DEPARTMENT OF
COMPUTER SCIENCE

CAROLINA RIBEIRO DIAS FERREIRA
Master of Science

CREATING IMMERSIVE AUDIO IN A HISTORICAL SOUNDSCAPE CONTEXT

COMPUTER SCIENCE

NOVA University Lisbon
November, 2021

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTMENT OF
COMPUTER SCIENCE

CREATING IMMERSIVE AUDIO IN A HISTORICAL SOUNDSCAPE CONTEXT

CAROLINA RIBEIRO DIAS FERREIRA

Master of Science

Adviser: Maria Armanda Simenta Rodrigues
Associate Professor, NOVA University of Lisbon

Co-adviser: Sofia Carmen Faria Maia Cavaco
Assistant Professor, NOVA University of Lisbon

Examination Committee:

Chair: Henrique João Lopes Domingos
Associate Professor, NOVA University of Lisbon

Rapporteur: Valter Nelson Noronha Alves
Associate Professor, Polytechnic Institute of Viseu

Adviser: Maria Armanda Simenta Rodrigues
Associate Professor, NOVA University of Lisbon

COMPUTER SCIENCE

NOVA University Lisbon
November, 2021

Creating Immersive Audio in a Historical Soundscape Context

To my family.

Acknowledgements

Firstly, I would like to start by expressing my gratitude to my advisers, Professor Dr. Armanda Rodrigues, and Professor Dr. Sofia Cavaco for all the support and guidance throughout the elaboration of this dissertation. It was a pleasure to be taught by you over the course of the past five years.

The work presented in this dissertation was funded by the Portuguese Foundation for Science and Technology under the projects PASEV (PTDC/ART-PER/28584/2017) and NOVA-LINCS (PEest/UID/CEC/04516/2019) in the form of a research grant, which I am very grateful for. I would like to thank the entire PASEV team for their support and feedback during the design phase of the proposed solution and its evaluation process.

A special thank you should also be extended to Sofia de Castro, Encontro de Sons, and all the actors involved in the interpretation of the audio clips kindly provided to the project. Your readily availability and amazing work were deeply appreciated.

Additionally, I would also like to thank all the volunteers who took their time to participate in the user study conducted during the evaluation and validation phase of this dissertation.

I want to express my deepest gratitude to NOVA School of Science and Technology (NOVA University Lisbon), and all its community members, for providing me with a warm welcome from day one, and for assuring a quality learning environment to allow me to grow as a person and as a professional. Particularly, I would like to thank the Computer Science and Informatics Department. From my colleagues to my teachers, everyone contributed to a good experience that I will most certainly always remember very fondly.

I want to leave many thanks to my family, in particular, to my brother, my mother, and both my grandmothers, for all their love, their patience, and the immeasurable amount of support throughout my entire academic journey.

Last, but definitely not least, I would like to thank my friends. You make everything a lot more enjoyable. I will always appreciate the time we spent working together, debating, joking around, growing, and learning together. There are a lot of fun moments spent on, and outside, the campus that I will never forget. I am very thankful for your friendship, our little adventures, and the support you gave me when I most needed it. A special thank you note to my longtime best friend, Gabriel Botelho, whom I have known for many years now. Thank you for being so patient and kind; for being a loyal project partner and above all, for the sincerest friendship and comradeship.

To my classmates, I wish you lots of success, both personal and professional. I hope you all do great things and I get to keep hearing from you. Thank you.

Abstract

A soundscape is the representation of an acoustic environment in a specific context and time. Soundscapes evolve over time according to the changes in social, cultural and political contexts, and they constitute a unique way of experiencing auditory memories of a specific period in history. Considering that the soundscape of a place is ever-changing and it can provide us with important information about a location's context in time, the study and preservation of soundscapes, in particular historic soundscapes, has gained relevance in the cultural heritage protection and promotion process. Soundscapes enrich the history of places and contribute to enhance the visitors experience in tourism contexts.

However, it is impossible to go back in time and record the acoustic experiences from centuries ago. For that reason, the closest we can get to experience an historic soundscape is by trying to recreate it. This can be done with the support of historic documentation and audio immersion technology.

We created a solution that allows users to construct scenes that provide immersive auditory experiences. The immersive audio files generated are obtained by combining different current sound recordings, and by composing an immersive auditory scene that can be recorded by the user. Here, we propose Immerscape, a tool that non-expert users can easily use to create immersive historical soundscapes. This work was validated by a user study that evaluated the tool's functionalities and its acceptance by users.

The proposed work is a contribution to the PASEV project, whose main purpose is to preserve and promote Évora's rich cultural heritage, in particular, their historic soundscapes corresponding to the period between 1540 and 1910.

Keywords: Soundscape, Immersive Audio, Spatial Audio, Immersive Tourism, HRTF, Applied Computing, Digital Heritage

Resumo

A paisagem sonora é uma representação de um ambiente acústico num contexto e período de tempo específicos. As paisagens sonoras evoluem ao longo do tempo de acordo com mudanças em contextos sociais, culturais e políticos, e constituem uma maneira única de experienciar memórias auditivas de um específico período histórico. Visto que as paisagens sonoras de um local são mutáveis e nos podem fornecer informação importante acerca do contexto do mesmo numa determinada época, o estudo e a preservação de paisagens sonoras, em particular de paisagens sonoras históricas, tem adquirido uma especial relevância no processo de proteção e promoção da herança cultural. Isto acontece visto que as paisagens sonoras enriquecem a cultura e história dos locais e contribuem para melhorar a experiência dos visitantes em ambientes turísticos.

Contudo, é impossível voltar atrás no tempo para gravar os contextos sonoros experienciado há séculos atrás. Por isso, o mais perto que podemos estar de experienciar uma paisagem sonora histórica passa por tentar recriar essa mesma paisagem. Para isso, devemos recorrer a documentação de suporte relativa à época e ao uso de tecnologias de áudio imersivo.

A solução apresentada nesta dissertação tem como objetivo permitir aos utilizadores construir cenas que ofereçam uma experiência auditiva imersiva. Os ficheiros imersivos são obtidos através da combinação de diferentes sons atuais, compondo-os numa cena auditória imersiva que pode ser gravada pelo utilizador. É apresentada Immerscape, uma ferramenta que utilizadores não-experientes podem facilmente usar para criar paisagens sonoras históricas. A ferramenta foi avaliada através de um estudo com utilizadores que avaliaram as suas funcionalidades e sua aceitação.

O trabalho proposto nesta dissertação é uma contribuição para o projeto PASEV, cujo principal objetivo é preservar e promover a rica herança cultural da cidade de Évora, com especial foco nas suas paisagens sonoras correspondentes aos períodos entre 1540 e 1910.

Palavras-chave: Paisagem Sonora, Áudio Imersivo, Áudio Espacial, Tourismo Imersivo, HRTF, Computação Aplicada, Herança Digital

Contents

List of Figures

Acronyms

3D	3-Dimensional 4, 22, 32, 34, 38, 42, 44, 46, 49, 51, 58
AAR	Audio Augmented Reality 33, 34, 39
AI	Artificial Intelligence 30
AR	Augmented Reality 4, 22, 34, 38
BVI	Blind or Visually Impaired 33, 39
EPSRC	Engineering and Physical Sciences Research Council 43
FCT	Faculdade de Ciências e Tecnologia 36
FFT	Fast Fourier Transform 11
GPS	Global Positioning System 32, 34
HRTF	Head Related Transfer Function 6, 15, 17, 18, 19, 27, 28, 30, 32, 33, 34, 42, 44, 46, 105
Hz	Hertz 11
IDE	Integrated Development Environment 46
ILD	Interaural Level Difference 17
IRCAM	Institut de Recherche et Coordination Acoustique/Musique (Institute for Research and Coordination in Acoustics/Music) 19
ITD	Interaural Time Difference 17
meSch	Material Encounters with Digital Cultural Heritage 31, 39
MIT	Massachusetts Institute of Technology 19
MR	Mixed Reality 22

NOVA	NOVA University 36
PASEV	Patriomonialização da Paisagem Sonora em Évora v, 3, 4, 5, 6, 9, 29, 32, 34, 36, 38, 40, 43, 44, 45, 46, 54, 104
PPA	Peak-To-Peak Amplitude 10, 12
SADIE	Spatial Audio For Domestic Interactive Entertaiment 19, 43
SARIM	Sound Augmented Reality Interface for visiting Museum 34
SDK	Software Development Kit 42, 43, 57, 58
UC	University of California 19
UGC	User Generated Content 30
UI	User Interface 23, 51, 53, 55, 57, 59
UK	United Kingdom 19
UNESCO	The United Nations Educational, Scientific and Cultural Organization 1
USA	United States of America 19
UX	User Experience 23
VR	Virtual Reality 22, 42
WFS	Wave Field Synthesis 21
WWI	First World War 31
XR	Extended Reality 22

1 Introduction

This chapter makes an introduction to the problem we intend to solve. It presents the motivation to further explore immersive audio as an option to create historical soundscape contexts. It also presents the research project where the proposed solution was integrated. In addition to that, it defines a set of objectives and contributions, and briefly describes the corresponding proposed solution to achieve them.

1.1 Motivation

Évora, the capital of the Alentejo Province, is a Portuguese city that holds, in its historic centre, more than twenty centuries of history. Its vast variety of architectural styles and monuments contribute to a cultural heritage that goes as far back as the Celtic times. Moreover, its cityscape is a unique place for understanding the influence exerted by Portuguese architecture. Évora presents one of the finest examples of a city of the golden age of Portugal, as it was not affected by the 1755 earthquake that destroyed Lisbon and erased part of that city's history. All of these factors contributed to UNESCO's recognition of the historic centre of Évora as a World Heritage Site [64].

The concept of cultural heritage is not limited to physical aspects, as it can include both the tangible and intangible [65]. While the presence of a specific architecture style can be considered an example of a city's tangible heritage, other aspects such as soundscapes and language are examples of intangible cultural heritage. Considering Évora's World Heritage Site status, and its historical significance, it is essential to find ways to preserve and promote its culture. One of the most effective ways to do this is through tourism. By creating an appealing touristic experience that shows the city's authenticity in a new interesting manner, more people will want to explore the city's heritage, keeping its legacy alive.

Music and sound, in general, have been used in touristic contexts for some time now, through their use in audio-guides, ambient music in exhibitions' settings, and to interact with the visitors through different interfaces in museums and other cultural activities.

However, with the advancements in technology, new ways of using audio in touristic contexts emerge, such as through mobile technology. Nowadays, mobile devices, especially smartphones, are more prevalent than ever in people's lives, and as their popularity keeps showing a growing tendency [58], new opportunities to improve touristic experiences through the use of mobile technology become more and more appealing. Additionally, over the years, smartphones, and other mobile devices, have acquired more processing power and the quality of the cameras and speakers they integrate have been substantially improved. This allows the creation of new touristic solutions that make use of the new modern devices' features to extend on the preexisting traditional touristic offer. The use of mobile technology in tourism, meant to provide visitors with additional, and better quality information, has shown to have a positive impact in the visitor's engagement and learning process [5, 6].

When exploring audio-visual contexts, sound is simply used as an integrated component of the complete captured scene, making us often neglect the capture of "sound memories". Image capture is the main source of memory collection and the capture of audio is often neglected, as it is most commonly used as a complementary feature of image capture through video. Sound recordings should be more valued. Besides the ability of sound to add information to a scene, it is also an effective contributor to the creation of a sense of presence [27]. Consequently, it has an ability to trigger strong instinctual emotions in the listener [56]. A good example of that is the use of sound, particularly of music, in films. Depending on the course of action of a movie, different sound effects and music will be used to complement and transform the visual information of a scene. In this context, the importance of sound to convey emotion can be easily put to the test by watching a movie without sound. Part of the emotion and effect of a scene is lost when, for example, you watch a horror film without the suspenseful background music that our brains interpret as the anticipation of danger. Due to all this, the emotional and informative nature of sound are important concepts when we consider immersion in virtual environments.

The use of audio can be more powerful than we realize, as it may give us specific information about a setting and directly influence the way we perceive our environment. Humans' perception of their environments is strongly related to their spatial awareness, which is influenced by hearing [27]. Since that, in most computer systems, the visual and auditory channels are the only ways to communicate with the user, both should be utilized to provide the user with the most complete and realistic experience possible.

In the natural world, most objects are able to produce sound. Consequently, urban environments have complex and ever-changing soundscapes. These can be eventually lost in time unless preserved. In order to fully experience the environment of a city, especially in a different context from its current one, there must be sound involved, as it provides us with an added degree of immersion that ultimately increases our sense of presence and might even be able to “take us back in time”. Sounds and music, in particular, are often connected to specific locations and generally have a major emotional impact on the listeners [27]. Due to the sound links to locations, removing sound also removes part of the sense of the life-likeness of an experience.

Therefore, in order to preserve Évora's historic soundscape, while contributing to create a new way for tourists to experience the city's cultural richness, we have developed an audio immersion tool that supports sound recreation of several historical scenes, associated to the city of Évora. The integration of the generated sounds into the PASEV project's database allows users to have access to an immersive historic sound experience associated to the sites they are visiting, when using one of PASEV's digital platforms. This way, the sounds will be available whether the user is using the application on site, as a complementary feature to the visit, or at home, when the user is just exploring the map and timeline available on the website, as well as the related multimedia content.

An additional benefit of this approach is that the immersive audio is available independently of the user's location, which might be interesting in situations such as the ones we are currently living, resulting from the COVID-19 pandemic. At home, or even in outdoor spaces, the users are able to use their own devices to enrich their touristic experience while enjoying Évora's cultural heritage. The immersive audio, paired with images or video, can also provide a great opportunity for users to independently explore audiovisual content related to the cultural sites without having to leave the comfort of their own homes.

1.2 Context

The PASEV project, which stands for *Patriomonialização da Paisagem Sonora em Évora*, or in english, "Patrimonialization of Évora's Soundscape" is a project set to contribute to the patrimonialization of Évora's soundscape. Its goal is to chart as many historical sound events as possible in the city of Évora, referring to the chronological period between 1540 (creation of Évora's Archbishopric) and 1910 (beginning of Republic in Portugal), while exploring new perspectives on the cultural manifestations of the city, as well as finding new interesting ways to experience the city's rich history [42].

The notion of soundscapes allows the understanding of musical reality based on a wide-ranging contextual perspective, as well as the reconstructions of historical contexts and the mapping of the presence of sound in a city's history. Évora's intense musical activity in churches, squares, theatres as well as other social events that have taken place over the centuries, make the city's historic soundscape an extremely rich one. Therefore, one of PASEV's main focus is the creation of a digital platform that gathers information about the city of Évora in several formats such as image, video and audio, associating these testimonies with the points of interest in a map of the city, that evolves with the progression of the timeline.

In addition to the compilation of visual and audio content, the platform aims to provide its visitors with an interactive and immersive experience of the city. In order to achieve that, the information is meant to be available through various virtual interactive interfaces, easily accessible to the general public. For that purpose, a web platform [49, 14], a responsive mobile application, interactive games interfaces [66], and an AR mobile application [44] have been developed in the context of the PASEV project.

The access to the platform through a mobile application (e.g. via smartphone), while visiting the city, can enrich the touristic routes by enabling the creation of themed road maps for visitors looking to explore the city by following thematic itineraries. Examples of these itineraries are a church organs circuit, or a road map following a historical religious ceremony, like the procession of Senhor Jesus dos Passos.

The development of the project involves a multidisciplinary team that includes contributors from several institutions and research units, including NOVA LINCS (FCT). This research unit is responsible for two main tasks [42]: the development of the aforementioned digital platform; and the domain analysis and determination of the geographic context of the city, a process that is responsible for collecting all relevant information on Évora's soundscape domain [49, 14, 66, 44, 32].

1.3 Objectives & Research Questions

As previously mentioned, PASEV is a project centred on charting as many historical sound events as possible in the city of Évora from 1540 (creation of Évora's Archbishopric) to 1910 (begining of Republic in Portugal). For this purpose, the project aims at providing audio content, related to the soundscapes of that period, to visitors through all its interactive interfaces.

However, making historical soundscapes available to the general public is a real challenge, due to the inexistence of original audio captures. An obvious solution to tackle this limitation is to capture or synthesize audio that resembles what might have been those historical soundscapes. Due to that, the PASEV team captured some current sound recordings characteristic of the city.

Since it is not possible to go back in time to experience or record Évora's historic soundscapes, the main goal of this dissertation is to provide the project's team with support to recreate these soundscapes, so that visitors can gain insight into what it was like to be at the city of Évora centuries ago.

For that purpose, our goal was the development of an audio immersion tool that can be used to integrate and convert sounds obtained from specifically selected recordings, into immersive 3D audio files to be reproduced over headphones. This process included the the composition and spatialization of sound in order to achieve the intended sense of user immersion, while only using the visitors' devices and headphones.

These immersive audio files are meant to be integrated into PASEV's database, and consequently become available on any of the project's, current or future, interactive interfaces. This integration provides the end-user with an improved and more involving experience when exploring the city of Évora. In addition to that, the immersive audio files also serve as records of the city's soundscapes, contributing to enrich its heritage, while preserving it.

Considering all the aforementioned, the research questions that this dissertation aims to answer are the following:

1. **How can we create immersive audio to represent a historical soundscape?**
2. **Can we create a tool that helps non-expert users to create immersive audio representations of historical soundscapes?**
3. **How easily can non-expert users create representations of historical soundscapes using the immersive audio tool?**

1.4 Solution Overview & Contributions

As briefly presented in section 1.3, the main goal of this dissertation is to create immersive audio that represents historic soundscape scenes, associated to the city of Évora. There are multiple ways in which this can be achieved. Artefacts such as books, paintings, ancient objects, monuments and other written documentation can provide us with considerable information to support the creation of a historical soundscape through simulation of the historical acoustic environments [67]. Still, since it may be hard to create a complete environment that simulates a scene from the past, historical soundscape designers may have to deal with and combine several independent recordings from different audio sources.

The auditory scene building process required the support from documentation about the social and political context, as well as environmental descriptions of the historic events we were trying to represent through the soundscapes. This information was provided by the history research experts part of PASEV's team.

The immersive audio created as a result of this dissertation is meant to be integrated into PASEV's database. Consequently, the audio files will be made available to the end-users through all of PASEV's digital interfaces. This includes access through the location cards in the web platform, through the geo-referenced markers, the augmented reality platform and mobile games interfaces being currently developed within the context of the project, as well as eventually any other type of platform that is connected to the database.

Audio processing is required to give spatial dimension to the generated audio, whether it is played on loudspeakers or through headphones. However, playing immersive audio on loudspeakers requires an environmental setup that allows the sound to be integrated into the listener's surroundings. On the other hand, when using headphones, the immersive sound is directly fed into the listener's ears. The end-user might want to listen to the immersive audio at home through the website or on site, for example, using the augmented reality application. Taking that into consideration, the audio immersion techniques used to create the historic soundscapes assume that, in order to get the full audio immersion quality, the users must wear headphones connected to the device where they are playing the immersive audio. Therefore, our proposed solution only focuses on delivering audio immersion through headphones.

In this dissertation, we aim to provide mechanisms that inexperienced users can use to create immersive historical soundscapes and generate immersive audio files that can be reproduced over headphones. Therefore, we propose Immerscape, a tool to create historical soundscapes from current audio recordings.

The tool can convert sounds obtained from specifically selected recordings, into immersive 3D audio files. This process includes the composition and spatialization of sound in order to generate audio files with the intended sense of user immersion in the historical scene, while only using the visitors' devices and headphones to reproduce them. Immerscape is designed for users with no programming skills such as historians and other non-engineers for whom it may be hard to deal with more technical audio development applications and integrated development environments (IDE). Using an interactive interface, Immerscape aims to provide simple mechanisms for users to build immersive scenes and record the audio from within those scenes without having to worry about any of the technical aspects of 3D sound generation.

The selected sounds samples will have to undergo some treatment as preparation for the composition of the scenes. The previously described audio immersion feeling is then simulated by the strategic combination of sounds and the application of the HRTF binaural filters used by Immerscape.

To demonstrate the technique, a working prototype of Immerscape was built as a standalone application. For validation purposes, the prototype underwent a user study using a subjective quality inference mechanism. This study focused on the functionality of the tool, the quality of the audio files generated by users while collecting their feedback for future adjustments. The evaluation of the Immerscape prototype was performed in a series of ten tasks.

Considering the proposal, implementation and validation of the proposed technique, a summary of the main contributions of this dissertation's work is presented below:

- **Identification, collection and analysis of the sounds needed to recreate specific historical scenes associated to PASEV's project.**

- **Proposal, design and implementation of an immersive audio tool using an audio immersion technique.**

- **Validation of the proposed technique through users' evaluations of the tool's effectiveness and functionality.**

- **The immersive audio files generated with the tool and during the implementation of the exploratory initial solution.**

1.5 Document Outline

This document is divided in 6 main chapters:

- **Chapter 1 - Introduction:** The chapter makes an introduction to the problem we intend to solve. It presents the motivation to further explore immersive audio as an option to create historical soundscape contexts. It also presents the research project where the proposed solution is integrated. In addition to that, it defines a set of objectives and expected contributions, and briefly describes the corresponding proposed solution to achieve them.

- **Chapter 2 - Fundamental Concepts:** The chapter provides the basic knowledge related to the concepts that are the foundation of this work. It is intended to give the reader a better understanding of what a soundscape is and what some of the basic principles of sound are. It also aims to explore fundamental concepts related to binaural hearing, audio immersion techniques, spatial audio, the integration of 3D sound in different contexts, as well as introduce some basic concepts of human-computer interaction.

- **Chapter 3 - Related Work:** The chapter presents examples of applications that implement similar and relevant characteristics to the ones we are exploring, as well as some related work and research done on topics concerning different audio immersion techniques that reflect the current state of the art in the spatial audio field. It also presents some of the previous work done on the research project that supports this dissertation.

- **Chapter 4 - User Generated Soundscape Tool:** The chapter gives a detailed description of the proposed solution to the presented problem. It discusses the initial approach taken and how the proposed solution evolved from there. It gives an overview of the technology and techniques used in the development of the system, as well as it discusses the design and development process and details of the prototype's implementation.

- **Chapter 5 - Evaluation and Results:** The chapter introduces the evaluation and validation of the the proposed solution. It presents the first user study performed on the prototype, its purpose and its results. Additionally, it includes the discussion of the results obtained and the analysis of the participants' feedback.

- **Chapter 6 - Conclusion:** The chapter overviews the developed work during the dissertation, discussing the main results of the proposed solution, its strengths and its weaknesses. It also briefly introduces the future work that should be done to improve the current version of the proposed solution.

2

Fundamental Concepts

This chapter provides the basic knowledge related to the concepts that are the foundation of this work. It is intended to give the reader a better understanding of what a soundscape is and what some of the basic principles of sound are. It also aims to explore fundamental concepts related to binaural hearing, audio immersion techniques, spatial audio, the integration of 3D Sound in different contexts, as well as introduce some basic concepts of human-computer interaction.

Most of the topics discussed in this chapter are based on the books *Fundamentals of Hearing* by W. Yost [69] and *Spatial Audio* by F. Rumsey [50]. Readers with knowledge of acoustics or experience working with sound can skip this chapter.

