings
TFM Proposal Draft
Reading Related Work
Experiment
Experiment Proposal
Gather and Choose Proposals
Pilot Baseline
Experiment Design
Scene
Stimuli Generation
Output Data Logs
Pause Routine
Interface
Tests and Checks
Frequency Test
Design
Implementation
Interface
Tests and Checks
Hardware Setup
Installation
Eye-Tracking
Player
Tests and Checks
Surveys
Design
Pre-Experiment
Post-Experiment
Disturb
Experiments and Surveys
Experiment Setup
Trial Participants
Trial Results
Experiments Participants Sample
Results Participants Sample
Discussion and Conclusion
Future Work
Research Physiological Measures
Avaliable Gadgets
Assessment and Selection
Documents
Progress Record
TFM Proposal
Report

Anexo B. Pre-Experiment Survey

[Daniel J.] Final Master Thesis: Pre-test Survey.

*Obligatorio

Proyecto de investigación: PROTECCIÓN DE DATOS

El siguiente cuestionario forma parte de un TFM que se llevará a cabo por Daniel Jiménez Navarro, dirigido/tutorizado por Ana Serrano Pacheu (IP del Proyecto o Profesor/a del Área de Informática e Ingeniería de Sistemas de la EINA), bajo la responsabilidad de la Universidad de Zaragoza.

El objetivo de este trabajo es observar el efecto de supresión audiovisual que se produce bajo unas ciertas condiciones en entornos inmersivos de realidad virtual.

Si Ud. tiene visión y audición normales o corregidas le invitamos a participar.

Contestar a esta encuesta no le llevará más de 5 minutos. Pero es totalmente libre de completarla o salir sin completarla, en cuyo caso sus contestaciones no se tendrán en cuenta.

Su participación es totalmente voluntaria y gratuita, no le ocasionará perjuicios ni más molestias que el tiempo que emplee en completar la encuesta, pero contribuirá a obtener el conocimiento que necesitamos. Sólo le pedimos que sus respuestas se ajusten lo más posible a la realidad.

Es un cuestionario totalmente anónimo, no incluye datos que permitan identificarle y todas las respuestas serán confidenciales. No obstante, al igual que ocurre cuando realiza una búsqueda en Google o utiliza sus servicios, esta empresa sí podría identificarle y recopilar sus datos. Por ello le informamos que puede acceder a la Política de privacidad de Google y revisar sus ajustes de privacidad en el siguiente enlace: https://policies.google.com/privacy.

En relación con esta encuesta, Ud. puede ejercer sus derechos en materia de privacidad directamente ante Google y, caso de no verlos satisfechos, podrá, si lo desea, dirigirse al investigador principal (741358@unizar.es) o al Delegado/a de Protección de Datos de la Universidad de Zaragoza (dpd@unizar.es) o, en reclamación, a la Agencia Española de Protección de Datos (www.aepd.es).

La Universidad de Zaragoza cuenta con una página donde ofrece amplia información respecto de este tratamiento y de su política de protección de datos, así como formularios para el ejercicio de sus derechos: http://protecciondatos.unizar.es/

Muchas gracias de antemano por colaborar con esta investigación rellenando la encuesta.

1. ¿Ha leído y acepta la Política de Privacidad de la Universidad de Zaragoza: https://protecciondatos.unizar.es/politica-de-privacidad y de Google: https://policies.google.com/privacy? *

Marca solo un óvalo.

◯ Acepto

2. I agree to participate in this research experiment. I understand the purpose and nature of this study and I am participating volutarily. I understand that I may withdraw from the experiment anytime without any kind of penalty or consequence. I consent to the use of the data generated from this questionnarie in the researcher's publications on this topic. Any personal information obtained throughout this study will remain confidential and will be released only under your specific permission. *

Marca solo un óvalo.

- I agree

3. I authorize the recording of this session for further study *

Marca solo un óvalo.

- I agree

4. Personal ID (Please, choose and remember a personal two-digit ID. It will be requested later.) *

____

5. Age *

____

6. Gender *

Marca solo un óvalo.

- Male
- Female
- Other
- Rather not to say

7. Birthplace (as concrete as possible) *

Marca solo un óvalo.

- Zaragoza
- Aragón
- Spain
- Otro: ____________

8. Education (highest level completed) *

Marca solo un óvalo.

- No formal education
- Elementary School
- High School or equivalent
- Baccalaureate Degree
- University Degree
- Master's Degree
- Doctorate (e.g. PhD) or higher
- Other

9. Do you have any vision problem? *

Marca solo un óvalo.

- Yes
- No

10. If you answered 'Yes' to the previous question, please specify what type of problem it is (e.g. poor distance vision, etc.)

11. If you have any vision problems, do you have them corrected? (e.g. by wearing glasesses or contact lenses)

Marca solo un óvalo.

- Yes, I am wearing glasses
- Yes, I am wearing contact lenses
- No, I do not have them corrected

12. Do you have any auditory problem? *

Marca solo un óvalo.

- Yes
- No

13. If you answered 'Yes' to the previous question, please specify what type of problem it is (e.g. age-related hearing loss, partial or total deafness etc.)

14. If you have any auditory problems, do you have them corrected? (e.g. by wearing a ear-mounted assistive device)

Marca solo un óvalo.

- ◯ Yes, I am wearing a ear-mounted device
- ◯ Yes, I have them corrected using other kind of assistive technology
- ◯ No, I do not have them corrected

15. Do you have any characteristics that make you fall into the neurodivergent group (e.g. dyslexia, autism, etc.)? If so, please indicate your condition

16. Add any information you consider relevant

17. Do you play videogames? *

Marca solo un óvalo.