2.1 Soundscape

The acoustic environment of a place, that is, the sounds present at that place, provides rich information about the living beings, type of setting (such as urban or rural), weather conditions, as well as other important characteristic of the setting like time. The acoustic environment of a place is ever-changing, and it evolves over time according to the changes in social, cultural and political contexts. As thus, it can provide relevant information about a location's context in time.

A soundscape is the representation of an acoustic environment in a specific context and time as perceived by humans. It is the combination of sounds that form an immersive environment, coming from different sources of elements that in space and time [57, 53].

Consequently, soundscapes are inescapable. Due to our inherent ability to hear, and the dynamic evolution property of any landscape, every place is characterized by a distinctive soundscape that uniquely describes that specific location in time [2].

In conclusion, soundscapes provide us with important information about a location's context. For that reason, the recreation of soundscapes in a specific place can help us travel back in time when visiting a location. Artefacts such as books, paintings, ancient objects, monuments and other written documentation can provide us with considerable information to support the recreation of a historic soundscape [67].

The study and preservation of soundscapes, in particular historical soundscapes, has gained relevance in the cultural heritage protection and promotion process [48, 65, 47]. As stated in PASEV's website [41], the notion of soundscapes makes it possible to understand musical activity from a wide-ranging contextual perspective. Therefore, if we consider Évora's rich musical presence over the centuries and the prevalence of other urban aspects, such as the everyday life activity from churches, convents, squares and theatres, soundscapes enable us to reconstruct contexts, that in some form, recreate the environment in the city at some point.

2.2 Sound

As previously mentioned, the existence of sound around us is constant and ever-changing, and its perception is dependent on each person's individual hearing capabilities. Sound provides us with important information about our environment. Among other things, it helps us navigate spaces and stay alert for potential danger, which makes our ability to determinate the sources of sounds one of our most important biological traits [69].

When we listen, the sound that reaches our ears, or more accurately, our tympanic membrane, also called eardrum (see Figure 2.1), is a complex sound that combines all the distinct sounds from different sources in our surroundings. In order to determine the different sources, our nervous system must translate (transduce) the various physical aspects of sound into a *neural code* for the complex sound field. This auditory neural information is processed and integrated with our other sensory systems, eliciting a response from humans that is the culmination of the hearing process [69]. In subsubsection 2.3.2.2, we will explain how the outer ear, especially the concha and pinna (represented in Figure 2.1), influence the way each individual hears sound.

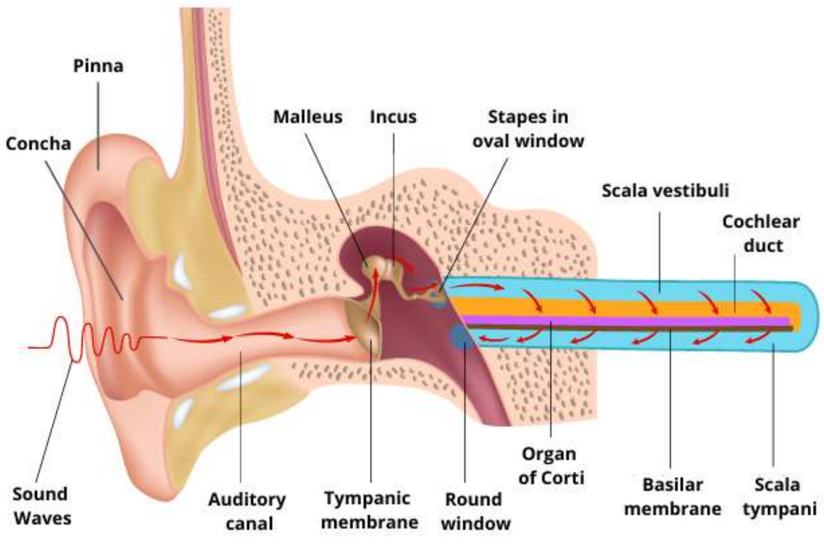


Figure 2.1: Anatomy of the Human Ear [20].

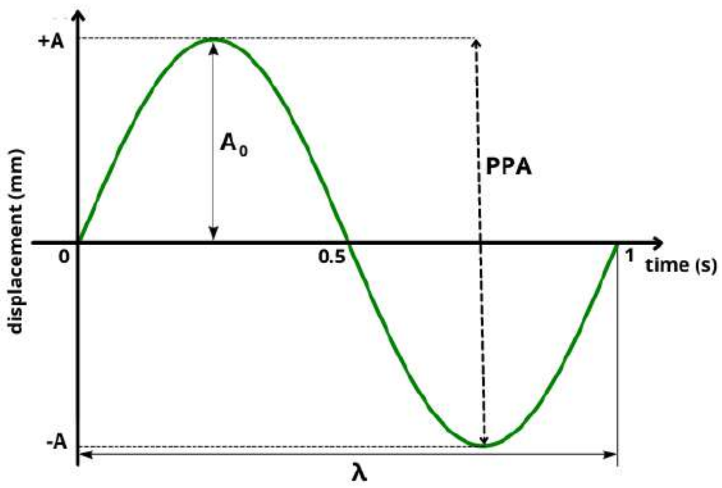


Figure 2.2: Graphic representation of a sine wave over one cycle [21].

2.2.1 Basic Notions of Sound

Any object with the properties of *inertia* and *elasticity* might be set into vibration, and hence, produce an audible sound. In other words, the vibration of an object is what makes sound possible, which in practice means that since almost every object has inertia and elasticity, almost every object can be set into vibration and produce sound. However, the real world contains sources of resistance that prevent objects from vibrating forever.

A sound is a composite vibration that can be resolved into the sum of several particular simpler vibrations, called sinusoidal vibrations. Sinusoidal vibrations are the basic building blocks of all vibrations and, hence, of all possible sounds.

A sinusoid, also called a sine wave, describes a particular relationship between displacement and time. Displacement simply refers to the distance an object moves from its initial position, and the sine wave describes the continuous, regular, back-and-forth displacement of a vibrating object. In Figure 2.2, we can see the graphic representation of the sine wave over one complete cycle. In this context, A_0 refers to the sine's instantaneous amplitude and PPA refers to the peak-to-peak amplitude of the sine wave. In addition, λ refers to the wavelength of the sine wave, which is the distance between adjacent identical parts of the wave [39]. The period of a wave is the time required to complete one full cycle, which means the period is the amount of time it takes a wave to travel the distance of one wavelength. For instance, considering that, in Figure 2.2, the wavelength is 1 second, we say that the period of the wave is 1 second.

There are three main properties that characterize a sinusoid: frequency, the instantaneous starting phase and amplitude [69]. Once we have specified the values of these three parameters (frequency, starting phase and amplitude), we have uniquely described a single sinusoid.

Since any vibration consists of a sum of one or more sinusoidal vibrations, we can completely describe any vibration by specifying the amplitude, frequency and starting phase of the various sinusoids that constitute the vibration. If a vibration is constituted only by a single sinusoid, it is called a pure tone or waveform, whereas if a vibration is constituted by more than a sinusoid, it is called a complex vibration or complex waveform [69].

Joseph Fourier derived important theorems for the flow of heat and his analysis can be applied to the types of vibrations we have for sound. Fourier derived a theorem that specified that any vibration can be resolved into a sum of sinusoidal vibrations, called a Fourier series, which can describe almost any arbitrary vibration. The amplitude, frequency and starting phase of the sinusoids that compose complex sounds can be derived from the Fourier Analysis.

When we use computers to process sound we are working with discrete signals. For that reason, in order to solve Fourier Analysis problems, an algorithm called Fast Fourier Transform (FFT) is often used as it calculates the Discrete Fourier Transform efficiently for a discrete signal [69].

Frequency

The frequency is a measure for the number cycles a repeating event is completed per unit of time [69]. It is usually measured in Hertz (number of cycles per second), and it is inversely proportional to the period of the wave (expressed in seconds), defined as the time needed to complete a cycle.

For example, considering the waveform in Figure 2.2, we can say the period is 1 second because it completed one cycle in one second and the frequency is 1 Hz. This means the vibrating object only oscillated once over its period time. As a consequence, the period may be used alternatively to the frequency to describe the oscillation:

$$frequency = \frac{1}{period} \tag{2.1}$$

These two terms may also be used to describe the repetition occurring in many non-sinusoidal vibrations (complex vibrations) as long as these vibrations are periodic (have a pattern that repeats itself in time).

Starting Phase

The starting phase indicates the point in the displacement cycle of the vibrating particle at the instant in time it begins to vibrate (and produce sound). It is usually defined in degrees of angle. A sinusoid is said to start at zero phase, or to start “in phase”, if when the time is equal to zero (t=0), the displacement D is equal to zero too. Two sinusoids are said to be “out of phase” with each other when they have the same frequency but different starting phases, like it is shown in figure 2.3 where θ represents the out-of-phase angle.

Amplitude

Amplitude generally describes the amount of vibratory displacement of a sine wave,

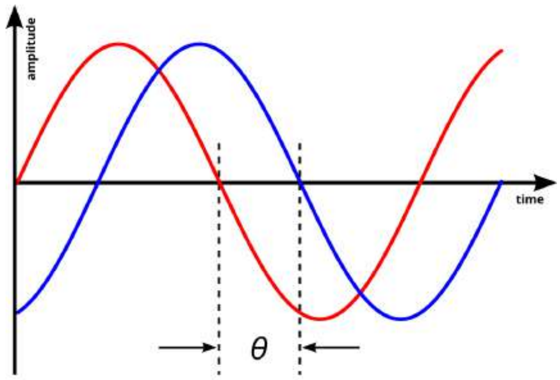


Figure 2.3: Two out of phase sine waves where Θ represents the out of phase angle [22].

or in other words, how far a vibrating particle moves from its initial position describes its amplitude. The time-varying displacement is called the instantaneous amplitude of a sine wave. It can be expressed by the following equation, where A corresponds to the maximum amplitude, f is the frequency of the wave, t, the time instance and θ, the starting phase:

$$D(t) = A\sin(2\pi f t + \theta) \tag{2.2}$$

Definitions of amplitude that do not vary with time include peak amplitude, peak-to-peak amplitude, and root-mean-square amplitude. The larger the displacement the higher the amplitude value. The peak amplitude refers to the maximum positive displacement the waveform achieves in one period, while peak-to-peak amplitude or PPA (see Figure 2.2) is the total distance from the maximum positive peak to the maximum negative displacement peak in one period.

For non-sinusoidal waveforms we might not wish to describe the amplitude with peak values. For instance, the complex waveforms might have much larger peak and peak-to-peak amplitudes than simple sinusoids, yet over the period of the complex waveform, very little displacement occurs. If we simply take the average of the instantaneous amplitudes over one period of a since wave, we will always have zero average because half of the wave has positive values and the other negative values. However, if we square each instantaneous amplitude, all the negative numbers become positive. By computing the root-mean-square (rms) amplitude, we can compare simple and complex waveforms.

Psychophysics is a scientific field that studies the correlation between physical stimuli and the sensations and perceptions they produce [8], including the human perception of these sound parameters. On a subjective, psychological basis, we generally associate changes in the amplitude as loudness and changes in frequency as pitch. We do not have a single subjective term for the starting phase because we are only sensitive to changes in starting phases under particular conditions. The relationships between the physical descriptions and the subjective descriptions of the parameters of a sine wave are somewhat complex.

For instance, although changes in amplitude are almost always perceived as loudness changes, variations in frequency, which are independent of intensity changes, can also result in perceived loudness changes. For that reason, when referring to the physical stimulus, we should only use the terms, frequency, amplitude, and starting phase, while when referring to how a person perceives the stimulus, we can use the subjective terms pitch, and loudness [69].

2.2.2 Sound Transmission

As explained before, in order for a sound to be heard, the vibration produced by the object must propagate until it reaches our ears. Sound can travel through any elastic medium that has inertia. In other words, sound can travel through air, water, steel, and so on, but not through vacuum since it does not possess those two properties. Considering air is the most common medium for sound transmission in most everyday situations, and it is how sound is transmitted in the contexts we are trying to recreate in this dissertation, we will consider air as the standard medium for sound propagation.

Air is composed of molecules in constant aleatory movement. However, the molecules of air do not move all the way from the vibrating object to the receiver's ears. Instead, they propagate the sound's vibration in a falling domino effect and in a spherical, three-dimensional, manner. When the air molecules vibrate, the vibration is propagated in a wavelike motion to the closest molecules, compressing them as the object vibrates. Consequently, this creates an area with a greater density of air molecules. This area of higher pressure is called a condensation area. When the vibrating object moves back, the density of molecules decreases, creating a rarefaction area. The pattern repeats itself until the vibration reaches the human ear and the hearing process is started (see Figure 2.4).

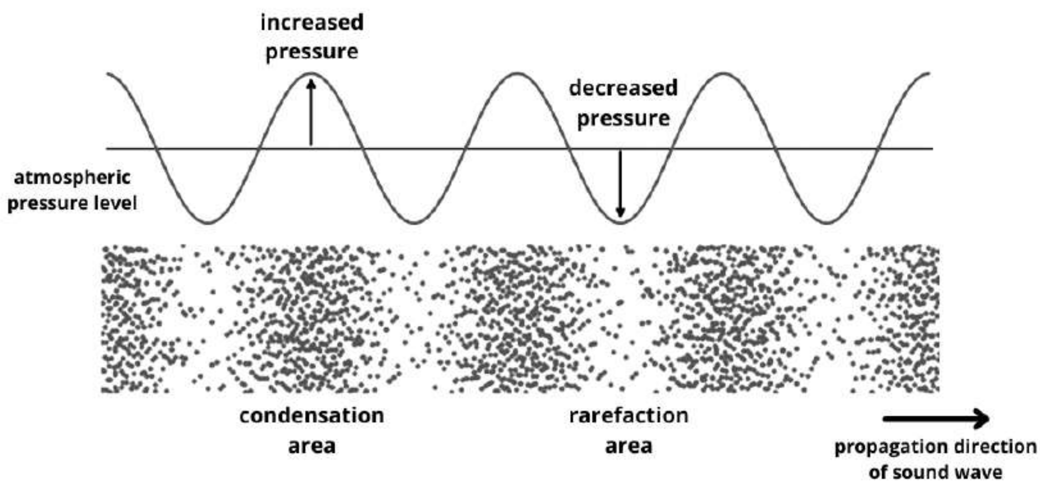


Figure 2.4: Propagation of Sound [55].

The farther away we are from a sound source, the softer the sound becomes, as sound waves, most often than not, encounter obstacles, and these impede the propagation of a sound wave. The impedance of sound transmission occurs whenever there is a change in medium (e.g., sound travelling through air finds a brick wall) and it affects sound intensity. The change in impedance causes a portion of the sound wave to be reflected, while the rest is either transmitted or absorbed through the new medium. In short, we must be careful when assuming a sound wave arriving at some point in a room is identical to the waveform that left the source [69].

Another important concept to introduce when considering sound interference during propagation is a sound shadow. A sound shadow is an area near an object with no wave propagation or very reduced wave magnitude. “Small”, “Medium” and “Large” objects are determined in relation to the wavelength (the distance between any two successive identical points) of sound. Objects that are much larger than the sound wave and with different characteristic impedance, reflect sound; objects that are much smaller do not affect wave motion; while “Medium” objects (approximately equal to wavelength) may produce sound shadows [69]. The fact that the human head can produce a sound shadow has importance implications for our ability to use our two ears to locate sound which will be further discussed in the Binaural Hearing section.

2.3 Binaural Hearing

Before we further explore the concept of Binaural Hearing, the natural human ability of hearing using two ears (two channels), we briefly introduce other types of audio: Monaural Audio and Stereophonic Audio.

Monaural Audio

Monaural sound, often shortened to mono, is the most basic sound recording and reproduction method. It refers to sound that is represented by one channel only. In the case of headphone use, or multiple channels output, the monaural audio is equally reproduced on all output channels.

Since its introduction, monaural audio has been widely replaced by stereophonic audio in most of its entertainment applications. However it is still remains relevant in some contexts, such as in its use in phones, radio and hearing aids [50].

Humans can localize the full three-dimensional direction of a sound with reasonably high accuracy, and although it has its own set of challenges, humans can even localize objects using a single ear, in a process that is known as monaural localization. While it is possible to locate an object using a single ear, much of the spatial information is lost, causing a decrease in accuracy in the representation of the environment [52].

Stereophonic Audio

Stereophonic sound, most often referred to as stereo sound, is a method of sound reproduction that creates a multi-directional audible experience. This is possible because stereophonic audio uses two or more independent audio channels to record and reproduce sound (see Figure 2.5). The reproduction depends on the configuration of two of more stereo output mechanisms, which can either be loudspeakers or stereo headphones [50].

A good example of this is the latter, where the left and right channels receive different output signals. This is an approach closer to reality than the one used in monaural sound reproduction, since the use of more than one channel to reproduce the audio gives the receiver the impression that sound is heard from various directions, as it is in natural hearing. Stereo audio is widely present in modern reproduction devices, especially those used in entertainment systems such as television, recorded music, computer audio and cinema [50].

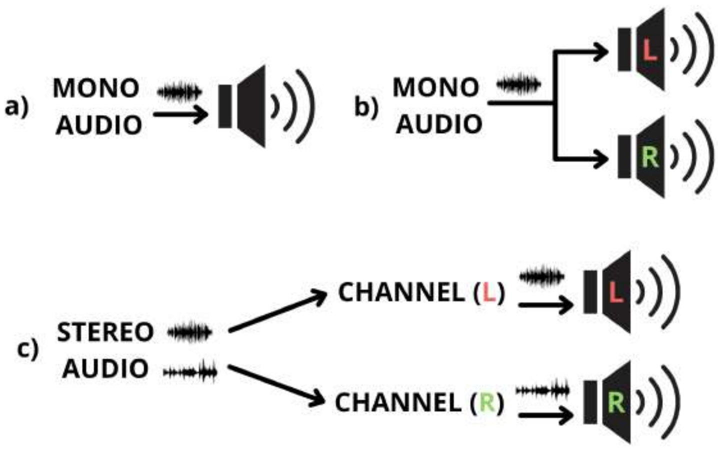


Figure 2.5: Mono vs Stereo Audio: Mono sound reproduced over one output channel (a); Mono sound distributed on two output channels (b) and Stereo audio, distributed through two output channels (c).

2.3.1 Binaural Audio

Binaural audio also sometimes referred to as binaural stereo, just like stereophonic audio, uses two channels to record and reproduce sound. Although it could be argued that, ultimately, all sound reproduction is binaural, since the end-listener uses two ears, the binaural audio term is usually reserved to describe signals that have been recorded or processed to replicate how a human would hear and locate a sound.

To sum up, the main difference between stereophonic audio and binaural audio is in the recording process. The method of recording binaural audio consists in using two microphones located in the ears of a real or dummy head (Figure 2.6) to replicate as closely as possible all the spatial cues received by a human during the hearing process. The microphones are placed at the same fixed distance from each other and positioned in the dummy head just like our ears would be. The simulated anatomical presence of the listener's ears, head and shoulders, create a field called head-related transfer function, or also commonly called, HRTF (see subsubsection 2.3.2.2 for further information) [50].

Figure 2.6: Example of a dummy head setup used to record binaural audio [36].

When reproduced over headphones, or other earpieces, binaural sound can recreate a remarkable sense of realism as it is able to provide spatial cues to the listener that resemble those created by the human body.

Nonetheless, the anatomical differences between the dummy head, where the sound was recorded, and the listener's head may destroy the subtle spectral and timing cues required for success (for more information, see subsection 2.3.2.2). In addition to that, the recording on the dummy head is done in a head-locked position, disregarding the important effect of head movements that allow natural listeners to resolve spatial confusions [50, 12].

2.3.2 Sound Localization

Most often than not, sound travels through environments that include a wide range of elements that affect the original sound signal emitted by the vibrating object. Due to the various forms of interference that exist in the real world, especially in closed spaces, the signal that reaches the human ear is often not the same original sound signal emitted by the source. Instead, it is a combination of waves and the results of their reflections and diffraction that arrives at the ears at different times and with different levels of intensity (see subsubsection 2.3.2.1 for further information).

The way the human body's anatomy affects the sound waves, is particularly relevant for determining the sound sources' locations. The human head acts as an additional source of reflection and diffraction, the process in which the sound wave is spread out as a result of passing through a narrow aperture or across an edge. By decomposing the sound, the human auditory system is capable of processing its physical parameters, correlating them with the spatial location of the source, and consequently, being able to localize sounds in the three spatial dimensions (represented in Figure 2.7) [69]:

- **The horizontal plane (azimuth)**, also known as the left-right dimension, where the interaural time and interaural level differences (see subsubsection 2.3.2.1) are relevant cues to localize sounds;

- **The vertical plane**, also known as the up-down dimension, where the spectral differences provided by HRTF (see subsubsection 2.3.2.2) are cues used for sound localization;
- Distance, or alternatively called, **the near-far dimension**. A powerful cue to determine the sound in distance is the sound's perceived loudness or level (e.g. loud sounds appear closer than other far sounds).

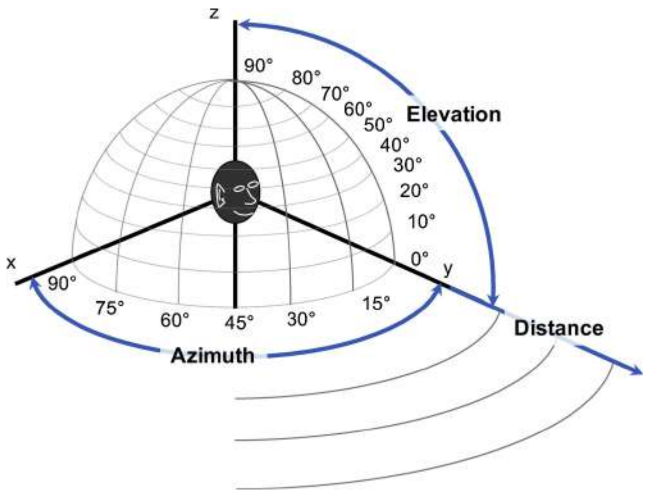


Figure 2.7: Polar coordinates used to locate a sound source in 3D space centered on the listener: azimuth in horizontal plane, elevation or height in vertical plane, and distance as depth [46].

2.3.2.1 Interaural Differences

Interaural differences is a term used when referring to the differences in signals received by the two ears, and it involves the detection of differences in time, or phase, amplitude and spectral cues between the two ears [69].

Depending on the person's position, and the angle of incidence formed between the source of the sound and the listener, sound can arrive at different times to the right and left ears, as it takes some additional time to travel to the more distant ear (binaural delay). We call this **interaural time difference (ITD)** (see left side of Figure 2.8).

The time differences cues are particularly registered at the start (onset) and end (offset) of sounds. They are relevant cues for stimulus location at low frequencies and for complex stimuli with low-frequency repetition. These differences between ears are maximal when the source is located either to the left or right of the listener (in direct line with one of the ears), and null when the source is located either to the center-front or center-back of the listener (source at the same distance to both ears).

Besides interaural time differences, there are also **interaural level differences (ILD)** (see right side of Figure 2.8), related to intensity differences of signals arriving at the two ears. These are the relevant cues for localizing sounds at high frequencies.

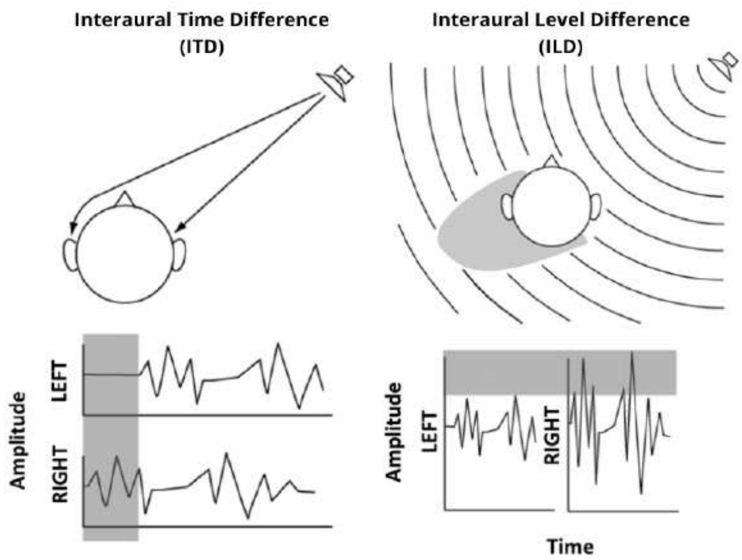


Figure 2.8: Interaural Time Difference (ITD) and Interaural Level Difference (ILD) [61].

The intensity of the sound signal is dependent on the interference it encounters while travelling from the vibrating source until the listener's ears. As mentioned previously, among the variety of objects that can interfere with the sound signal, is the listener's head. Although the human head is not a considerable barrier to low-frequency sounds, it can also originate reflection and diffraction. These events have an impact on the incident sound wave, changing it while originating new ones (the reflections) that will eventually also reach the listener's ears, and consequently, cause significant interaural level differences.

Additionally, just like any other object, the human head can produce a sound shadow (reduced wave motion area) if its size is close to the sound's wavelength. Since a signal's wavelength is directly proportional to its frequency, the interaural level difference caused by a sound shadow is dependent on the signal's frequency [69, 50].

2.3.2.2 Head Related Transfer Function (HRTF)

As explained before, most often than not, the sound signal emitted by a vibrating object is not the same signal that reaches the listener's ears. In fact, the listener's body, head, shoulders and even pinna interfere with the sound wave during its transmission, acting as a small sound shadow [69]. In addition, the shape of the pinnae's components (helix, concha, etc.) causes reflections of the original sound wave, that when added to the original sound result in spectral changes to the final sound signal. These spectral changes depend on the direction of the sound and humans learn to relate these spectral changes to this direction. The combination of these spectral changes are called head related transfer functions (HRTF). Thus, the HRTFs describe how the human anatomy, namely the listener's torso, head, and pinnae change the amplitudes and phases (adds a phase shift) of a sound's individual frequency components (that is, the individual sinusoids that compose the sound), as the sound travels from the source to the ears [50].

The spectral cues caused by the HRTFs are particularly important in human sound localization, as they provide localization cues for the vertical plane, the azimuth and for monaural localization [69]. Nonetheless, they vary from person-to-person. This happen because humans' pinnae vary quite widely in shape and size. Therefore, so do each person's HRTFs, making it difficult to generalize the spectral characteristics suitable for a large group of people. When experimenting with feeding another person's HRTF to someone, while blocking the own person's pinnae effect, researchers observed that the listener's localization ability was noticeably reduced [50]. However, after a short while, the listener seemed to be able to adapt to the new information and regain some of their ability to locate sound sources using a different HRTFs set.

Sound systems, such as headphones, because they are directly in the ear or right over the pinna, tend to create an 'in-head' localization perception. As a solution for this, researchers attempted to design headphones that stimulate the concha resonance.

The concha resonance, created by the main cavity in the center of the pinna (see Figure 2.1), is believed to be responsible to give a sense of externalization, or in other words, gives us the perception that the sounds' sources are outside our heads and not inside. Headphones that were able to stimulate the concha would hence improve 'out-of-head' perception during headphone listening.

The HRTFs related individuality has implication for binaural audio processing. This motivated researchers to work on HRTF characterization, in order to find important features for directional perception, which can be generalized.

For binaural reproduction to work well, the HRTFs of sound sources from the environment must be accurately recreated at the listener's ear upon reproduction. Using methods such as principal components analysis and feature extraction, researchers have been able to generalize HRTFs that work reasonably well for a wide range of listeners [50]. Since the HRTFs are such a crucial element of realistic spatial sound creation, many researchers have worked on gathering HRTF measurements in databases that can be used by developers working with audio. The SADIE Binaural database, developed by the University of York (UK) [68] is an example of a repository containing information and resources that can be used in order to build binaural output. However, it is not the only HRTF related collection available. Other examples include the KEMAR dataset provdided by the MIT Media Lab (USA) [31], the CIPIC HRTF, a public-domain database of high-spatial-resolution HRTF measurements made available by UC Davis (USA) [9] and the Listen HRTF database which is a result from the measurements performed in the context of the Listen project by the Room Acoustics Team at IRCAM (France) [24].

2.4 Audio Immersion

Our everyday lives are full of complex three-dimensional sound experiences that help humans make sense of their surroundings. Often, our real life sound environments are composed of multiple sound sources, spread across different areas, and as seen previously, humans have the ability to locate them in space [50].

Spatial audio is the term used to describe any audio that allows us to determine where sound is coming from. It can contain cues in all three dimensions (width, height, depth), allowing the listener to pinpoint if the sound is above, below, to the left or to the right, or completely surrounding the listener. While stereophonic audio allows to determine left and right sources, it does not give the listener a complete spatial perception, and that is what spatial, immersive, audio aims to do [59, 50].

The accurate spatial reproduction of sound can create an enhanced sense of realness, especially in virtual environments. The most common, and simplest, way of reproducing sound with some spatial sense is by reproducing it using two-channels. However, immersive audio systems, such as the surround sound systems, that use more than two-channels to reproduce sound, have evolved over the years to provide us with an even better immersive experience. The quality of the immersive sound experience relies on a precise reconstruction of a sound field. In order to achieve that, and although the techniques available all present some practical constraints, approaches like the Ambisonics (see subsection 2.4.2) and Wave Field Synthesis (see subsection 2.4.3) have been developed as methods to provide accurate spatial reproductions of sound. Nowadays, immersive audio has applications in a variety of industries, including entertainment, air traffic control, guidance systems and, with special relevance, the gaming industry [11].

2.4.1 Sound Panning

Sound Panning is the distribution of a sound signal into a new stereo or multi-channel sound field determined by a pan control setting [50]. This is typically done using a pan control console that has analog controls with position indicators to distribute the incoming sound into the output channels. However, there is also audio mixing software where the same process of sound panning can be done. For example, panning all the way to the left, places the sound only on the left speaker of a stereo pair, while panning all the way right, places it solely on the right side of the stereo pair. It is also possible to place the sound somewhere between the two extremes [60]. To sum up, panning changes the intensity of the signal in the different channels. In order to simulate sound coming from the right, the intensity of the sound on the right channel should be higher. This works because humans use ILDs (see subsubsection 2.3.2.1) to locate sound. The interaural level differences is what sound panning tries to emulate with the controlled distribution of a sound signal into a new sound field [4].

However, this technique presents some limitations. It only works for the azimuth plan (for higher frequencies) and it does not allow to distinguish if the sound is coming from the front or the back in some cases. For example, if one sound is coming towards the listener from a 45º angle and another sound is coming from a 45º+90º angle (90º to the right ear), the listener cannot distinguish the two directions [69].

2.4.2 Ambisonics

Ambisonics is a surround sound approach that places the listener inside a 360-degree sonic sphere, simulating a spatial sound field where information seems to come from all directions and not just from two speakers in front of the listener. Developed in the mid-1970s by a group of British academics, including Gerzon, Fellgett and Barton, Ambisonics is an extension of Blumlein's M&S stereo, while being compatible with both stereo and mono audio, and using just a four-channel recording called 'B-Format'. It uses whatever speakers are present in the surround sound system, involving any number of reproduction channels, to project a sound image around the listener, decoding it from its 'B-Format' format.

Despite its elegance, the original system was quite possibly ahead of its time and therefore, it did not achieve commercial success. However, ambisonic techniques have evolved, and since then been adopted by content platforms, such as YouTube and Facebook, as their default immersive audio format over headphones, potentially marking ambisonics resurgence in the commercial sphere [50, 17].

2.4.3 Wave Field Synthesis

Wave field synthesis (WFS) is a spatial sound field reproduction technique that places virtual sound sources in real space, giving humans the same auditory perception as if it was a real sound field. This happens because the synthesis is precise and physically correct, in a way that we cannot distinguish the synthesized source from a "real" wave field, at frequencies below the upper frequency limit of the system.

In order to create the sound field, the concept works with the wave theory and the generation of wave fronts using a large number of small and independently controlled loudspeakers (see Figure 2.9), which are fed with corresponding signals calculated by algorithms based on the Kirchhoff-Helmholtz integrals.

One of the clear advantages of this approach is how it allows the listener to perceive the accurate placement of the virtual sound sources independently of their position in the active sound field, not relying on the listener being in the "sweet spot" like other commercially available surround sound techniques. For that reason, WFS presents itself as a good solution for sound reproduction systems where it is possible to use a large number of loudspeakers, particularly in audiovisual areas such as concert halls or cinemas [26, 7].

Figure 2.9: Wave Field Synthesis experimental setup with an array of small loudspeakers [23].

2.5 Augmented Reality

Augmented reality (AR) is an interactive experience of a real-world environment enhanced by overlaying computer-generated information onto the physical world. AR hardware can take various forms including devices we already use in our everyday lives, such as handheld displays (like our smartphones) or devices you can wear (like headsets and glasses), giving them new and more interactive opportunities to be helpful throughout our day (e.g. pointing our camera at something to search the web visually) [16, 19].

AR falls under the umbrella term of extended reality (XR) and it is often compared with virtual reality (VR) and mixed reality (MR). The terms refer to the different levels of virtual component an extended reality experience can have, as represented in Figure 2.10. While in AR digital elements appear over the real-world views, in VR, users are isolated in a virtual world, typically requiring the use of fully immersive headsets devices for the purpose. Finally, mixed reality (MR) is designed to combine AR and VR elements; creating an immersive, semi-isolated user experience that interacts with real world elements by bringing them into the virtual world [16].

As explained in previous sections, most humans have a natural and innate ability to perceive sound spatially, in a process referred to as binaural hearing (see subsection 2.3.1). Considering that, in the context of this project, we intend to provide immersive audio through both a website and an augmented reality applications, we propose that the best way to deliver the 3D sound to the user is through the use of headphones, in order to provide an experience that is as close to real life as possible, while somewhat abstracting the listeners from all the other sounds in their surroundings [29].

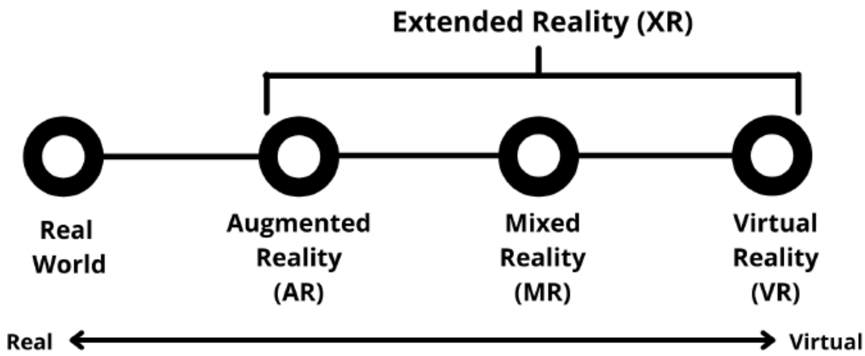


Figure 2.10: Extended Reality Spectrum.

2.6 Human-Computer Interaction

Human-Computer interaction is a field of study focused on the relationship established between human users and computational systems. The connection between the two is made through an interface. Therefore, the effectiveness and usability of interfaces influence directly the quality of communication between humans and the systems they interact with [3].

In order to better understand Human-Computer interactions and to determine what principles can maximize the effectiveness of computer systems when communicating with humans, we must explore two main fields: User Experience Design (subsection 2.6.1) and User Interfaces Design (subsection 2.6.2).

Although User Experience (UX) and User Interfaces (UI) are both essential aspects of human-computer interactions and they are inertly related, it is important to distinguish between the two. While UX focuses of the overall experience, UI design is mainly concerned with how the interfaces look and function.

2.6.1 User Experience (UX)

We should distinguish between User Experience (UX) and Usability as UX is a much broader concept that besides concerning usability also considers another quality attribute, utility.

2.6.1.1 Norman's Design Principles

According to Don Norman, all systems should be intuitive and easy to use besides working correctly. In his work, Norman proposes six design principles [25] that should be considered while developing user-centered interfaces:

- **Visibility:** The available functionalities should be visible as this enables the user to know what to do next and how to use the system.

- **Feedback:** The system should always provide feedback to the user. The feedback lets the user know if the action was successful or not.
- **Constraints:** This refers to the design concept of restricting certain interactions at a given moment, narrowing the user's options.
- **Mapping:** The controls used to interact should closely resemble their effect (e.g. the up and down arrows on the keyboard are used to represent the up and down movement).
- **Consistency:** In order to improve the user's learnability and the recognition of actions, interfaces should have similar operations and use similar elements for achieving similar tasks.
- **Affordance:** The way things look on a interface should hint at how users should interact with them (e.g. a mouse button invites pushing).

2.6.1.2 Usability

Usability can be described as the ease of use and acceptability of a system meant for a group of users carrying out specific tasks in a specific environment. Its 'ease of use' affects the user's performance and satisfaction, while the 'acceptability' determines whether or not the product is used at all. In other words, usability measures how efficiently, effectively and at what level of satisfaction, a user interacts with a system [38].

The definition of usability encompasses different aspects and it can be defined by five main components [54]:

- **Learnability:** This can be defined as the amount of time it takes a target user to become sufficiently familiar with the system in order to perform the actions necessary to complete a task.
- **Efficiency:** Refers to the ease and amount of time, it takes for an advanced user to perform tasks.
- **Memorability:** It refers to the user's retention of the interface over time. In other words, the ease with which users can use a design after a period of not using it.
- **Errors:** It refers to the rate of errors made by users, the kind of errors made while using the system, as well as how easy it is to recover from them.
- **Satisfaction:** This subjective measure refers to the level of the user's enjoyment while using the system.

Sometimes, when designing systems and the way they interact with users, compromises must be made as it can be difficult to excel in all the usability and design principles presented. Considering trade-offs and prioritizing some principles over others is part of the developers' job, keeping in mind the main goals of the system being designed [25, 54].

2.6.2 User Interface (UI)

A user interface is the interaction medium between the system and its user. User interfaces can include both hardware and software components.

Users interact with the interface through input controls, informational, navigational, and other components while the system communicates with the user by returning feedback to the users' decisions and actions. Therefore, the goal of User Interfaces design is to guide the user.

2.6.2.1 Usability Heuristics for User Interfaces Design

Computer research leader in user experience , Jakob Nielsen, presented in 1994 ten broad user interface design rules ("heuristics") that have since then remained relevant in the field [37, 3]:

- **Visibility of system status:** Users should be informed about what is happening through feedback so they can learn the outcome of their previous interactions and adjust their behavior;

- **Match between system and the real world:** The design of the interface should follow real-world conventions (*Mapping*), making information appear in a logical and natural way;

- **User control and freedom:** An "emergency exit" should be given to users to leave unwanted actions quickly and easily;

- **Consistency and standards:** Interface designs should follow platform and industry standards as the consistency helps users to know exactly what they should expect from a design;

- **Error prevention:** Designs should eliminate error-prone conditions or present confirmation options to users to prevent certain errors from occurring;

- **Recognition rather than recall:** The users' memory load should be minimized by making information visible or easily retrievable when needed;

- **Flexibility and efficiency of use:** Shortcuts, hidden to new users, may make the interaction more efficient for an advanced user;

- **Aesthetic and minimalist design:** Irrelevant or rarely needed information should not be displayed in interfaces as it can "clutter" the design;

- **Help users recognize, diagnose, and recover from errors:** Error messages should be expressed in plain language, clearly indicating the problem and suggesting a solution the user can follow;

- **Help and documentation:** The best systems are self-explanatory, not requiring any documentation or training. However, if users need assistance to complete a task, concise and concrete documentation should be provided.

2.6.3 Design Process

An iterative design process is composed by different stages, meant to be cyclically repeated to improve a project in order to tailor it to its target audience [3]. As presented by Shneiderman et al. [54], an iterative design process can be divided in the following stages (Figure 2.11):

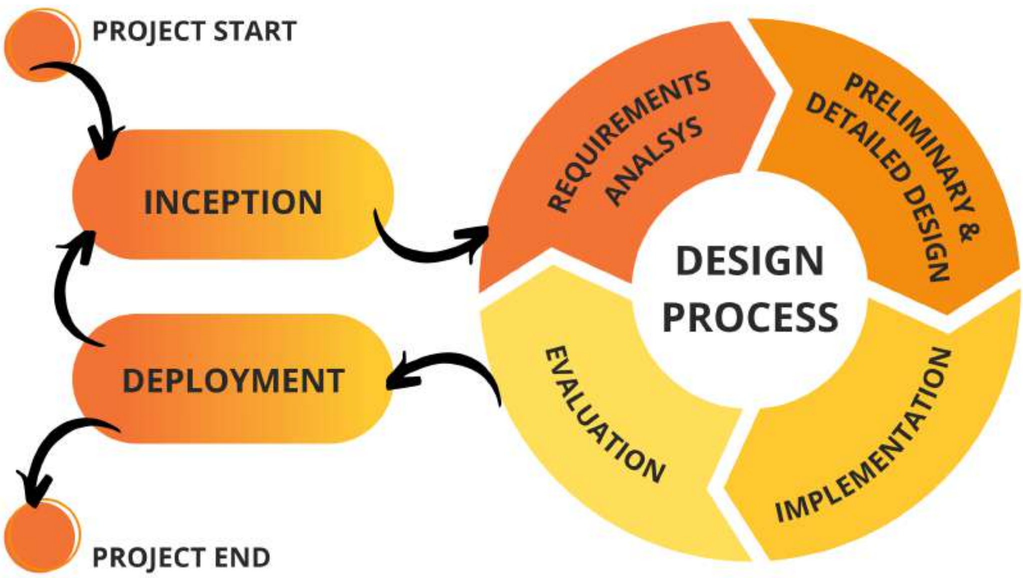


Figure 2.11: An iterative design process approach for interaction design [3].

- **Requirements Analysis:** Learn about the product we want to make and the target users. There are functional and non-functional requirements. Functional requirements define behavior that should be implemented while non-functional define requirements needed to support the application;

- **Preliminary Design:** Analyse the requirements and plan accordingly. Define the system's architecture and a plan a high-level design that helps to visualize the product;

- **Detailed Design:** Determine the specific user interaction and how the system should work specifically. It might be useful to create use cases that can be used to test and understand the desired workflow;
- **Implementation:** Turn the detailed design plans into a functioning system. The implementation of the design is progressive and gets better with each iteration of the process;
- **Evaluation:** The implementation is tested and validated. This can be done in various ways, such as user tests or expert reviews of the system. The feedback of this stage should be considered into a following iteration.

2.7 Summary

This chapter presented some of the fundamental concepts, the base of our work, with the intention of giving the reader a better understanding of the topics discussed in the following chapters.

First, we started by explaining what a soundscape is (section 2.1), since that is what we will try to recreate in the development of this dissertation. We presented some basic notions related to sound such as its physical aspects (section 2.2), how it propagates and how it can be represented and decomposed.

In the following section, we made an introduction to binaural hearing (section 2.3). This included exploring the concepts of monaural, stereophonic and binaural audio, as well as the way our brains interpret and locate sounds in our environment (section 2.3.2). Considering that we aim to create spatial audio, we focused on presenting the most relevant cues for sound localization: interaural differences and HRTFs.

In addition to that, we also presented some audio immersion techniques (section 2.4): sound panning, ambisonics and the wave field synthesis. The techniques showed different ways of using the spatialization of sound to provide a sense of immersion to the listener. However, these techniques have not been used in the context of this dissertation due to their practical constraints.

We briefly explained what augmented reality is (section 2.5), taking into consideration that the immersive audio created in the context of this dissertation will also be used in an augmented reality application.

Finally, and considering that we developed an interactive tool that is user-orientated, we presented some basic, yet relevant, concepts concerning the interaction between humans and computers (section 2.6). This included discussing User Experience, User Interfaces and the design process. The iterative design process discussed in this chapter was the basis method for the development of this dissertation. This concepts will be important when presenting Immerscape in Chapter 4 and when discussing the evaluation results in Chapter 5.

3

Related Work

This chapter presents examples of applications that implement similar and relevant characteristics to the ones we are exploring, as well as some related work and research done on topics concerning different audio immersion techniques that reflect the current state of the art in the spatial audio field. It also presents some of the previous work done on the research project that supports this dissertation.

3.1 Spatial Audio in Cultural and Educational Environments

As previously mentioned in section 2.4, spatialized audio has a wide range of applications in various areas of interest. Considering the context and objectives of this dissertation, we will present some projects that incorporate similar and relevant concepts, such as audio immersion and the application of HRTFs. We will also consider projects that make use of audio in learning or entertainment environments.

3.1.1 St. Paul's Cathedral and Paul's Churchyard Project

The gothic St. Paul's Cathedral in London was for centuries the center of religious life in London. Paul's Cross, a preaching station in the north east corner of the churchyard that surrounded the cathedral, was the site of public sermons on Sundays. Unfortunately, this space, together with the cathedral that defined it, was swept away by the Great Fire of London in 1666 [67].

The virtual Paul's Cross project[1] aims to help explore public preaching in early modern London, enabling the general public to experience a Paul's Cross sermon as an event unfolding in real time in the context of an interactive occasion. The project uses an architectural modelling software and acoustic simulation software to give access experimentally to a particular event from the past, the Paul's Cross sermon that John Donne delivered on Tuesday, November 5, 1622. These visual and audio tools allow the integration of the physical traces of pre-fire St. Paul's Cathedral with the surviving visual

[1] Available at: https://vpcp.chass.ncsu.edu

record of the cathedral and its surroundings to create a visual model. In addition to the visual model, an audio model was also created to enable users to listen to the sermon delivered, with the possibility of being positioned in different places in the crowd, and in the presence of different sizes of congregation [67].

Consequently, an important part of this historical reconstruction is the digital acoustic model that, in order to be created, required the collaboration of architects and acoustic engineers. Together, they assessed the acoustic properties of buildings and sites, so that the acoustic engineers could create an acoustic model, providing an immersive experience of spaces that do not exist anymore.