- ◯ Yes
- ◯ No

18. If you answered 'Yes' to the previous question, how much time do spend playing videogames on a daily basis?

Marca solo un óvalo.

- Low (one hour at most)
- Moderate (between one and three hours)
- High (more than three hours)

19. Did you hear about Virtual Reality (VR) before? *

Marca solo un óvalo.

- Yes
- No

20. Have you ever used a Virtual Reality (VR) device? *

Marca solo un óvalo.

- Yes
- No

21. If you answered 'Yes' to the previous question, how often?

Marca solo un óvalo.

- Low (rarely, about 5 times in total)
- Moderate (occasionally)
- High (on a daily basis)

22. If you have ever used a VR device, check those that apply

Selecciona todos los que correspondan.

- [ ] I have tested computer-type devices (HTV Vive, Oculus, PlayStation VR...)
- [ ] I have tested devices that use a smartphone
- [ ] I use virtual reality devices on a daily basis

23. Have you ever experienced eyestrain, sickness, headaches, or nausea when using VR? *

Marca solo un óvalo.

- Yes
- No
- I have never used VR

Anexo C. Post-Experiment Survey

[Daniel J.] Final Master Thesis: Post-test Survey.

*Obligatorio

1. Please, write the personal two-digit ID you chose previously *

2. Did you have any problem wearing and using the VR glasses (HMD)? *

   Marca solo un óvalo.

   - Yes
   - No

3. If you answered 'Yes' to the previous question, please specify what problems did you encounter (e.g. size, malfunction...)

4. In general, have you found easy to perform the experiment? *

   Marca solo un óvalo.

   - Yes
   - No

5. If you answered 'No' to the previous question, please specify what kind of issues did you encounter (e.g. operativity, scene, comfort, practice...)

6. Throughout the experiment, were you able to distinguish the visual targets easily? *

Marca solo un óvalo.

- Yes
- No

7. If you answered 'No' to the previous question, please specify what made it difficult to distinguish the visual targets (e.g. location, color, shape, size...)

8. Would you include other types of visual targets? *

Marca solo un óvalo.

- More complex shapes
- Scene objects
- Videos
- Other

9. Write any comment or suggestion that you would like to point out or find it interesting regarding the visual targets (if any)

____

10. Throughout the experiment, were you able to listen the audio sources easily? *

Marca solo un óvalo.

- Yes
- No

11. If you answered 'No' to the previous question, please specify what made it difficult to listen the audio sources (e.g. volumen, location...)

____
____
____
____
____

12. Write any comment or suggestion that you would like to point out or find it interesting regarding the audio sources (if any)

____

13. Throughout the experiment, have you felt any kind of dizziness or discomfort * (headache, eyestrain...)?

Marca solo un óvalo.

- Yes
- No

14. Throughout the experiment, were you able to predict when a audio source would sound and a visual target would spawn? *

Marca solo un óvalo.

- Yes
- No

15. Please, rate your feeling of being in the virtual environment on the following scale from 1 to 5, where 5 represents your normal experience of being in a real place. I had the feeling of "being there" in the virtual environment: *

Marca solo un óvalo.

	1	2	3	4	5	
Not at all						Completely

16. How long during the experience the virtual environment was like reality for you? *

Marca solo un óvalo.

	1	2	3	4	5	
Never						All the time

17. Throughout the experience, which was stronger in general: your feeling of being in the virtual environment or of being in another place? *

Marca solo un óvalo.

	1	2	3	4	5	
Virtual Environment						Another Place

18. When you look back on your experience, do you remember the living room space more like images you saw or more like a place you visited? *

Marca solo un óvalo.

	1	2	3	4	5	
Images I saw	⬭	⬭	⬭	⬭	⬭	Place I visited

19. In order to improve the experiment and the user experience, please report any comment, suggestion or review that you may have

Anexo D. Discomfort Survey

[Daniel J.] Final Master Thesis: Discomfort Survey.

*Obligatorio

1. Please, write the personal two-digit ID you chose previously *

2. Session *

   Selecciona todos los que correspondan.

   - [ ] Before
   - [ ] After

3. According to your current condition, indicate the degree of the following symptoms: *

   Marca solo un óvalo por fila.

	None	Mild	Moderate	Severe
General Discomfort	◯	◯	◯	◯
Tiredness	◯	◯	◯	◯
Headache	◯	◯	◯	◯
Eyestrain	◯	◯	◯	◯
Blurry Vision	◯	◯	◯	◯
Dizziness	◯	◯	◯	◯

Anexo E. Frequency Test Results Table

		FREQUENCY TEST	
User ID	Date	Lower Limit (Hz)	Upper Limit (Hz)
1	30/05/2022	200	13500
2	31/05/2022	200	13300
3	31/05/2022	150	11800
4	31/05/2022	120	12650
5	01/06/2022	150	13200
6	06/06/2022	100	15500
7	06/06/2022	200	14600
8	07/06/2022	200	10200
9	07/06/2022	200	13000
10	07/06/2022	100	15700
11	07/06/2022	200	13000
12	07/06/2022	200	14200
13	08/06/2022	300	14800
14	08/06/2022	200	12000
15	08/06/2022	200	15000
16	08/06/2022	200	14300
17	09/06/2022	300	13000
18	09/06/2022	200	12700
19	09/06/2022	200	8300
20	09/06/2022	200	11000
	Mean	**191**	**13087,5**
	Standard Deviation	**110,03**	**1824,96**

Figure E.1: Frequency Test results table. For each user, the lower and upper frequency values related to the hearing range limits are displayed. The mean and standard deviation for both frequency values are calculated as well.