Thanks to this project, the acoustic properties of the space have been recreated, enabling people to experience a Paul's Cross sermon as a public event unfolding in real time (see Figure 3.1), with the presence of large crowds of people, and in a space filled with the ambient noises, such as the sounds of birds, horses, dogs and church bells. In order to do this, random occurring sounds from the kind of sources documented in the historic record were used to recreate the ambient noise. Besides that, the acoustic model takes other aspects into consideration, such as how people arrived at their places for the event, how the crowd gathered and settled, how the ambient noise was accommodated, and the reactions people had to the sermon.

This project is an excellent example of the use of immersive audio to recreate historic scenes. Similarly to PASEV's situation, there are no audio recordings of that time period. Due to that, it was necessary to combine different current recordings and spatialize them to be able to create the simulated audio that represents the immersive historic soundscape.

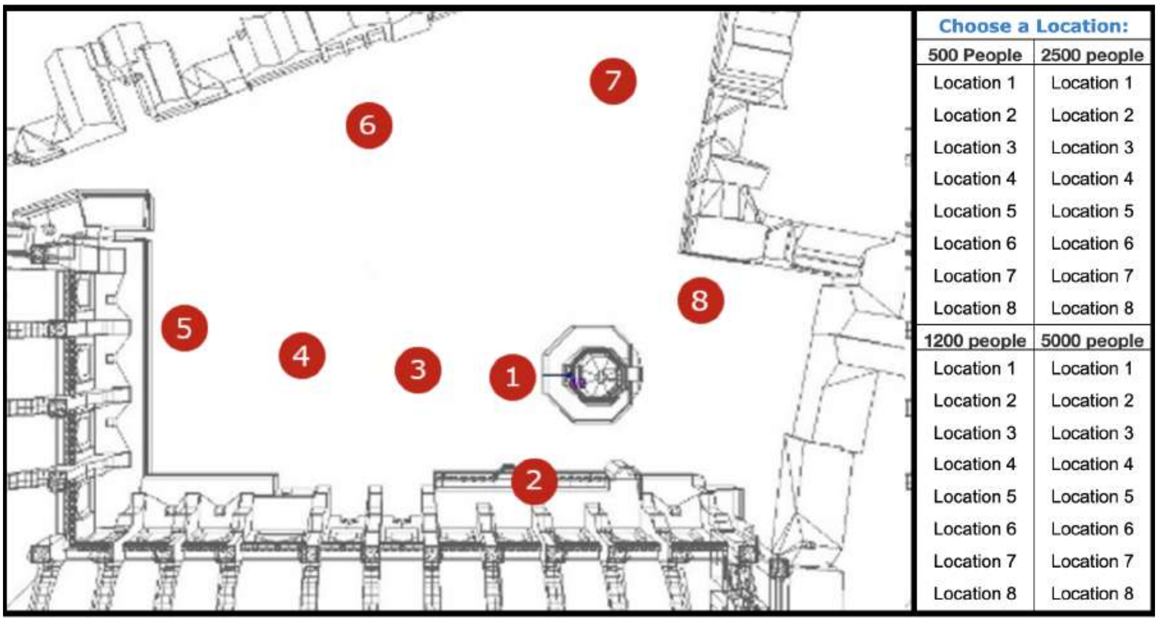


Figure 3.1: Interface of the Virtual Paul's Cross Project audio experience. The user can select the location and the size of congregation and it presents the soundscape of the event according the parameters selected [45].

3.1.2 COGNITUS Application

COGNITUS[2] is a media AI software platform that offers mechanisms to engage the public and provide high fidelity professional productions based on user-generated content (UGC). It works as a platform for crowdsourcing and enhancing video for broadcast. For this purpose, a mobile application is used to record and upload video to servers where it is processed using media AI machine learning techniques.

The platform was developed in response to an era of multimedia capture and consumption trends. The general public has access to mobile devices whose techniques have evolved to be able to capture multimedia content with reasonably good quality, while providing the mechanisms needed to share it via social media platforms. However, usually, the available UGC media has poor technical and editorial quality as it is captured on the spot by amateurs.

One of the key areas of COGNITUS is immersive sound. Using the users' recordings, and applying its audio enhancing technology, COGNITUS aims to create a realistic and immersive reproduction of the original soundscape of an event captured via the devices of multiple spectators. The audio processing is composed by a series of steps including the synchronization of multiple audio files from the same event. Then, the audio is spatialized by introducing reproduction from different angles and applying HRTFs [18].

In the context of this dissertation, we intend to achieve a similar sense of audio immersion by applying some of the same techniques. However, because we will not be using multiple recordings of the same audio, we will not need to synchronize the tracks.

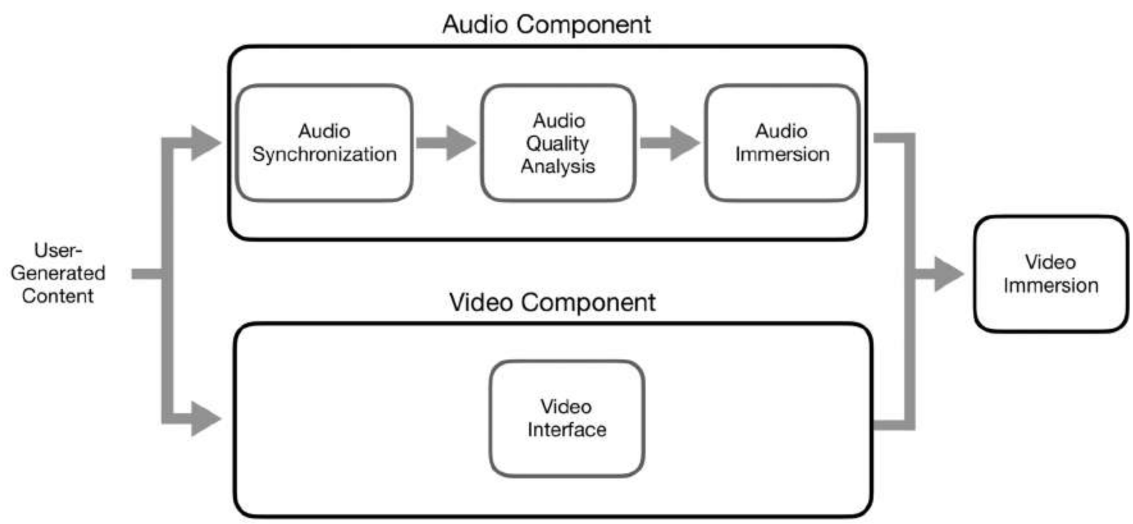


Figure 3.2: COGNITUS audio immersion technique scheme [18].

[2]COGNITUS Project: http://cognitus-h2020.eu

3.1.3 The Invisible Show Exhibition

The Invisible Show[3] exhibition, promoted by Culturgest, takes the visitor on an aural journey through the 20th century. The event provides an experience where the visitor does not *see* an exhibition but rather *hears* it. The choice of the period of time for the exhibition is also relevant, considering that the 20th century is marked by the first time humans had access to memories through audio recordings.

The exhibition takes sound as the base for all the projects presented. Whether it involves the use of the voice or it is the construction of atmospheres, it is the vehicle used for the conceptual exploration of the invisible exhibition. Integrating contributions from many artists, the event takes the visitors on a historic journey in which sound is translated into space, immersing the visitor in the auditory experiences presented.

This project shows how audio, including spatial audio, can be used in cultural contexts to provide visitors with immersive experiences that take them back in time. It also reinforces the idea that sound can be used as the main channel to deliver information to the visitor.

3.1.4 meSch Project

The meSch[4] (Material EncounterS with digital Cultural Heritage) project proposes to design, develop and deploy tools for the creation of tangible and interactive experiences that add new and relevant digital cross-media information to the one already available in the physical dimensions of museums [33].

The meSch platform enables creators and cultural heritage professionals to compose digital content to be embedded in smart objects and spaces without the need for specialized technical knowledge. This approach is grounded on principles of co-design, involving the participation of designers, developers and stakeholders in the process of creation and evaluation. To study the effectiveness of the meSch technology with the general public and heritage stakeholders, the project conducted three large-scale case studies in three different museums.

An important aspect the meSch project tried to understand is how the form of the interactive object can affect the visiting experience. In a study carried out on the historical location of the trenches of WWI, meSch determined that the tangible interaction is changed by the form of the object we interact with [43]. For this experiment, the study did not use any specific new technology. Instead, it simply changed the object the visitors hold: a book versus a belt. At the site, embedded audio recordings of stories were played automatically as a response to the visitors' actions. The visitors were able to select a different story to listen to by changing the bookmark on the book or by placing another card in the belt. As a result of this study, it was shown that the meaning behind the object held affected the visitors' expectations [43].

[3] https://www.culturgest.pt/en/whats-on/exposicao-invisivel-eng-20/

[4] meSch Project: https://www.mesch-project.eu/

The introduction of sound in the exhibitions at key moments (e.g the visitors' presence is sensed when they enter a space), and the delivery of information through sound recordings are two interactive ways that have showed to have good engagement results with visitors. Since the PASEV project also has a touristic component, these are interesting concepts to further explore in its context.

3.1.5 Demor Project

Demor[5] is a location-based 3D audio shooting game that aims to be accessible to both blind and sighted players through its use of audio as the key feature for entertainment and interaction. The project had as its main goals, not only to provide entertainment to the player, but also to be a tool that attempts to improve the inclusion of visually impaired people in the "sighted" world.

Although it is no longer available, the game presented an interesting way of using 3D sound to engage the user, by putting the player in the center of an audible shooting game that used the properties of spatial audio to convey location. The players could physically move through their surroundings based on the auditory cues they heard from the relative positions of objects around them.

In order to play this game, there was some required hardware. The players were equipped with a backpack that contained a laptop, a GPS receiver, a head tracker (used to capture the player's head movements), stereo headphones and a modified joystick. All the sensors included in the setup were used to continuously update the spatial audio feedback given to the user, in order to accurately represent the player's position in the virtual space and create an immersive auditory virtual reality context.

One of the clear downsides of this project was the amount of hardware it needed to correctly transpose the environment through sound. One of the reasons for this was the constant tracking of head movements. The Demor system tracked the movements of the player's head and continuously updated the spatial audio the player listened to according to it. In the context of this dissertation, we want to create static immersive audio files intended to be heard by different people. Due to that, we do not use any head tracking devices as we do not update the spatial audio in real time. Instead, we simulate an entire scene and record the spatial audio from start to finish.

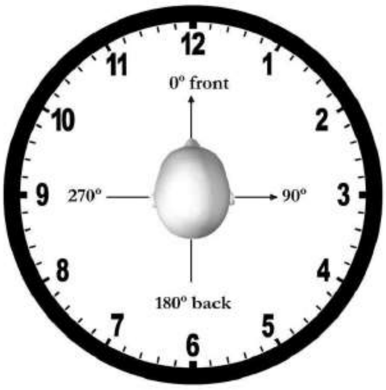
3.1.6 Navmol Application

Navmol is an application that uses spatial audio to help blind and visually impaired (BVI) students to navigate, interpret and edit the representation of molecular structures by using an analog clock reference system [1]. Once an atom in a molecule is selected, Navmol uses a speech synthesizer to communicate to the student, through headphones, information about the neighbor atoms. The sound signal is modified using HRTFs in

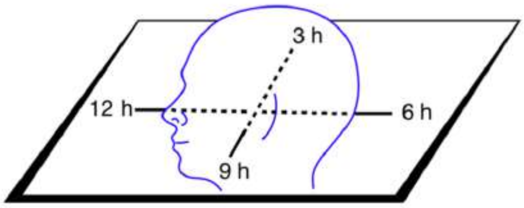
[5] More information available at: http://blendid.nl/index8803.html?id=7

order to spatialize the speech, providing the necessary cues for the listener to determine the position of the neighbor atoms based on the direction of the voice.

Considering that HRTFs play a major role on the quality of the spatialization of the sound, some user tests were conducted to choose the HRTF that best suited the user from a set of five HRTFs. In addition to that, the students had a training phase in which they tested sound originated from each one of the twelve positions present in the analogue clock system (see Figure 3.3).



(a) Clock system from above.



(b) Clock system from a side perspective.

Figure 3.3: Representation of the head (facing 0○ in the azimuth) and the direction of sound according to an analog clock used by Navmol [1].

The results of the tests showed that the spatialization of sound was sufficient to accurately perceive the planar structure of molecules, and in consequence, help BVI students in their study of molecular structures [1].

This application shows how spatial audio can be a learning tool. It presents audio as the main communication vehicle to delivery information for people who cannot see. In addition, it shows how our brains can effectively interpret location through sound and how it can be enough to help us navigate without our vision power. In the context of this dissertation, the generated spatial audio will be mostly used to complement the visual experiences the visitors have while exploring the city of Évora.

3.2 Audio Augmented Reality Projects

Other type of projects that present relevant information for the context of our dissertation are projects that are based on audio augmented reality (AAR) systems. The predominant focus on delivering information visually, in augmented reality systems, led Montan to develop a low-cost audio augmented reality prototype [35].

With that intention, Montan created a headset device that included a head movement tracker, meant to be used in a study context, as a museum interactive audio guide. The tracker in the headset registers the head's position and the system renders the audio fed to the listener accordingly. In this system, the 3D sound and the environmental acoustics are rendered in real time using HRTFs, which take into consideration the relative position and orientation of the listener. The results of Montan's tests showed that his prototype performed well for simple environments, producing quality acoustic effect and good positioning of sound using an AAR system supported by inexpensive hardware [35].

Similarly to what was done by Montan, Sundareswaran et al. [62] also implemented a wearable 3D audio system, developed for mobile security applications. The system provided alerts and communicated spatial informational cues to mobile personnel about the specific location of suspected/detected threats and waypoints for navigation to target locations. The spatial sound is processed using the information received by the system's sensors and a custom-built HRTF engine. The system mounted on a vest and helmet configuration includes wireless data communications (802.11b), voice-over-IP, GPS and a head tracker among other technology. In tests, the users were able to successfully navigate to the indicated positions of virtual entities [62].

Another context where the applications of audio in augmented reality technologies has been researched is in tourism contexts, especially in museums [34, 28]. In a recent study conducted by Fakhour et al. [28], an adaptive gesture-based audio augmented reality system was evaluated in a real museum environment, the Musée des arts et Métiers[6] in Paris, where the usefulness and enjoyment of the model were confirmed. In this study, spatialized sound was used to create the effect of a 3D space and consequently originate an immersive environment. An advantage presented for using an AAR system for tourism guidance and immersion is its capability of minimizing the degree of attention required from the users to interact with the environment, while providing them with interesting information.

In the scope of this study, Sound Augmented Reality Interface for visiting Museum (SARIM) was developed as a novel immersion and adaptation system [28]. It enables users to get personalized and position-dependent information, while they naturally explore the museum exhibitions. After conducting the experiments in the real museum environment, the positive results showed that the system is usable and achieved good levels of engagement with the users. In addition, it is also relevant to note that SARIM is being updated and extended to be used in outdoor situations such as in medieval cities, for example, the medieval city of Fez, which is considered an open-air living museum [28].

The AAR systems presented showed how spatial audio and AR can work together to provide the user with better, and overall more engaging, experiences. In the context of the PASEV project, spatial audio developed in the context of this dissertation will also be integrated in the AR solution developed for the project.

[6] https://www.arts-et-metiers.net/

3.3 Soundscapes Creation Software

As seen above, soundscapes and spatialized audio can be used in various areas of interest, but it is not always easy to create them. Some soundscape generation software include the software developed by the SoDA (Sound Design Accelerator) project [10]. Yet, this solution is targeted towards sound designers and non-specialized users may find it difficult to master. Similarly, Salo et al. proposed a soundscape platform focused on audio augmentation for museum contexts, ADAM [51]. The platform is accessible through a mobile application so that visitors can use it to create user generated soundscapes to enrich their experience. While the platform was designed for non-expert users, the soundscapes created are not immersive, which is one of the main goals of Immerscape.

In addition, there are also some professional immersive sound software alternatives. The Dolby ATMOS [13] (see Figure 3.4) and the L-ISA Studio product [30] are both examples of immersive audio software that are currently used in professional contexts. Likewise, Sound Particles [40] is a CGI-like software for Sound Design. It is an immersive audio software capable of using particle systems [15] to generate thousands of sounds in a virtual 3D world. It is mostly used in audio post-production to compose complex audio for film, videogames, TV and VR. However, the aforementioned software products are targeted towards expert users (sound engineers). This is a drawback we have set to solve with Immerscape. Immerscape aims to be an easy-to-use tool that can be used by non-expert users to create complex immersive historical soundscapes without requiring engineering knowledge, audio mixing or programming skills.

Figure 3.4: Dolby ATMOS Renderer interface for audio creation.

3.4 PASEV Project & Previous Work

As it was previously explained in the contextualization of this dissertation (section 1.2), the work developed is to be integrated in PASEV, a project focused on the patrimonialization of Évora's historic soundscapes. NOVA-LINCS (FCT NOVA) has been involved in the project as a partner research centre, contributing to build and continuously improve PASEV's digital platforms.

Currently, besides the back-office platform, there is a working version of PASEV's web page already available online (Figure 3.5). The online web platform has, as its main feature, an interactive map that allows visitors to explore multimedia content (images, audio, etc.) related to the points-of-interest of the city (identified in the figure by the colorful markers). This points-of-interest concept extends on the traditional textual descriptions of the landmarks, as it was designed to provide information to the user through visual and auditory elements.

Figure 3.5: PASEV web platform homepage [14]

When the users select a point-of-interest, by clicking on one of the colorful markers, the interface shows them an informative card about the location, similar to the one in Figure 3.6. This card is how the textual and multimedia content of a location is presented to the users. In the example shown in Figure 3.6, the audio section of the card was selected, followed by the selection of the spatial audio tab.

This results in showing all the spatial audio content related to Évora's Cathedral to the users. Therefore, this is how the immersive audio created in the context of this dissertation is presented in the web interface once it is integrated into PASEV's database.

Besides simply exploring the map through the markers, the web interface also includes an "EXPLORE" button (see top bar in Figure 3.5) that suggests thematic itineraries to the users. In addition to that, the platform also provides other filtering options that allow visitors to search locations and events according to their preferences and interests. This

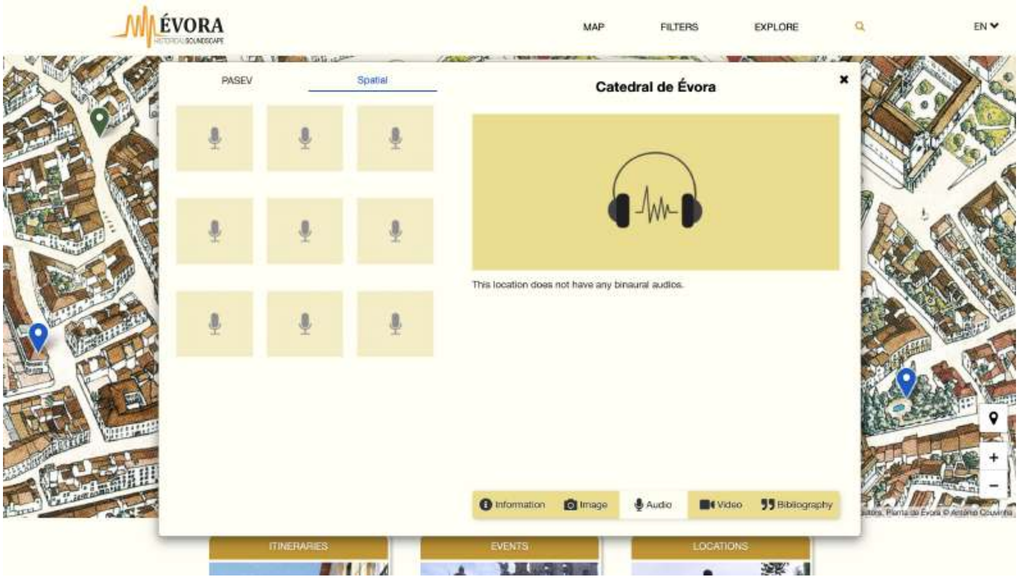


Figure 3.6: Spatial audio section in Évora's Cathedral location card [14].

includes the search mechanism (accessed by selecting the magnifying glass next to the "EXPLORE" button, shown in Figure 3.5) and the filtering interface shown by selecting the "FILTERS" option (see Figure 3.7).

Figure 3.7: PASEV landmark and timeline filtering interface shown when the button "FILTERS" is selected in the homepage [14].

As presented in Figure 3.7, the filtering mechanism includes a timeline structure that allows the visitor to filter out points-of-interest by century, as well as according to four main filtering divisions (above the timeline in Figure 3.7), which are:

- **Religious Spaces:** Represented by the blue filter, these landmarks are easily identified in the map (Figure 3.5) by the blue markers. It filters out religious spaces such as churches, convents and Évora's Cathedral.
- **Buildings:** Represented by the red filter and matching red markers in the map (Figure 3.5), these points represent buildings with a cultural and historic relevance such as theatres, museums, libraries, schools and so on.
- **Exterior Spaces:** Represented by the green filter and corresponding green markers on the map (Figure 3.5), the exterior spaces include all of Évora's outdoors points-of-interest including squares, gardens and even the famous Roman Temple, the Diana Temple.

- **Monuments:** Represented by the orange filter and the equivalent orange markers on the map (Figure 3.5), this category identifies historic monuments such as the Água de Prata aqueduct.

Some of the work previously done by other colleagues involved in the PASEV project included the database structuring process, the gathering of multimedia resources [49, 14], the domain analysis of Évora's soundscapes [32] and the integration of all the already available information in the platform. The work that resulted from those tasks served as support and contextualization for the rest of the project's development, including the recreation of the historical soundscapes generated in this dissertation.

One other contribution to the PASEV platform included a system that assists the creation of gamification solutions that include interactive narratives in the urban space. The solution was implemented to support the development of games or applications that include motivational capabilities and enhance the use of the concept of augmented reality. This AR component enables the visualisation of synthetic elements, that may involve auditory or musical pieces, allied to itineraries as drivers of interactive narratives, and in order to make the exploration of the city more appealing [66].

Additionally, another colleague developed an AR application that takes advantage of visitors' location to dynamically adapt the information the visitors are presented with during their selected itineraries, whether it is on site, or by navigating the interactive map remotely. This application allies the ubiquity of mobile devices with the need to explore historical sites in an immersive way by providing its users with 3D visual and audio content [44].

Considering that the itineraries available through the aforementioned applications are related to the historical soundscapes, the use of sound in this context is also of upmost relevance. Therefore, the spatial sound created as a result of the work developed in this dissertation, will be integrated in PASEV's database so it can be available in all its associated interfaces.

3.5 Summary

In the first section of this chapter (chapter 3), some examples of applications that have interesting features related to the solution we are proposing, or the context in which it was developed, were presented.

Firstly, we introduced the St. Paul's Cathedral and Churchyard Project (subsection 3.1.1) that uses spatial audio in a historic cultural context to allow visitors to "travel back in time", just like we intend to do in the context of this dissertation. Similarly to PASEV's situation, there are no audio recordings of that time period. Due to that, it was necessary to combine different current recordings and spatialize them to be able to create the simulated audio. However, this project only focused on producing immersive audio for one specific case. It required the involvement of sound engineers to model the acoustic environment and mix the sounds. In this dissertation, we aim to present a solution that enables non-expert users to take control, be creative and create immersive audio for a broad type of soundscapes without having to acquire any highly specific sound knowledge.

Then, we presented the COGNITUS application (subsection 3.1.2), which uses user-generated content to create immersive audio that can then be used in a virtual reality environment, to try to recreate musical events. In the context of this dissertation, we intend to achieve a similar sense of audio immersion by applying some of the same techniques. However, because we will not be using multiple recordings of the same audio, we will not need to synchronize the tracks.

Additionally, we introduced other cultural projects that make use of spatial audio: The Invisible Show Exhibition (subsection 3.1.3) and the meSch Project (subsection 3.1.4). The Invisible Show Exhibition showed how spatial audio can be used in cultural contexts to provide visitors with immersive experiences and to deliver information to the visitor, while meSch showed that having audio as a way to interact with visitors has had good engagement results. Both projects reinforce the motivation of this dissertation, showing that having a way of easily producing immersive audio can be an advantage to many audio non-experts.

Audio, especially spatialized audio, is not only used in cultural or touristic contexts like the aforementioned. Both the Demor project (subsection 3.1.5) and the Navmol application (subsection 3.1.6) showed us ways in which audio can be used as the primary source of information when users cannot see, as it is the case for BVI people. Applications like Demor and Navmol show that our brains can effectively interpret location through sound and how it can be enough to help us navigate without our vision power. However, in the context of this dissertation, we want to create static immersive audio files, intended to complement the visual experiences the visitors have while exploring the city of Évora. Due to that, we do not use any head tracking devices as we do not update the spatial audio in real time. Instead, we simulate an entire scene and record the spatial audio from start to finish.

Additionally, we presented some AAR projects (section 3.2). This type of projects were included in the review because the immersive audio created as a result of this dissertation will also be used in augmented reality settings.

Considering that this dissertation is mainly about a soundscape creation software tool, it was relevant to explore what are some of the options currently available (section 3.3). We mentioned some professional software products, some of which, although very fitting to create immersive experiences of good quality, fall on the same drawback: being targeted towards sound engineers and sound design professionals. From the soundscape software mentioned, ADAM presents itself as a mobile application that visitors can use to create user generated soundscapes to enrich their experience. While it was designed for non-expert users, the soundscapes created are not immersive, which is one of the main goals of Immerscape. In this dissertation, we aim to present a solution that is suitable for non-expert users but that allows the creation of immersive audio.

Finally, in the last section of this chapter (section 3.4), we presented the work previously done by other colleagues in the context of the PASEV project. Their work served as support for the context of this dissertation and it is what supports the integration of the immersive audio in the PASEV database and, consequently, all of its digital platforms.

4

User Generated Soundscape Tool: Immerscape

This chapter gives a detailed description of the proposed solution to the presented problem. It discusses the initial approach taken and how the proposed solution evolved from there. It gives an overview of the technology and techniques used in the development of the system, as well as it discusses the design and development process and details of the prototype's implementation.

In chapter 1, the problem we intend to solve in this dissertation is discussed. The chapter presents the motivation to further explore immersive audio as an option to create historical soundscape contexts and it introduces the proposed solution and its objectives briefly. Throughout this chapter, the answers to two of the research questions initially made will be presented:

- **How can we create immersive audio to represent a historical soundscape?**
- **Can we create a tool that helps non-expert users to create immersive audio representations of historical soundscapes?**

We have also seen, in previous chapters (chapter 2 and chapter 3), that audio processing is required to give spatial dimension to audio, whether it is played on loudspeakers or through headphones. Taking that into consideration, the audio immersion techniques used in the solution proposed in this dissertation will assume that, in order to get the full audio immersion quality, the users must wear headphones when playing the immersive audio generated.

Looking to answer the research questions, we propose a software tool, Immerscape, that aims to provide simple mechanisms that inexperienced users (people with no background in sound mixing or sound engineering) can use to easily create immersive soundscapes. The main goal of this tool is to generate immersive audio files that can be reproduced over headphones.

The following sections give a detailed description of Immerscape and its development process. They explain how Immerscape was designed and what functionalities it offers its users so they can generate immersive audio that represents soundscapes.

Firstly, an overview of the technology used for the development, specifically the audio immersion technique used by the tool (section 4.1) is given. In addition, the initial approach to the problem will be discussed, as well as how the solution evolved into a user generated soundscape tool - Immerscape (section 4.2). Lastly, the chapter gives an overview of the tool's development process and presents a discussion of all the details of the solution's prototype design process (subsection 4.3.1 and subsection 4.3.2).

4.1 Technology Overview

Immerscape was developed using the technology tools for 3D audio available in Unity[1], a cross-platform game engine developed by Unity Technologies. This engine can be used to create three-dimensional, two-dimensional, virtual reality and augmented reality games, simulations and other experiences. Unity is widely popular for video games development, but it is also used in other fields of application, such as in the film, architecture, engineering and automotive industries [63].

Considering that we wanted to simulate historic soundscapes, audio immersiveness was achieved by spatializing the sounds with HRTFs. We used Google's Resonance Audio Software Developer Kit[2] to achieve this effect. The Resonance Audio SDK is available for multiple platforms including Unity, iOS and Android. Resonance Audio is a powerful spatial audio SDK that delivers high fidelity at scale, which is critical to realistic experiences in contexts like videogames, VR, or as we intend to do in this dissertation, in the context of augmented reality experiences.

The Resonance Audio's SDK and visualization plugins work with popular tools that help to simulate how sound waves interact with human ears and their environment. The technology was built to be optimized for limited mobile computational resources, providing high quality with a positive cost-effectiveness output for all its contexts of use. In addition to that, Resonance Audio internally projects all sound sources into a global high-order Ambisonic sound field, which allows applying the head-related transfer functions (HRTF), just once to the entire sound field, rather than to all the individual sound sources within it.

The Resonance Audio SDK does not compromise on quality. It is optimized to spatialize hundreds of simultaneous 3D sound sources, allowing the developer to control the spatial resolution by adjusting the ambisonic order, always ensuring audio fidelity even in mobile contexts.

[1] https://unity.com/

[2] https://resonance-audio.github.io/resonance-audio/

At the heart of Google Resonance Audio SDK are the SADIE (Spatial Audio For Domestic Interactive Entertainment)[3] Project's binaural filters. SADIE is an EPSRC funded research program conducted at the Department of Electronics at the University of York, and it is aimed at improving immersive sound rendering in the home. In order to achieve that, the SADIE project created new algorithms that optimize audio rendering. It also provides a repository containing information and resources related to spatial audio, including their binaural filters [29, 68].

4.2 Initial Solution

Initially, to meet PASEV's needs, the creation of simulated soundscapes for some of Évora's historical events was planned. The first challenge of this process was determining which historical events PASEV wanted to recreate and gather sufficient information on them to be able to recreate the auditory environment. The King's Acclamation was the first event for which we created an immersive soundscape. The process was then expanded to three other events. Figure 4.1 shows the setup for one of those scenes, The King's Funeral.

The historians in PASEV's team provided a written document with information about the four different historic events, including the title of the event, the date, a small description of the event, a small description of the environment where it took place and the type of sounds involved in the scene. Besides the details and descriptions regarding the historical scenes, the PASEV team also provided a collection of audio clips to be used during the composition of the scenes.

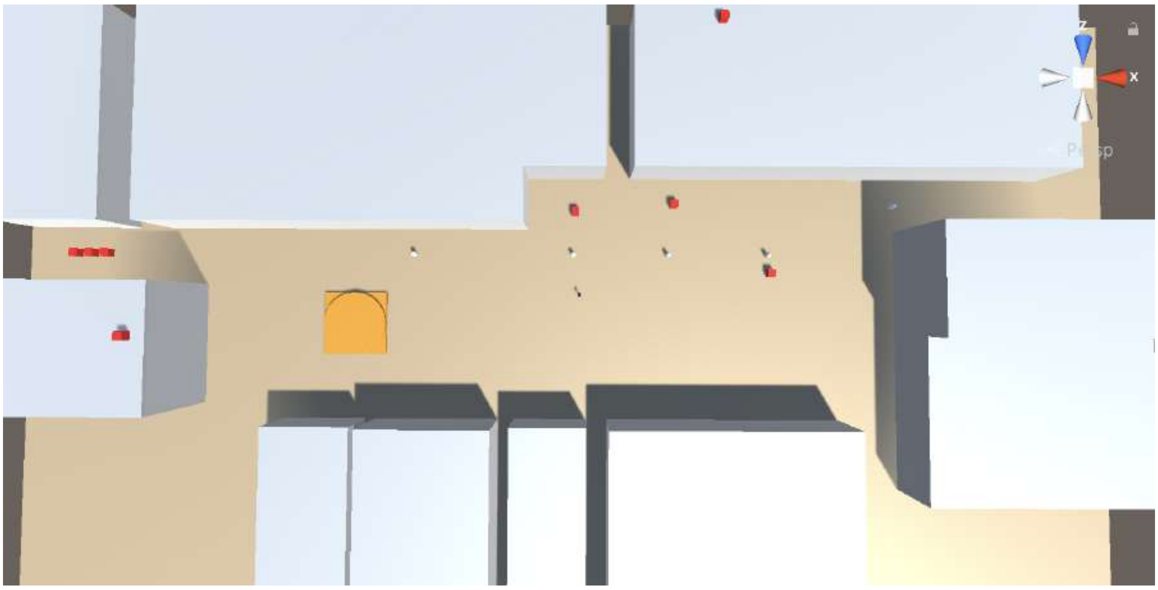


Figure 4.1: Top view of the composition of one of the historical scenes recreated: The King's Funeral Ceremony. The scene takes place at Giraldo Square (Évora).

[3]https://www.york.ac.uk/sadie-project/

For better understanding, the information provided was translated into storyboards that illustrated the scene (see Appendix A for more details). This project thus began with the prospective coding of four different historic events using the Unity 3D tools to create the immersive audio. The coding of each of the four examples involved the following different stages:

- **Requirements Analysis:** As aforementioned, we determined, for each one of the four examples, what the necessary requirements were. This process involved determining which audio clips would be necessary for each historical scenario. It required determining the type of environment and the spatial dimensions of the place where the action occurred. Lastly, the process involved determining the order of the events and the progressive flow of the auditory scene.

- **Environment Setup:** The second phase involved the creation of a spatial environment using the Unity 3D tools for each one of the scenes. It also involved determining if there were any resonance or reverberation proprieties to the environment that needed to be set up. None of the examples required any major environment resonance or reverberation setup as the scenes were all set in open outdoors environments.

- **Audio Treatment:** The quality of the audio clips provided was not ideal and so, because of that, most of the files underwent some treatment using Audacity (a free audio editor). This included the cutting of clips, normalization of the volume and combination of sounds, in some cases.

- **Composition of the Scene:** This phase involved the creation of all the different audio sources and the audio listener within the scene. In addition, it included the basic editing of audio sources and of their properties, as well as the appliance of the HRTF technique to convey audio immersion to the sound generated.

- **Creation of Mechanisms:** For each of one the scenes, and according to what was defined for the flow of the auditory scene (see Appendix A), different components (behaviors) were created, using C# scripts, to be applied to the different audio sources (such as adding movement to some sources or adding trigger events).

- **Recording of the Scene:** Once the scenes were complete using the Unity Editor, they needed to be recorded in an audio format. This was done using an imported package, Unity Recorder, which allows the developer to record the game view within the editor in various formats, including standard audio formats, such as WAV.

- **Evaluation of the Technique:** At last, the recordings were sent to two members of the PASEV team for some feedback on the immersion provided to the clips. The historical experts' comments helped improve the quality of the scenarios' implementation in a subsequent iteration. Upon making the advised changes, the team determined that the clips provided a good sense of immersion and that the technique showed promise.

During the initial development of the examples, it was obvious that there were repeated mechanisms (behaviors) being used by different audio sources, and that it could be useful to have an interactive platform where those mechanisms could be reused to create different soundscapes. This also showed to be an opportunity to allow inexperienced users to be able to create different soundscapes on their own. Immerscape was thought of as an easy medium of creation for users with no programming skills who otherwise would not be able to create these behavior components or add them to the audio sources.

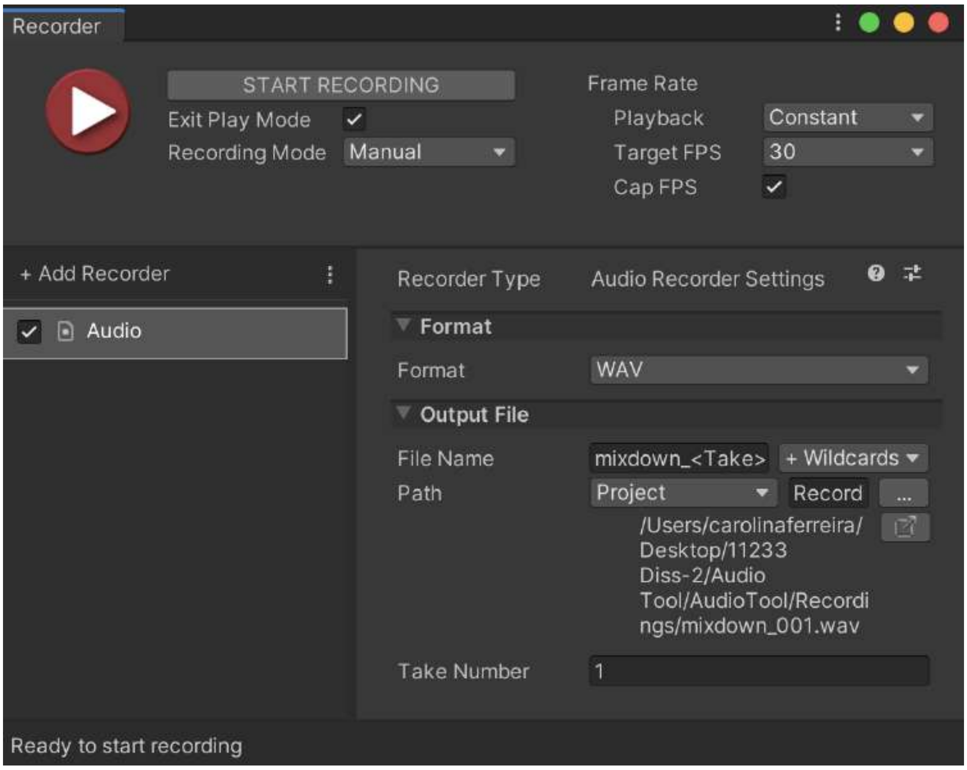


Figure 4.2: Unity Recorder Window in the Unity Editor.

In addition, one of the major inconveniences of using the Unity development environment for the examples recreation was the audio recording process. Although using the Unity Recorder (see Figure 4.2) allowed to record the scene's spatial audio in the editor mode, it was not possible to listen to the audio while it was being recorded. In order to make the recordings, the scenes had to be played in the Editor mode and timed so we knew when to stop the Unity Recorder. As so, it became one of the main goals of the aforementioned interactive platform to provide an easy-to-record mechanism which allowed the user to listen to the immersive audio as it is being recorded.

To sum up, Immerscape was developed as an auxiliary tool to generate immersive audio clips like the ones that represent the historical soundscapes the PASEV project wanted to offer the visitors of the city of Évora. However, it became our mission, during the development, to design the system so that its capacity was not limited to the specific context of the project.

4.3 Immerscape

Immerscape is a tool to generate audio immersive clips. From the definition of the environment to the recording of the soundscape created, Immerscape aims to allow its users to have autonomy while creating the audio clips themselves. While the tool was developed in the specific context of the PASEV project, which addresses Évora's historical soundscapes, it can be used to generate other soundscapes.

The tool was developed to be used as a back-office tool for organizations, institutions or individuals who wish to compose immersive auditory scenes and generate the corresponding standard format audio files from them. The process of creation of the immersive audio using Immerscape should be supported by the use of headphones. Headphones have two channels that directly input one into each ear. Consequently, the signal that reaches the ears is not influenced by the listener's surrounding environment [69]. The use of headphones is also relevant when audio immersion is conveyed using the HRTF technique (see subsubsection 2.3.2.2 for more information).

The tool can convert sounds obtained from specifically selected recordings, into immersive 3D audio files. This process includes the composition and spatialization of sound in order to generate audio files with the intended sense of user immersion in the historical scene, while only using the visitors' devices and headphones to reproduce them.

Immerscape is designed for users with no programming skills such as historians and other non-engineers for whom it may be hard to deal with more technical audio development applications and integrated development environments (IDE). Using an interactive interface, Immerscape provides simple mechanisms to build immersive scenes and record the audio from within those scenes, without having to worry about any of the technical aspects of 3D sound generation.

To achieve better results, the sounds samples selected by the user, must undergo some treatment beforehand as preparation for the composition of the scenes. The previously described audio immersion sense is then achieved by the strategic position and combination of sounds, and the application of the HRTF binaural filters provided through Immerscape.

Figure 4.3 illustrates a description of a historical soundscape that can be generated with Immerscape. This soundscape intends to recreate a historic event that happened in the city of Évora in 1656: the acclamation of the new king, Afonso VI. The orange bars in the figure represent the audio tracks of this soundscape. This soundscape is composed of several tracks that start and stop playing at different times, and that are spatialized to simulate the sound sources' movement and different positions within the environment.

The user can associate sounds to the audio objects and define when those sounds start and stop playing. The position of the audio objects in the scene defines their relative

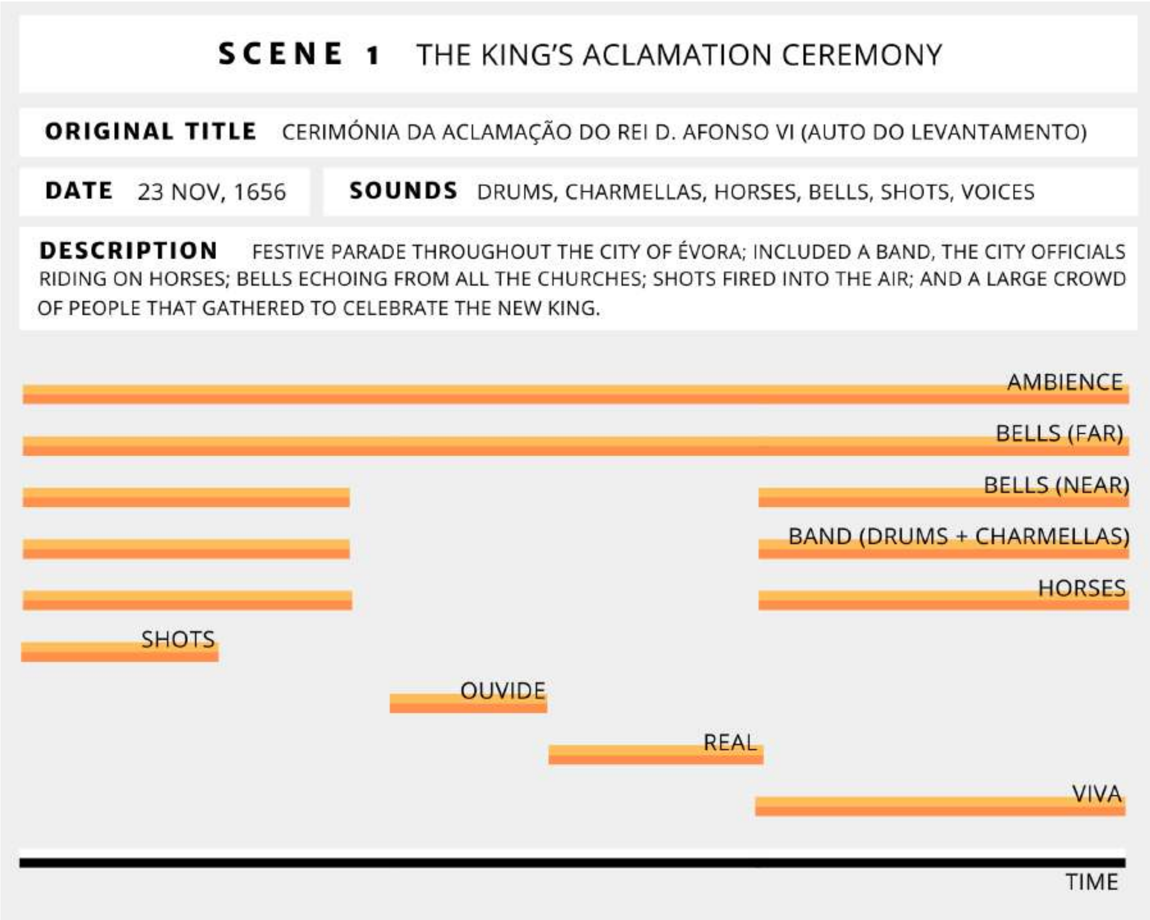


Figure 4.3: Acclamation Timeline.

position to the listener. When the listener hears those sounds with headphones, he or she will hear the sounds as coming from those directions.

In addition, the audio objects can move in the scene. As a result of this movement, the listeners hear the sounds moving in the 3D space around them, adding to the immersion feeling provided.

4.3.1 Analysis and System Design

Immerscape is the proposed solution to the main research questions made in section 1.3. The following subsections will contribute to elaborate on the iterative design process (subsection 2.6.3) followed during development.

It was important to complete at least one full iteration during the period of the dissertation; that is, going over each one of the different five design stages at least once. However, it is understood that in order to create a better software product, the process should be repeated a few more times until a solid result is achieved. The design process included the Requirements Analysis phase (subsubsection 4.3.1.1), the Preliminary Design phase, the Detailed Design phase (subsubsection 4.3.1.2 and subsubsection 4.3.1.3), the Implementation phase (subsection 4.3.2) and lastly, the Evaluation phase (chapter 5). In this subsection, the first three mentioned stages of the design process are presented.

4.3.1.1 Requirements Analysis

During the Requirements Analysis phase, we must learn about the product and its target users. The capture of requirements is an important part of all software engineering methodologies [3] but often, this activity has special focus on the functional requirements of the system - what the system must be able to do. These requirements usually focus on the needs of all stakeholders (anyone having any type of relation/interest in the project).

Since Immerscape evolved from the initial solution presented previously (section 4.2), the stakeholders did not express any specific requirements. The stakeholders just knew they wanted immersive audio files representative of soundscapes, but they had no requirements on how to obtain them.

Immerscape looks to answer two of the main research questions presented in section 1.3 and as so, the system's requirements were defined as an answer to "**how we can help non-expert users to create immersive audio representations of historical soundscapes?**".

From this broad question, several requirements were identified. Firstly, we determined that the users would have access to Immerscape through a desktop application. It was also determined that the system should provide users with a simulation environment where sounds can be added, edited, composed and recorded to create immersive scenes. This must be achieved through mechanisms, created and added to the user's projects, without them having any expertise in audio engineering. The system must provide an interface to interact with the sounds being composed and give users simple ways to do so.

To sum up, the system must implement the set of requirements that stem from the generalization of the mechanisms needed to create the historical soundscapes of Évora. The core requirements of Immerscape, which enable it to produce a wide variety of soundscape scenarios, are the following:

- To be able to select a three dimensional sound environment;
- To have an immersive presence within the scene's environment;
- To be able to navigate the environment;
- To create audio sources;
- To edit audio properties of each audio source;
- To provide movement to each audio source;
- To provide a mechanism of triggering sound events;

- To play and record the soundscape associated with the scene composed.

From the core requirements presented above more specific features were derived that represent a broader view of what the tool can do. One of the main goal of the evaluation process (presented in chapter 5) is to receive the end-users feedback on the current features and determine if any major redesigning needs to be applied to the current implementation. This allows us to present a solution to the end-users which stakeholders had not previously considered. This enables users to test the solution and give them an opportunity to participate in the design process, once they have a better understanding of the system. Ideally, the requirements of the system will be refined with each iteration as presented in subsection 2.6.3.

Other desirable characteristic of Immerscape is its modularity, room to scale and its general audio creation purpose. The system's requirements stem from the generalization of the mechanisms needed to create historical soundscapes of Évora, which allows the tool to go beyond the scope of the project.

4.3.1.2 Feature Design

Taking into consideration the requirements that were defined, it was necessary to plan Immerscape's features design accordingly. That process included planning a high-level design of Immerscape. From the list of core requirements, we derived some other features that complete the functionality of the application and ensure the completeness of the workflow. These features are introduced below, and they are described in detail in subsection 4.3.2:

New Project Creation The user must be able to create a new project. A project environment is where the user can built the acoustic scenes. In order to create a new project, the user must select one of Immerscape's pre-defined scene environments.

Environment Navigation The user can navigate the project environment with the player. The player can move around the 3D simulated environment as if the user was inside the acoustic scene. The player must be able to move back and forward, rotate its position and look up and down.

Camera Switch When inside the project environment, the user is given, by default, the player view. However, the user can switch to an alternative view, the up-view mode, where the acoustic scene can be seen from above, with a wider perspective.

Create Audio Object The user must be able to easily create an audio object within the project environment and position it on any horizontal surface of the scene. The audio object represents a sound source. Each audio object will only have one audio clip associated.

Create Point The user must be able to easily create a mark point within the project environment and position it on any horizontal surface of the scene. The point represents a mark for the audio object's movement or for triggering events. Multiple movement routes or multiple trigger events can be defined using the same Point.

Edit Audio Object The user must be able to edit the audio object's proprieties, including if the audio clip plays on start, if it is muted, if it plays on loop, its volume, its altitude/height on the scene, whether it has an associated tag, the radius of audio reach and the audio gain.

Add Movement to Audio Object The user must be able to add movement to each audio object. While some audio object may be static and remain in the same position throughout an acoustic scene, other audio sources might change positions. Due to that, the user must be able to describe movements for audio objects by determining their route and the speed at which they move.

Add Trigger to Audio Object The user must be able to add sound events that are triggered at some point after the scene starts. While some audio objects may play at the scene's start, others might only play later in the acoustic scene. Due to that, the user must be able to determine what trigger event is activated, when the trigger is activated and what Point activates it.

Move Audio Object The user must be able to adjust the audio object's position within the scene. As the player, immersed in the scene, the user must be able to rearrange the different audio objects, within the scene, by picking them up and moving them to another location.

Play the Scene's Audio The user must be able to play the acoustic scene, at any given moment, to listen to the soundscape that is being generated. This functions as a testing mechanism during the construction of the scene.

Record the Scene's Audio The user must be able to record the acoustic scene. When the user wants to record a clip of the current project's soundscape, the user must give the clip a name. The recording of the scene must play the audio in the player mode, while it records it from the player's perspective. The user must be able to listen to the reproduction of the scene while it is being recorded and stop the recording at any moment.

Save Project The user must be able to save the progress made on any project. In order to save a project, the user must give it a name. The users must be able to save the project at any moment during the construction of the scene and they must be asked if they intend to save the project, whenever they try to leave the project environment.

Reopen/Edit Saved Project The user must be able to find any saved project within the tool. The projects must be listed by update date, and when clicked, the project selected must be reopened. The user must then be able to continue to complete the project's data or to edit any aspect of the current version of the project.

Listen to Recorded Audio The user must be able to find any of the project's soundscape recordings made within the tool. The recordings must be ordered by date of creation. When selected, the recorded clip must play; and when selected again, while playing, the clip must stop.

Help At any point, while in the project environment, the user must have access to a help panel with quick instructions, to be able to access the tool's main functionalities.

From this design stage resulted some low-fidelity mock-ups (available in Appendix B) that illustrate the initial plans for the tool's interface. During the first iteration of the development process, the interface was not the main focus. Therefore, Immerscape was designed to be simple and functional. The plan was to provide end-users with a reasonably good interface to test the primordial prototype of the application. The user feedback, given during the Evaluation phase, helped determine the interaction and visual improvements that needed to be worked on in the following iterations.

4.3.1.3 System Architecture

Once the set of core features were determined, the system's architecture was defined. According to this structure, we determined the specific user interactions and how the system should work specifically.

The system was organized as shown in Figure 4.4. Immerscape's system organization was designed to take full advantage of the Unity Editor properties and its built-in assets: Scenes, Game Objects, Prefabs, Audio Components, its UI-tools, among other elements. In this subsection, we will discuss the different parts that compose the system and explain briefly how they interact with each other. For further details on how they work together to meet the system's requirements, see Implementation (subsection 4.3.2).

Pre-Defined Environments

Within the editor, scenes are where other assets are contained. The progression between scenes allows for a multiple screen application. The scenes represent the application's environment and, in Immerscape's case, with the help of the Google Resonance SDK, scenes were also used to define the various pre-defined 3D audio environments which can be used to build soundscapes on. The pre-defined environments can be selected by the users when they create a new project and environments represent the physical and spatial properties of the place where the acoustic scene takes place.

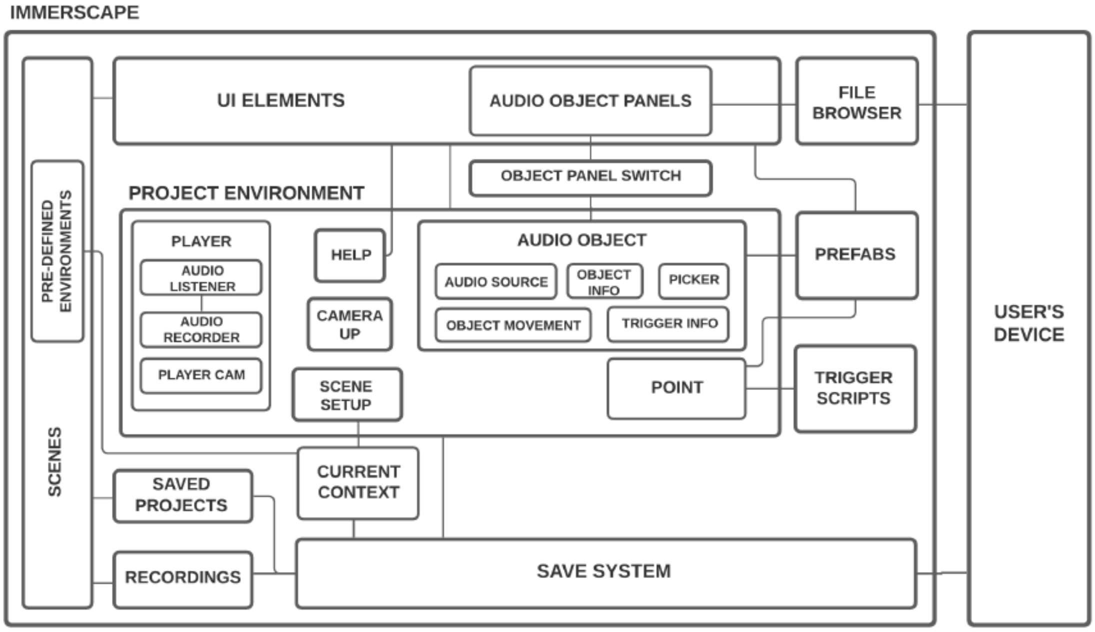


Figure 4.4: Immerscape Organization.

For example, Figure 4.5 shows the the Audio Room scene hierarchy (on the left), the disposition of the elements shown in the Scene view of the editor (middle), and the Inspector (on the right) that shows the components associated to the element selected in the editor, in this case, the ResonanceAudioRoom.

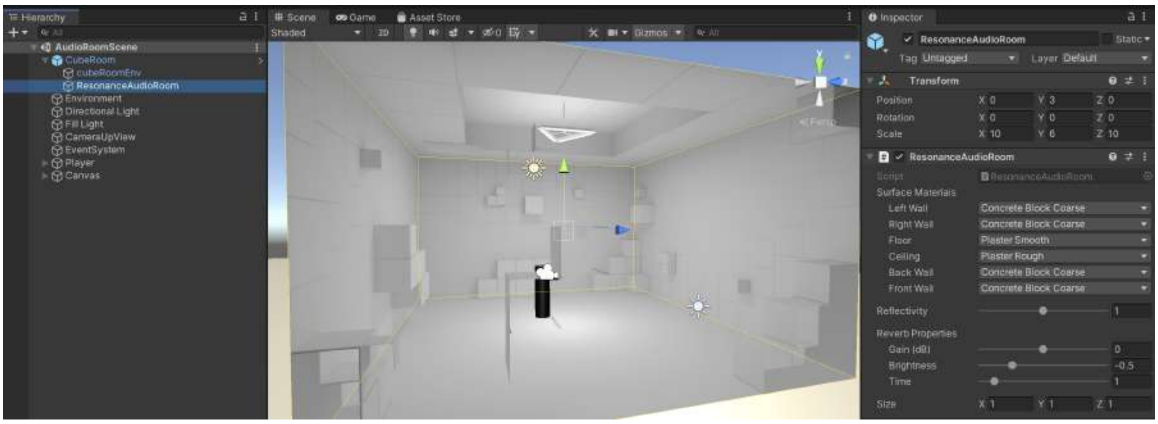


Figure 4.5: Example of one of the pre-defined environments (Unity Editor view).

Prefabs

Unity has a Prefab system that allows developers to create, configure, and store a GameObject, complete with all its components, properties and child GameObjects, as a reusable asset. Immerscape takes advantage of the Prefabs system to create the Audio

Object, the Player, the Point and some UI elements frequently used in scenes. Examples of UI elements that are a Prefab object include the confirmation windows and the item elements (see Figure 4.6) for the saved projects or the recordings listings. Prefabs can be instantiated and initialized through scripts. This allows for the dynamic evolution and update of the interactive interface.

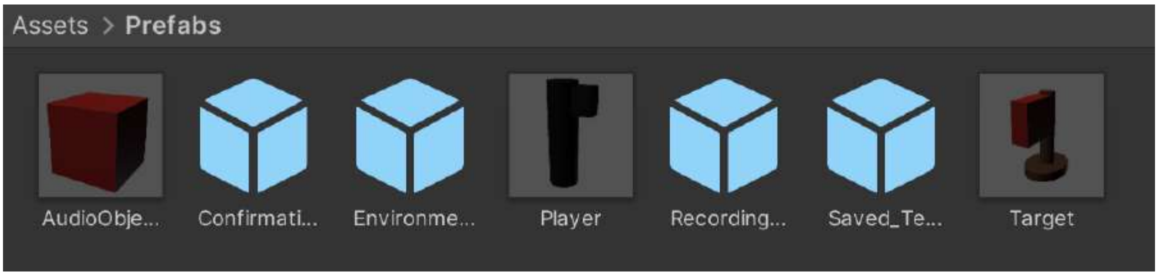


Figure 4.6: Example of prefabs used in Immerscape (Unity Editor view).

Player

The Player is the user's eyes and ears in the environment. It can move around the environment, look around it, as well as switch between the Player's view camera and an up-view camera. The up-view mode can be useful to determine the position of the elements in the scene.

The Player is responsible for composing the scene and record it, as they hear it with its immersive spatial awareness (obtained by the application of the Google Resonance binaural filters). This is possible because the Player has, among others, an Audio Listener component, which picks up the sounds emitted by the Audio Objects' audio sources. The engine can then simulate the effects of a source's distance and position to the listener and play it to the user accordingly.

When assessing how to make the initial requirements work, we determined that, inside the project environment, the Player needed to be able to create two types of elements: an Audio Object and a mark Point.

Audio Object

The audio source component in the Audio Object enables sound reproduction. Other scripting objects hold specific information about the Audio Object's movement, its triggers and other editable properties. This information is displayed on the Audio Object's panels when they are selected, and so that the Audio Object's state can persist when the project is saved and then loaded again. The Audio Object must be edited when it is created. Otherwise, it is just a non-influential physical object inside the scene. When editing it, the Player can upload an audio clip to each Audio Object and adjust parameters such as volume, altitude from the ground, when to play, among others.

Point

The Point element is a physical object which may be instantiated by the user, within the project's scene, but that, initially, has no behavior attached, serving only as a reference point. This object can provide support to define the Audio Objects' movement inside the scene and/or to define trigger events.

Audio Object Panels

Each Audio Object has three different panels associated: an Edit Panel, a Movement Panel and a Trigger Panel. These panels are the main interaction medium between the scene elements and the user (represented by the Player).

- **Edit Panel:** Each Audio Object's core proprieties, such as the volume and the clip it plays, can be edited through the Edit Panel.
- **Movement Panel:** Each Audio Object can define a route of movement, marked by spatially located Points. This panel allows the user to define the route and the speed at which the Audio Object should move.
- **Trigger Panel:** Each Audio Object can define multiple triggers using multiple different mark Points. A trigger can be defined by selecting which specific mark Point activates it, and by picking one of the trigger events available. When a trigger is defined for the Audio Object, the corresponding Trigger Script is added to the Point where it should be triggered.

Audio Recorder

Through the Player, the user can use an Audio Recorder component which records the audio that goes through the Audio Listener component. During the initial solution's development, one of the inconveniences of using the Unity Editor for the creation of the PASEV project's soundscape examples was the recording process. Although there is a Unity Recorder plugin that allows the recording of in-project sounds within the editor, it was not possible to listen to the audio, while it was being recorded. This Audio Recorder component allows the user to do just that.

Camera Up

While the user is inside the project environment, the primary camera activated is the Player Camera that represents the immersive view of the scene. However, Camera Up provides an alternative view. The user can switch the view to an up-view where the acoustic scene can be seen from above. This was included because it can be particularly useful, when the user is composing the scene, to have a wider view of the environment and to determine the position of an element relative to others.

File Browser

The File Browser is mainly a communication layer between the project environment and the user's device. The Unity Standalone File Browser (see Annex I for more information) was imported into the project. It is a simple wrapper for native file dialogues on Windows, Mac and Linux. It allows the user to upload the audio clips into the Audio Object through its Edit Panel.

Object Panel Switch

The UI Panel objects have to be attached to a Canvas object in Unity. The Canvas object is the main UI element that contains all the other UI elements, such as Panels, buttons, text elements, among others. When designing the system, we needed a component to active and deactivate the three different panels each Audio Object has. The Object Panel Switch reacts to the different key commands and open and closes the Audio Object Panels accordingly, depending on the Audio Object selected.

Trigger Scripts

The Trigger Scripts are the collection of scripting components developed to represent the different mechanisms that can be used to trigger audio events within the scene. They describe generic actions that can be associated to Audio Object and specific mark Points. As deemed necessarily, more scripting components, representative of other events can be developed, added to the collection of Trigger Scripts and made available to the end-users by integrating them into Immerscape's interface.

Help

Although Immerscape is fairly simple and all functionalities are mostly accessible directly to the user in the main screen of the project environment, a help panel was included. When activated, this panel shows the user the basic commands and shortcuts to perform all of Immerscape's functions. It includes a link to a tutorial slider that goes over the steps of each main functionality.

Current Context

The current context component keeps Immerscape's context. It is especially relevant to be able to save and load projects. It's an important element of the custom Save System developed for Immerscape. It keeps general information about the current scene and if in the project environment, other relevant information such as the numbers of audio elements and the date of update.

Scene Setup

The Scene Setup is the responsible for the interpretation of the current context and if needed, responsible for the loading of a saved project when it is reopened.

Save System

At any point during the scene's construction, the user can save the project simply by giving it a name. The save system is responsible for saving the project's context and all the elements added to the environment in local binary-encoded files. The Save System is also responsible for creating and managing the Immerscape folder (created in the Documents folder of the device). Each project is given its own folder where all related files are aggregated.

Figure 4.7 shows the file directory organization for Immerscape's persistent storage. The first level of the Immerscape folder is divided into two sub-folders: the "saves" folders, where the data resulting from the projects created is saved, and an "audio-samples" folder where the user can storage the clips that they intend to use for the projects. The second level (through the "saves" folder) is organized in recordings (the audios generated), projects and projects-data (that store all the information related to the saved projects).

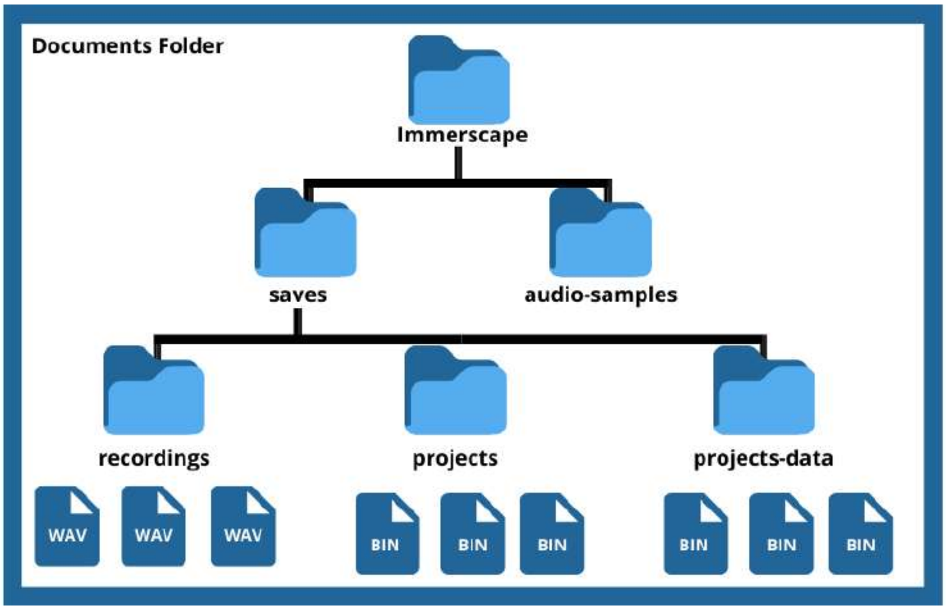


Figure 4.7: The file directory structure of Immerscape's custom Save System.

Saved Projects

Once it is saved, the project can be found in the list of all saved projects in a screen accessed by Immerscape's main menu. In the Saved Projects screen, the user can be reopen any project and edit it again.

Recordings

Similarly to the Saved Projects, all recordings created inside the projects' environments can be found within the application. The list of all recordings can be found in a screen accessed by Immerscape's main menu. In the Recordings screen, the user can play and listen to all the recorded audio clips.

4.3.2 Implementation

Pre-defined Scenes are populated by GameObjects, the most important concept in the Unity Editor. Every object within the scene, whether it is the Player, the Audio Objects, or even the UI elements are GameObjects and act as containers for functional Components that represent their behavior.

To support Immescape's requirements, GameObjects were created with different combinations of components. Examples of these components are, the Unity built-in component types such a Transform component (to position, rotate and scale the object in the environment), or imported components such as the 3D audio components, provided by the Google Resonance SDK.

However, Unity also allows developers to create their own components using the Unity Scripting API. The C# scripts created can then be attached to all different kinds of GameObjects to provide them with some type of specific behavior.

The aforementioned concepts were the basis of Immerscape's development using Unity. In this subsection, it will be discussed how the main functionalities were implemented.

New Project Creation

Using Immerscape, it is possible to maintain multiple under-development projects. When a user creates a new project, they are asked to select one of the pre-defined working environments (e.g., the user might want to work in our default empty outdoors environment or in our indoors resonance-prone audio room - see Figure 4.8). Currently, Immerscape's prototype offers six different pre-defined environments to choose from:

- **Default Scene:** The default scene is an outdoors environment with a simple plane surface and no specific audio properties.

- **Audio Room Scene:** The audio room scene is an indoors environment. It is an empty room where resonance properties are applied to the sounds played. This environment and the resonance components used to compose it were provided by the Google Resonance SDK.

- **Cathedral Scene:** The cathedral scene is an indoors environment. It is a model of a cathedral where resonance and reverb properties are applied to the sounds played. This environment and the resonance components used to compose it were provided by the Google Resonance SDK.

- **Cave Scene:** The cave scene is an indoors environment. It is a model of a cave where resonance and reverb properties are applied to the sounds played. This environment and the resonance components used to compose it were provided by the Google Resonance SDK.

- **Giraldo Square Scene:** The Giraldo square scene is an outdoors environment. It is a simple 3D model of the Giraldo Square in Évora. It was mainly developed to have a sense of space when building the historical scenes in the context of the project. No resonance or reverb proprieties were applied to the environment.

- **Park Scene:** The Park scene is an outdoors environment. It is a simple 3D model of a public park in Évora. It was developed to gain a sense of the surrounding space when building the bandstand inauguration scene in the context of the project. No resonance or reverb proprieties were applied to the environment.

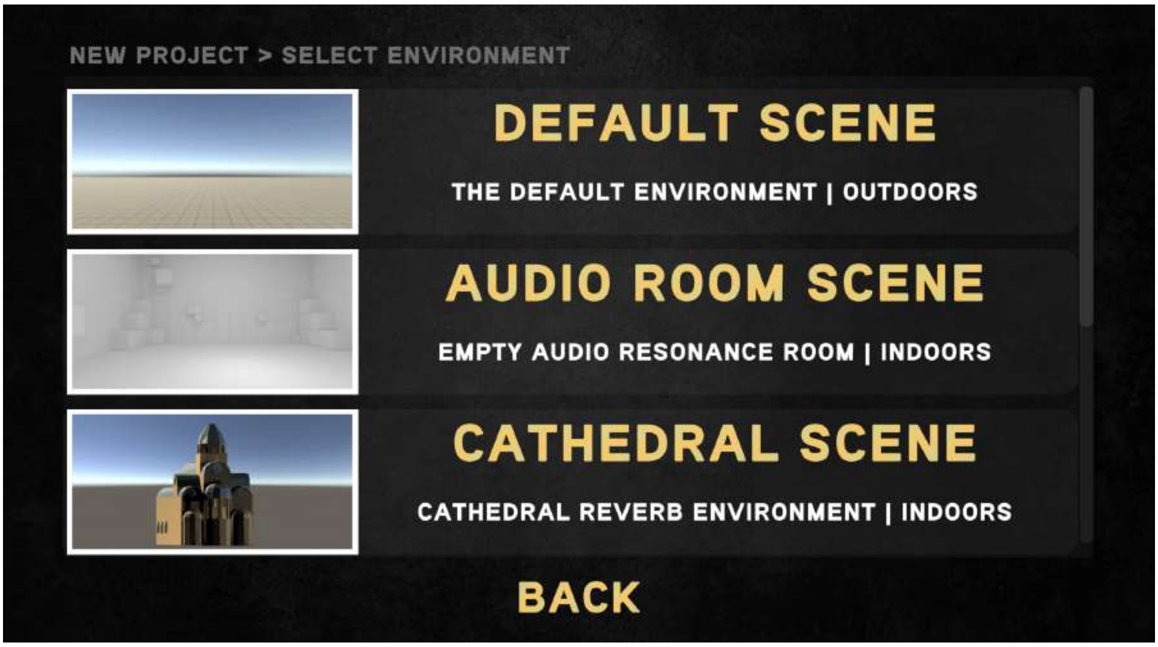


Figure 4.8: The Immerscape screen where the user can choose a scene to work on. The tool includes template scenes and some predefined scenes from different sites in Évora.

The environments that were not imported from the Google Resonance SDK, and modified based on the provided scenes by the plugin, were built using the packages Unity ProGrids and Unity ProBuilder. Some of the pre-defined environments could use some aesthetic improvements to look more realistic.

However, 3D modelling in Unity was not in the scope of the work developed in this dissertation. More complex 3D models should be designed by professionals in the area, and once imported to the Unity project, they could be easily be made available to the Immerscape users.

Once a pre-defined environment is selected, the project environment is open and the current context is set to the ongoing project. The same interface tools and same player capacities are provided to all environments.

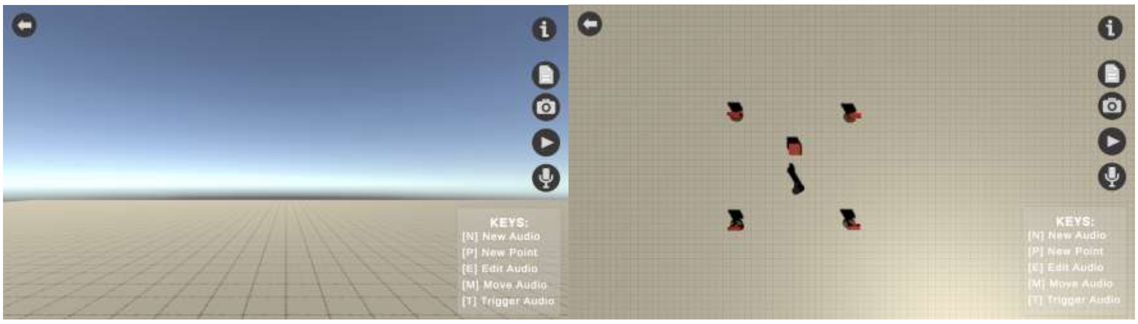
Environment Navigation

Once an environment is selected, the user is immersed into it in character form, which represents the user within the scene: the Player. The Player simulates the user's position in the world, as if they were there. The Player is able to see and hear what is around them and take part in it. In order to do that, the Player was given a CharacterController component, moving capacities, among others. Using the computer keys, the Player is able to move back and forward and rotate its body to change direction.

Additionally, and using the mouse wheel, the Player is able to change the camera angle to look up and down. These functionalities were added so the Player can move within the scene and position themselves in relation to the Audio Objects differently during the audio reproduction and recording process. The Player movement capacities were important to have since the Player represents the user in the environment and most people can freely move in real environments.

Camera Switch

While building scenes in Unity, it is possible to add as many different cameras as necessary. During the first design phase of the tool's prototype, it was envisioned that the tool would use two different cameras (Figure 4.9): the Player Camera, representing the Player's eyes, and an Up-view Camera, that looks at the scene from above.



(a) The Player view perspective.

(b) The Up-view perspective.

Figure 4.9: The different perspectives the user can experience with Immerscape

Immerscape aims to be simple to use and adding a second camera, the Up-view camera, to all environments was done with the intention of making the creation of the Audio Object and Point easier. We considered that the wider "lens" used allows Immerscape's users to have a better idea of the relative positions of the elements in the scene, and it provided a better view to adjust their positions in the scene. Since the Up-view Camera cannot capture the entire environment from the height it is set at, the camera "follows" the Player as it moves around the environment.

Only one camera can be active at once, so when one is activated, the other must be automatically deactivated. That is why the mechanism is implemented as a "switch". The mechanism can currently be activated by an UI button and a key shortcut, key C.

Only the Player Camera has an Audio Listener component (that represents the Player's ears), so when the Recorder is activated in the Up-view mode, the camera switch is performed automatically. To sum up, the Up-view Camera is meant to be used while the Player is in "editing mode" and not during the recording process.

Create Audio Object

A Prefab object of the Audio Object with all its base components was created. When in the project environment, the user can point with the cursor to a surface and with a simple press the N key to create a new Audio Object in that exact position (see Figure 4.10). Each Audio Object has an associated numeric identifier, defined upon creation.

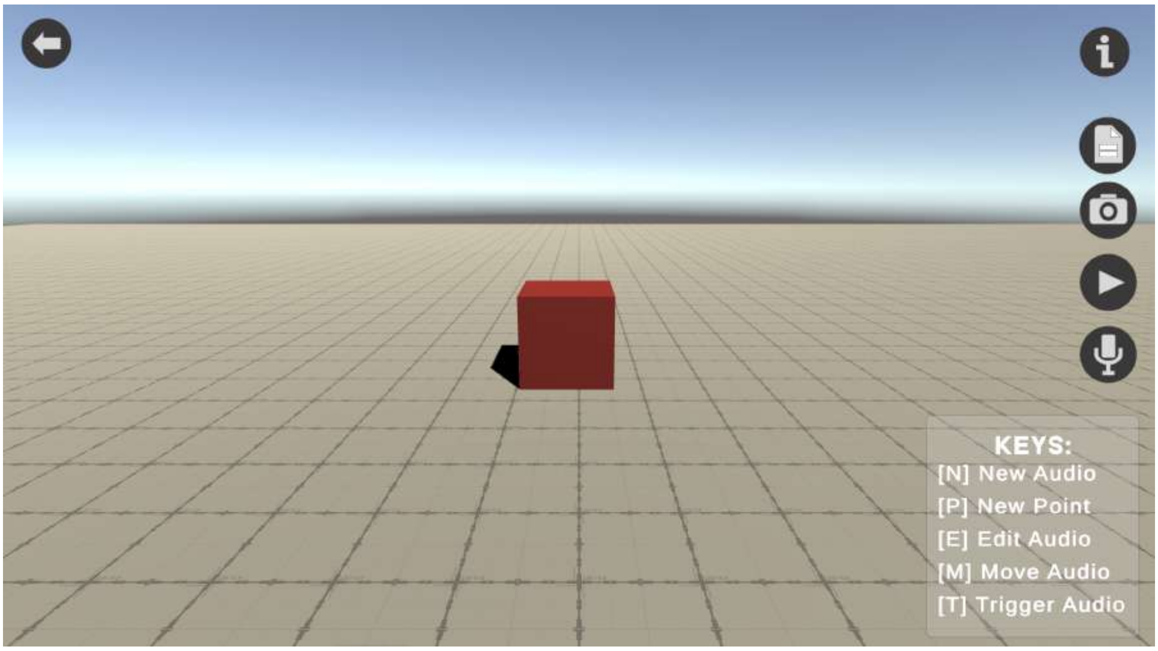


Figure 4.10: The Audio Object, represented by the red cube, when it is created in the environment.

The Audio Object is created with a few components that are common to all Audio Objects, such as the Audio Source, a Rigidbody, a BoxCollider, the Picker, the Object Info component, the Object Movement component and the Trigger Info component. The last three are components that keep the state of the panels for each object. These components are necessary so the information is persistent when switching between the panels of different objects. Additionally, those same components are used to save the properties of each Audio Object when the project is saved. They are also what allows Immerscape to later reload the saved objects.

Create Point

A Prefab object of the Point was created with a BoxCollider component attached that represents a surface around the point that "detects" collisions. The property that indicates that the element is a trigger is selected, indicating that the Point can act like a trigger object. This is the basis to some of the trigger events.

When in the project environment, the user can point with the cursor to a surface and by simply pressing key P, the user creates a Point in that exact position (see Figure 4.11).

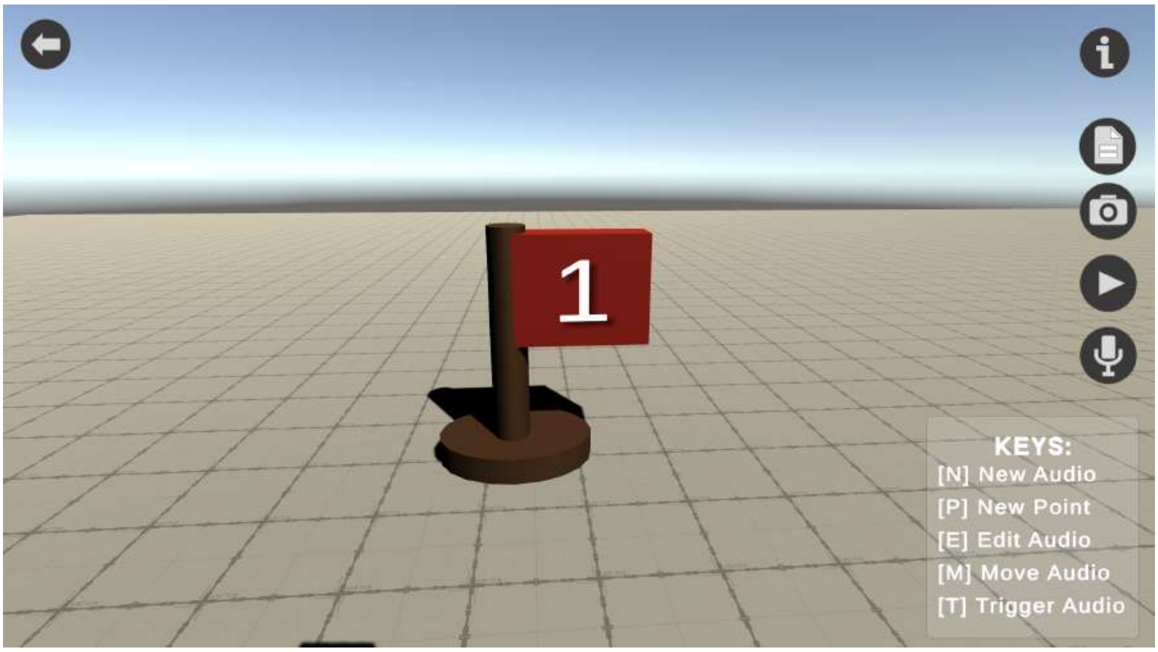


Figure 4.11: The first Point (Point 1) created in the project environment.

When created, the Point does not have any sort of behavior or properties associated, with the exception of a numeric identifier that is defined upon creation. The identifier is used to refer to the Point in the Audio Object panels. All the behavior associated with a Point must be added after its creation by the user, through Immerscape's interface.

Edit Audio Object

To edit an Audio Object, the user must point the cursor to it and simple press the key E. This will open the object's Edit Panel that displays all the information concerning the Audio Object's core properties. The state of the parameters that are adjustable through the Edit Panel is kept in the Audio Object's Object Info component. In this panel (Figure 4.12), the user can edit the following parameters for each Audio Object:

- **The Audio Clip:** The user can upload the audio clip that should be played by the Audio Object. This clip is immediately associated to the Audio Source component.
- **Mute:** The user can define if the Audio Object is muted.
- **Play On Start:** The user can define if the Audio Object's clip is to be played at the start of the scene's reproduction.

Figure 4.12: Audio Object 1's Edit Panel.

- **Loop:** The user can define if the Audio Object's clip is to be played in loop; meaning that clip might be successively played multiple times before the scene is stopped, depending on its length.

- **Tag:** The tag is a textual reference added for scalability of the system. While it is not used currently by any of the triggers available, it was part of one of the triggers envisioned for Immerscape. The tag represents a "type" for the Audio Object. For example, if we have multiple Audio Objects that all represent horses' sounds, that we want to play at the same time, it might be useful to have a trigger that defines playing behavior for a group identified by the tag "HORSE". Currently, the trigger behavior described is achievable but the user must define the same trigger event for all the Audio Objects individually.

- **Altitude:** The user can define the height from the ground they want the Audio Object to be at. In order to define this, the gravity property of the Rigidbody component that to be disabled. This property was included so the Player (user) could experience sound that comes from above them.

- **Volume:** The user can define the Audio Clip's volume.

- **Min Dist:** The user can define the Audio Clip's minimum distance limit that can be set for volume roll-off.

- **Max Dist:** The user can define the Audio Clip's maximum distance limit that can be set for volume roll-off.

- **Gain:** Audio gain is calculated as a ratio of output power to input power. The unit of measurement is decibels (dB). The user is given a restricted range of gain that can be applied to the Audio Object's clip.

Add Movement to Audio Object

To add movement to an Audio Object, the user must point the cursor to it and simple press the key M. This will open the object's Movement Panel that displays all the information concerning the Audio Object's movement route. The state of the parameters that are adjustable through the Movement Panel is kept in the Audio Object's Object Movement component. In this panel (Figure 4.13), the user can edit the following parameters for each Audio Object:

Figure 4.13: Audio Object 1's Movement Panel.

- **The number of points:** The user can define the number of points that define the movement. The movement is limited to a max 10-points routes. During the designing phase of the system, this was determined to be enough considering that to define a linear route, it is only needed one point.

- **Speed:** The user can define the speed at which the Audio Object moves.

- **The points to move to:** The user can determined the order the Audio Object follows the points available in the scene. Only the first [number of points] Points will be considered as target points for the movement route.

- **Status:** The user can apply or remove the movement defined to the Audio Object. This is controlled by a boolean parameter.

The movement of Audio Object is added when the scene is played if the Audio Object's status is "SET TO MOVE". The movement script takes into consideration the number of points indicated and the order defined, using them sequentially as targets for the Audio Object's movement. Making good use of the number of points and the list of selected targets, it is possible to define either simple routes (routes that start and stop) or cyclic routes (routes that are played on loop).

Add Trigger to Audio Object

To add a trigger event to an Audio Object, the user must point the cursor to it and simple press the key T. This will open the object's Trigger Panel that displays all the information concerning the Audio Object's associated trigger events. The state of the parameters that are adjustable through the Trigger Panel is kept in the Audio Object's Trigger Info component. In this panel (Figure 4.14), the user can select from the available Points in the environment which one activates the trigger, and then, the user must check off what trigger event is activated.

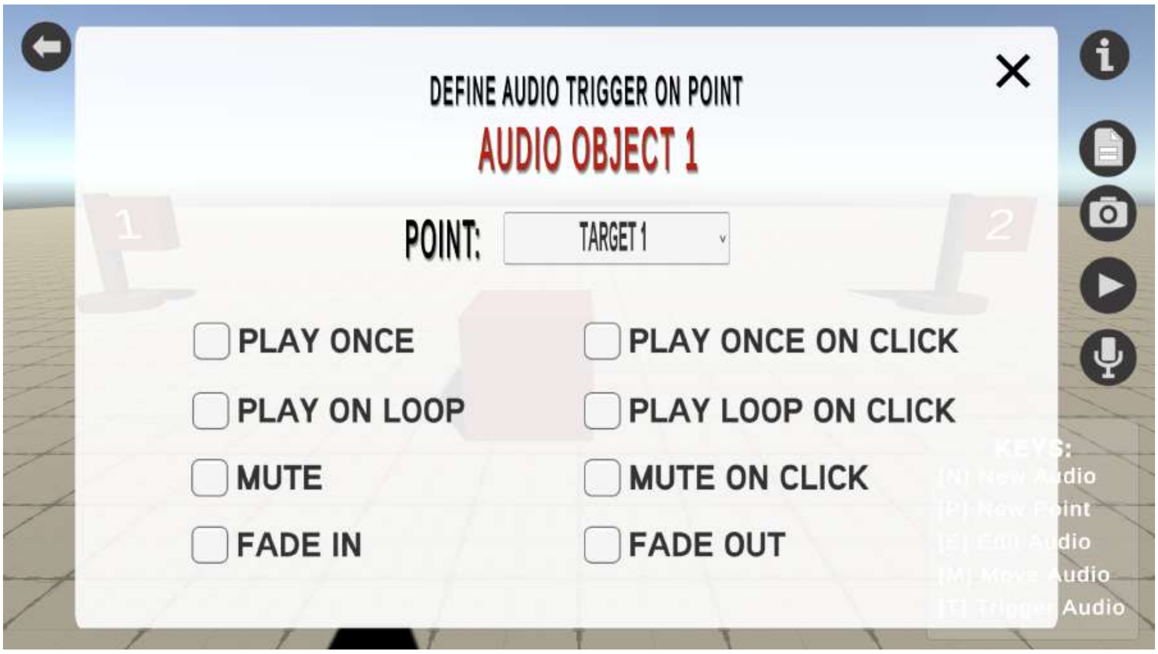


Figure 4.14: Audio Object 1's Trigger Panel.

Currently, Immerscape has the following Trigger Scripts available to add to the Audio Objects:

- **Play Once:** The trigger event gets the audio source to play its clip just once. The trigger is activated when an Audio Object reaches the selected target.
- **Play On Loop:** The trigger event gets the audio source to play its clip on loop. The audio clip is played until another trigger event mutes it, the player mutes the Audio

Object or the reproduction of the scene is stopped. The trigger is activated when an Audio Object reaches the selected target.

- **Mute:** The trigger event mutes an audio source. The trigger is activated when an Audio Object reaches the selected target.
- **Fade In:** The trigger event gets the audio source to play its clip; starting with the volume at 0 and progressively increasing the volume of the audio playing until it reaches the maximum value set for the Audio Object. The trigger is activated when an Audio Object reaches the selected target.
- **Fade Out:** The trigger event gets the audio source's clip to progressively stop playing; starting with the volume at Audio Object's maximum and progressively decreasing the volume of the audio playing until it reaches 0 and the audio source is muted. The trigger is activated when an Audio Object reaches the selected target.
- **Play Once On Click:** The trigger event gets the audio source to play its clip just once. The trigger is activated when the user clicks on the selected target during play or recording mode.
- **Play Loop On Click:** The trigger event gets the audio source to play its clip on loop. The audio clip is played until another trigger event mutes it, the player mutes the Audio Object or the reproduction of the scene is stopped. The trigger is activated when the user clicks on the selected target during play or recording mode.
- **Mute On Click:** The trigger event mutes an audio source. The trigger is activated when the user clicks on the selected target during play or recording mode.

This makes it possible so endless more types of triggers events can be added to Immerscape, following the same process. The Trigger scripts are reusable components so that the same behavior can be added to multiple Audio Objects. It also makes it so that the users can create the soundscapes workflow without having to code any behavior.

Move Audio Object

The Audio Object is created in a position selected by the user. However, when composing the scene during the development of the initial solution (section 4.2), it was determined that it was useful to adjust the Audio Object's position in the scene. This functionality was implemented with a Picker component.

The Audio Object are created with a Picker that allows the Player to pick them up by clicking with the cursor on them and hold it (see Figure 4.15). The Audio Object selected is raised as it was being picked by the Player and as the Player moves in the environment, the Audio Object is carried around; enabling the user to re-position it in the scene. While implementing this mechanism, it was important to consider the Audio Object's defined altitude value. If the Audio Object had an altitude greater than zero (was up in the air),

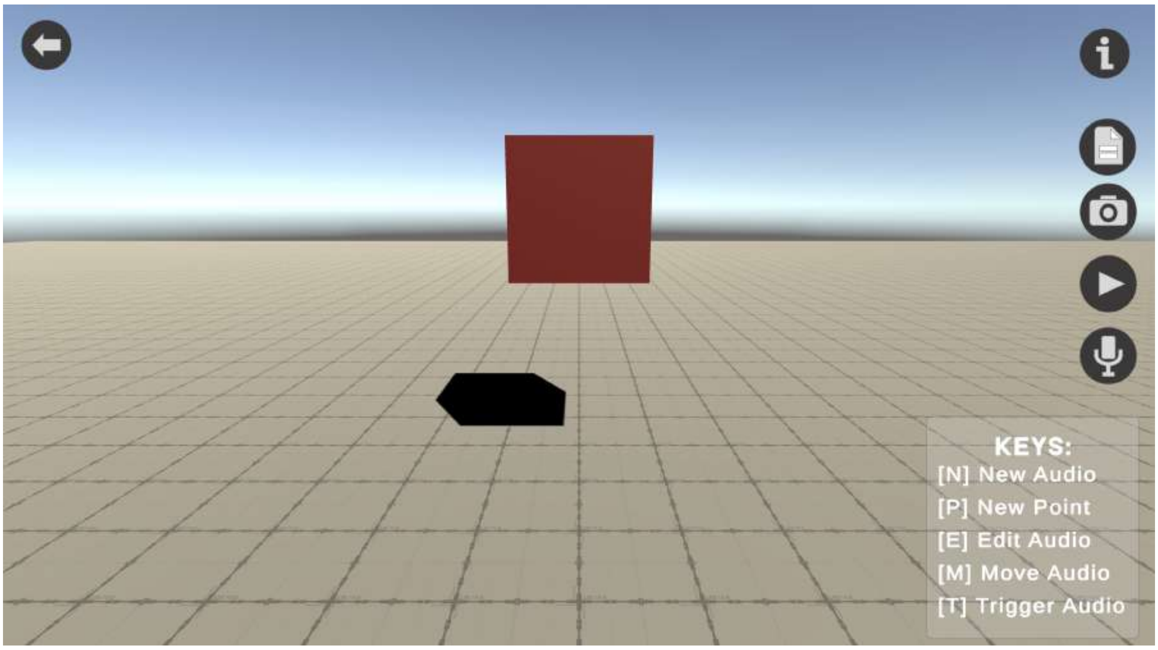


Figure 4.15: The Player picking up the Audio Object to move it.

the Player was picking it up from the air and not the ground. It was important to make sure that once the Audio Object was "dropped" by the Player in their new position, instead of hitting the ground, the Audio Object returned to their previously defined altitude. This functionality is also accessible when the camera is in the Up-view mode.

Play the Scene's Audio

While building and testing out scenarios during the development of the initial solution, we quickly realized that it would be useful to separate the Play and Record functionalities. The Play functionality can be used to test out the usability of the scene, while composing it (determine if our actions meet our expectations).

When the Play functionality is activated (by clicking on the Play button), Immerscape checks all the Audio Objects to add the necessary missing behavior to them, such as the movement component, to those which had movement added to them. The Play script mechanism is also responsible for starting to play all the Audio Objects who checked off the "Play On Start" option.

The Play button functions as a toggle. This means that it "transforms" into a Pause button when the playing of the scene is activated. Similarly to what was done on start, once the scene's reproduction is stopped, the added components are removed.

Record the Scene's Audio

Being able to record the soundscape composed by the user was a crucial functionality of the application. In order to accomplish it, we explored three different alternatives: the Unity Recorder Scripting API, the Google Resonance SDK Recorder and the open source

.NET audio library, NAudio. All the mentioned technologies presented some problems of integration. The Unity Recorder Scripting API, the Google Resonance SDK Recorder were only available to record inside Unity projects in the Editor mode; while the NAudio version showed incompatibility problems with the version of Unity being used during development.

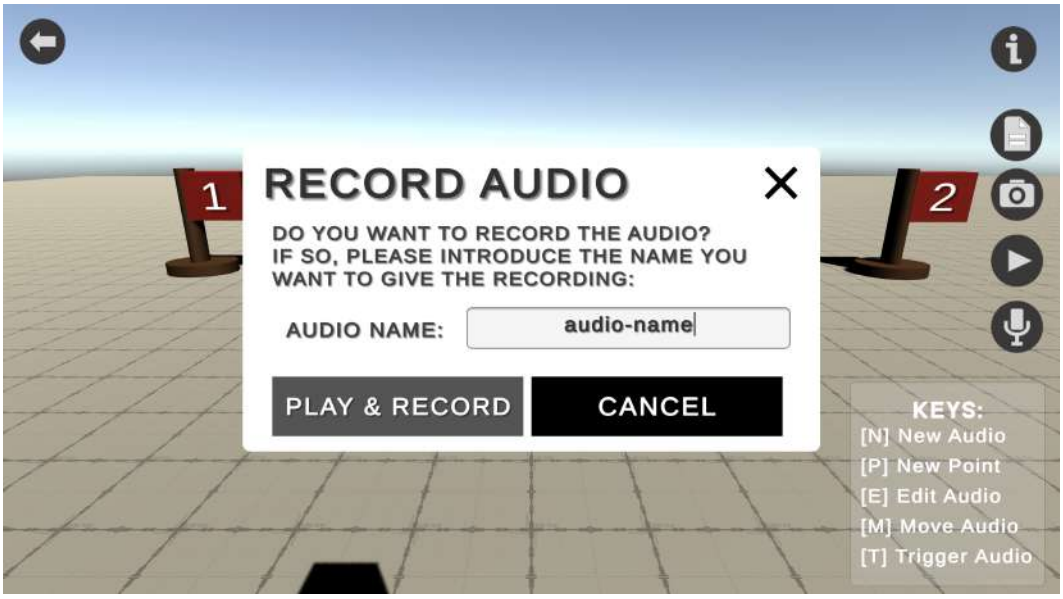


Figure 4.16: Dialogue window before the recording process starts. The user must insert a name for the audio clip that will be generated.

Considering how important it was to record audio with Immerscape, a scripting component was created, Audio Recorder. The script was developed based on a script to save a WAV file (SavWav.cs). When the recorder functionality is activated, it requires the user to introduce a name for the clip that will be recorded (see Figure 4.16). Before starting the recording process, the Recorder makes sure that the active camera is the Player Camera and that it activates the Recorder before playing the scene.

Once the recording process is started, the Audio Recorder component script is activated, and it listens to everything the Audio Listener picks up through an audio filter, writing it into a file until the recorder is stopped by the user (see Figure 4.17).

The file is written into the *recordings* folder inside the application's folder (*Documents\Immerscape*). This is where the user can later retrieve the audio file to use it. In the current version of the prototype, if given the name of an existing audio file in the recordings folder, Immerscape simply overwrites it.

Save Project

At any point, when in the project environment, the user can save the ongoing project. This option is available by clicking the save button available on the project's environment interface, or it is activated when the user tries to quit the project environment. Whenever

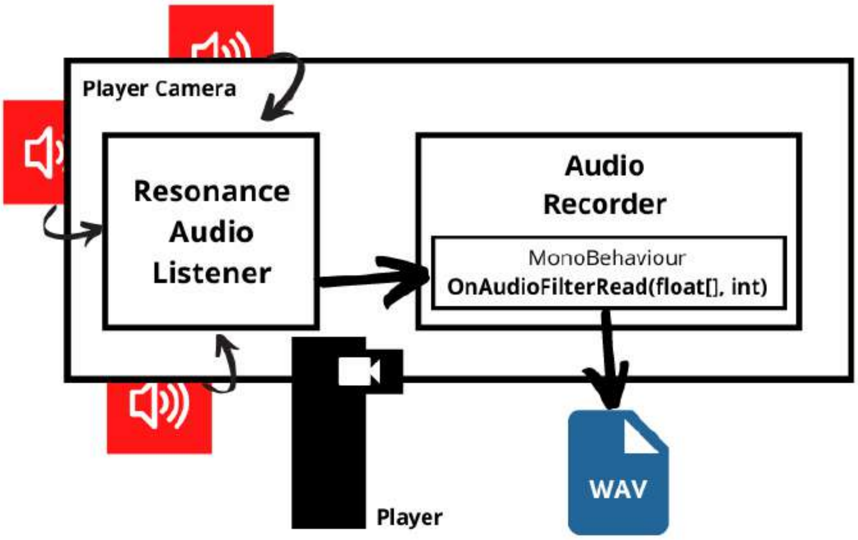


Figure 4.17: The scheme of the implementation of the Audio Recording process. OnAudioFilterRead implements a insert a custom filter writes the audio samples processed by the Player's Audio Listener into an audio file.

the user tries to quit the project environment, the user is asked if they intend to save the progress made. Then, the user can opt by saving the progress or not (see Figure 4.18).

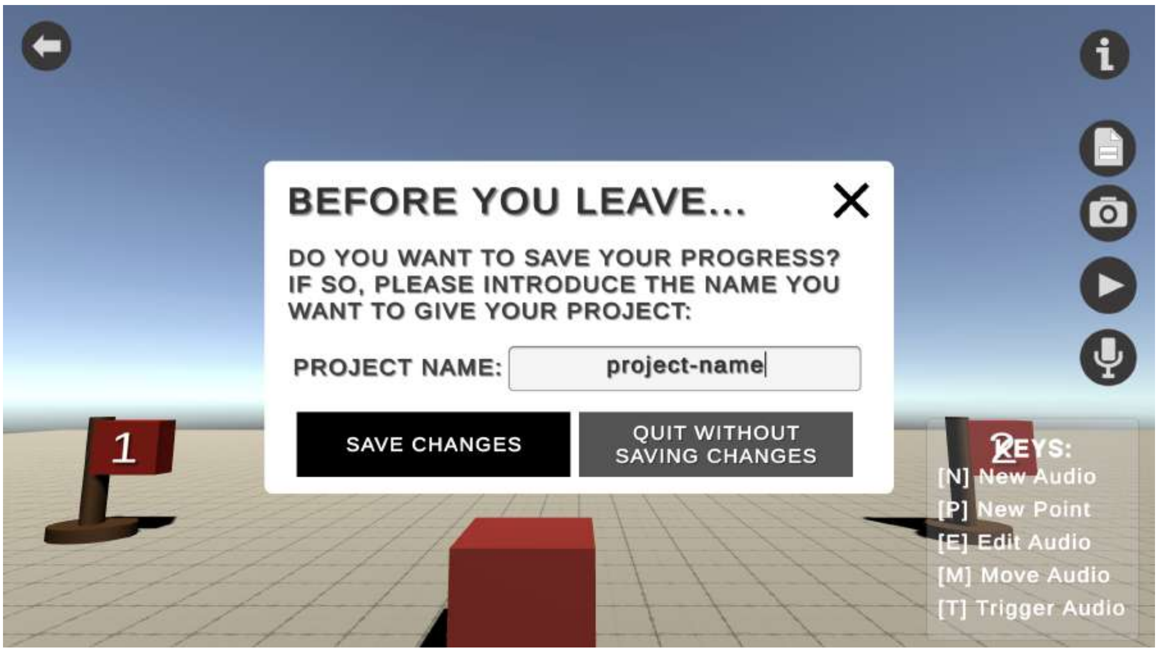


Figure 4.18: Dialogue window before the user can quit the project environment.

The Unity Engine provides a class, Player Prefs, that stores Player preferences between sessions. It is an option for developers to store persistent state as it can store string, float and integer values into the user's platform registry. However, due to the reasonable complexity of Immerscape and its potential for expansion, a custom Save System was implemented instead, using Serializable data files. The Save System is responsible for saving and loading the different saved projects inside Immerscape.

Figure 4.19: Dialogue window before the project is saved. The user must insert a name for the current project.

When users want to save a project, they are required to give it a name (see Figure 4.19). A folder for the project is created on the application's persistent path, and it is on this location that all the binary encoded files concerning the project will be saved. The Save System saves a file per Audio Object and Point, as well as one file for the general information about the project's scene and context at the time it was saved.

Reopen/Edit Saved Project

Outside of the project environment, in Immerscape's main menu, the user can opt to reopen and edit a saved project. When this option is selected, the user is shown a list of all saved projects (see Figure 4.20). The projects are ordered by the date of the last update, from most recent update to the least recent update.

When one of the list items is selected, the Save System reads the scene setup files and it opens the referenced pre-defined environment, loading all the saved elements into it and setting the Current Context to the loaded project's context. This enables the user to resume editing the project as if it had never been closed.

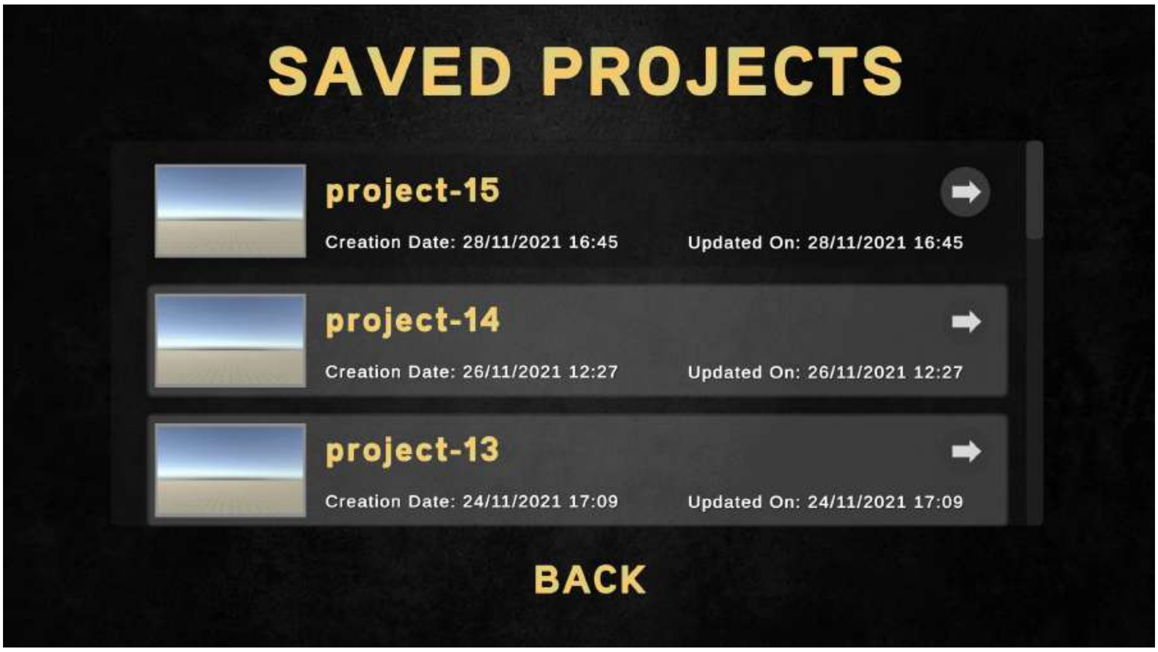


Figure 4.20: The Saved Projects screen. The user is hovering the project-15 list entry. If the user clicks on it, project-15 will be reopened.

Listen to Recorded Audio

Similarly to the "Reopen/Edit Saved Project" functionality, in Immerscape's main menu, the user also has access to a Recordings screen (see Figure 4.21). When the users opt to go to this screen, they are shown a listing of all recordings made with Immerscape. The recordings are ordered by date of creation, from the most recent to the least recent recording.

When one of the items is selected, the audio file is retrieved from the recordings folder and played so the user can listen to it. By deselecting it, the audio playing is paused.

Help

The help mechanism was incorporated as last resource or learning tool for a new user. Immerscape was developed with the aim of allowing users to quickly become familiar with it, and be able to make good use of all its features and capabilities.

Despite having most commands directly displayed on the interface, an information panel was added. This panel can be activated by clicking on the information icon or by pressing the key I. The help/information panel shows users a summary of all the shortcuts they can use to perform most of the tool's functionalities.

Additionally, a link to a tutorial slider was added to the information panel. The tutorial slider goes over the core functionalities of Immerscape, providing the user with some simple instructions on how to navigate the tool's functionalities.

Figure 4.21: The Recordings screen. The user is hovering the audio-15 clip list entry. If the user clicks on it, the audio clip will start playing.

4.4 Summary

This chapter presented the development of a user generated soundscape tool, Immerscape, that is the proposed solution to find answers to the research questions raised in this dissertation. The different sections of the chapter show how Immerscape was designed to help non-expert users to easily create immersive audio that represents a historical soundscape.

The chapter started by giving an overview of the technology and techniques used in the development of the system, which allowed to create the immersive audio tool (section 4.1). We discussed the initial approach taken to create the immersive audio (section 4.2) as well as explained how the proposed solution, Immerscape, evolved from there.

Finally, we introduced Immerscape. Firstly, we gave an overview of the application and its main purpose (section 4.3). Then, we presented the analysis and design process of the system (subsection 4.3.1): we specified the requirements the system must abide by (subsubsection 4.3.1.1), we introduced the tool's main features (subsubsection 4.3.1.2) and we presented the system's architecture (subsubsection 4.3.1.3). Lastly, we detailed Immerscape's implementation process while explaining how it operates (subsection 4.3.2).

5 Evaluation and Results

This chapter introduces the evaluation and validation of the the proposed solution. It presents the first user study performed on the prototype, its purpose and its results. Additionally, it includes the discussion of the results obtained and the analysis of the participants' feedback.

Once a working version of Immerscape's prototype was achieved, it became important to evaluate its current state before continuing the development as it was important to answer the last research question:

- **How easily can non-expert users create immersive audio representations of historical soundscapes using Immerscape?**

The evaluation process was not only important to determine if the solution was feasible, but also to validate the available features, to collect user feedback on the implementation and determine possible future upgrades to the current functionalities. Therefore, the user study conducted during the evaluation process used a subjective quality inference mechanism.

The following sections of this chapter detail the testing protocol followed while conducting the user tests. The chapter presents the participants' profiles, the testing conditions, and the tasks that the participants were asked to complete. Once the testing process is described, the chapter analyses the data collected, presenting the results while discussing some of the user feedback obtained.

5.1 User Study

In order to assess Immerscape's functionalities, the quality of the audio content generated, the immersiveness of the audio generated as well as the ease of use and acceptance by users, a user study was conducted on its first working prototype.

Since the work presented in this dissertation was developed during the COVID-19 pandemic, the user study was performed remotely over Zoom. Immerscape ran in the researcher's computer while the participants tested the tool using Zoom's remote control in their own computers. For the test, we also used Zoom's shared audio functionality which allowed the participants to listen to the audio being reproduced in the researcher's computer during the test. The participants were required to use either headphones or earbuds.

The study's protocol consisted of a set of ten small related tasks that went over Immerscape's most relevant features and that resulted in the completion of a trivial project that represented a simple soundscape. In the beginning of the test, the participants were given a quick overview of the tool's functionalities and some documentation, including a testing guide, a consent form and a questionnaire (further details are presented in Appendix C). For all tasks, the participants received written instructions.

During the test, the participants were asked to fill a questionnaire made available online. The questionnaire was divided into thirteen different sections: one concerning the participant's profile, another one concerning the testing environment, ten other sections with task specific questions for each one of the ten tasks, and finally a last section for an overall evaluation of Immerscape.

Before the start of the testing process, the participants were asked to fill the sections about their profile and the testing environment. At the the end of each task, the participants filled the section corresponding to the task they had just performed. After completing all tasks, the participants had the opportunity to fill in an additional form section about general aspects of the tool, as well as ask any questions or make any suggestion about the tool. Any other relevant comments, observations and suggestions made by the participant during the testing process were also registered by the researcher and considered as user feedback.

5.1.1 User Characterization

The test pool for Immerscape's user study is formed by seventeen test subjects. The participants were divided in two distinct groups based on their study background: Developers and Non-Developers. All volunteers were attributed a numeric identification reference in order to guarantee data protection.

The participants with some background in software development or engineering were considered as part of the Developers category while people with other academic backgrounds were considered part of the Non-Developers group. The groups had an unequal number of testing subjects. From the total of 17 participants, 10 were considered Developers while 7 were considered Non-Developers.

It was important to have the two groups, not only to test if there is any significant difference of use between Developers and Non-Developers, but also to receive feedback from users with different perspectives on software. Considering that the user study was conducted on Immerscape's first working prototype, it was relevant to have the feedback of other software developers who, by nature of their experience, have a more technical perspective regarding the software and might have more specific feedback when it comes to the implementation of the features.

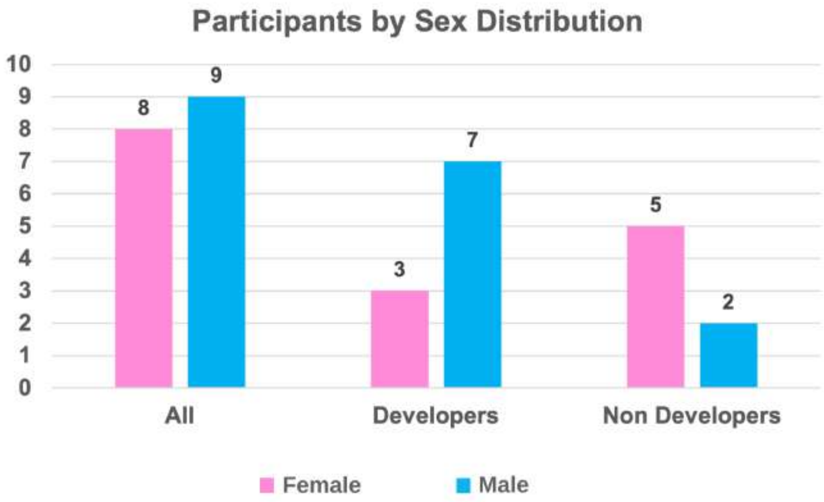


Figure 5.1: Distribution of participants by sex.

Out of the total of 17 participants, 53% are male and 47% are female. The Developers group was constituted by 70% male participants and 30% female participants. On the other hand, the Non-Developers group was constituted by 72% of female participants and 29% of male participants. Figure 5.1 shows the distribution of participants by gender.

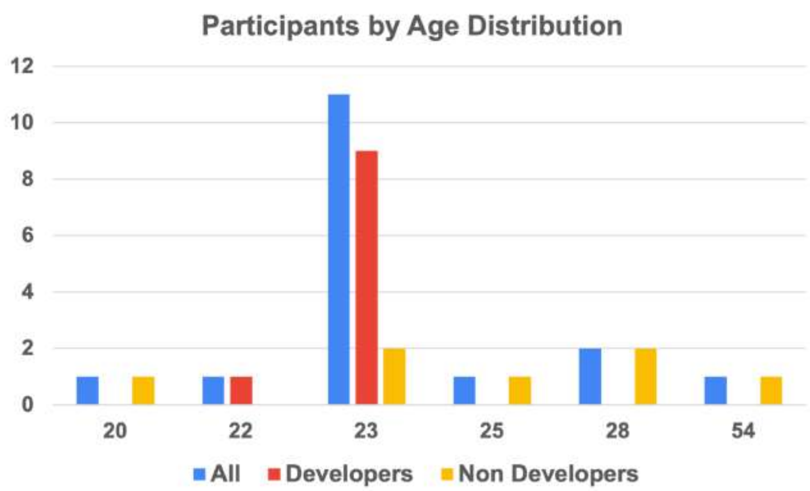


Figure 5.2: Distribution of participants by age.

The test subjects' ages ranged from 20 to 54 years old. The Developer group participants ages ranged between 22 and 23 years old, while the Non-Developers group participants presented a wider range of ages. Figure 5.2 shows the distribution of the participants by age.

The participants were questioned about permanent and temporary hearing problems and only one of the participants (Developer) declared having a temporary hearing condition (tinnitus in his right ear). The remaining participants asserted having no hearing problems that could affect their participation at the time of the test.

In addition, the participants were asked if they had previously used any audio creation software, to which 47% of the participants answered yes, naming the tools they had used, and 53% of the participants answered no, claiming to have never used an audio creation software tool before. The participants that had knowledge of other audio software explained having used it for personal or school related projects. Most of the participants named Adobe Audition and Audacity as the audio software they had used previously. Only one of the participants, a Non-Developer with background in Musicology named other audio software (Logic, FLStudio, CSound, MAX/MSP and Ableton).

The volunteers were required to wear either headphones or earbuds for all tasks; Figure 5.3 shows the distribution of participants by the type of device used during the testing process.

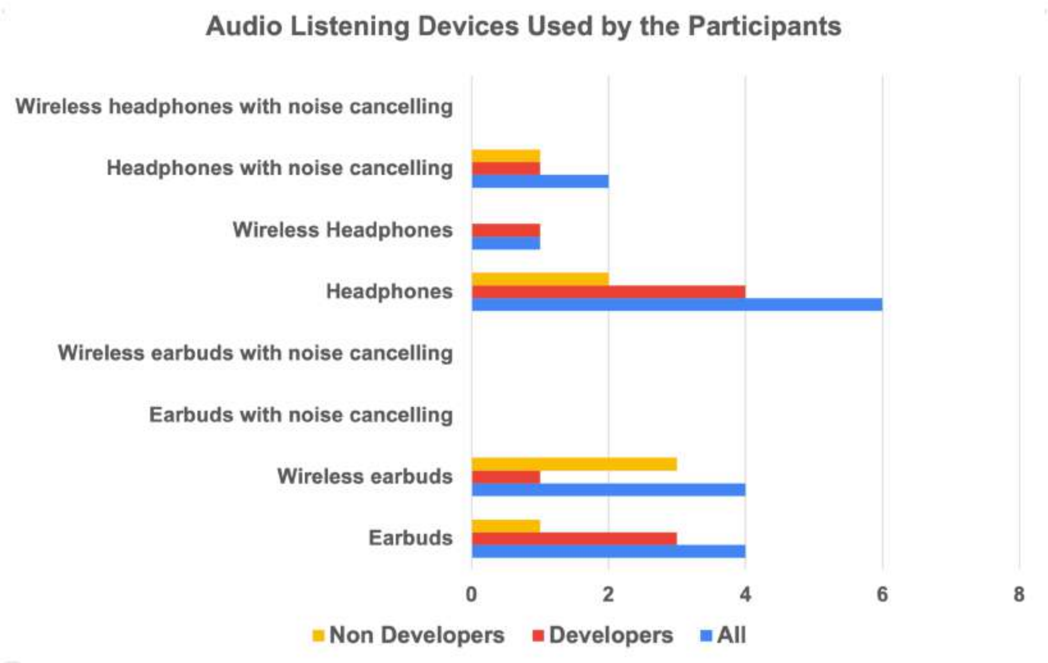


Figure 5.3: Distribution of participants by the audio output devices used during the testing process.

Lastly, the participants were queried on the levels of noise in their testing environment. Only two participants reported doing the test in an environment with a considerable level of noise. The remaining participants assured that the background noise was either non-existent or minimal (not considerable). Figure 5.4 shows the distribution of the participants by the reported level of noise of the testing environment.

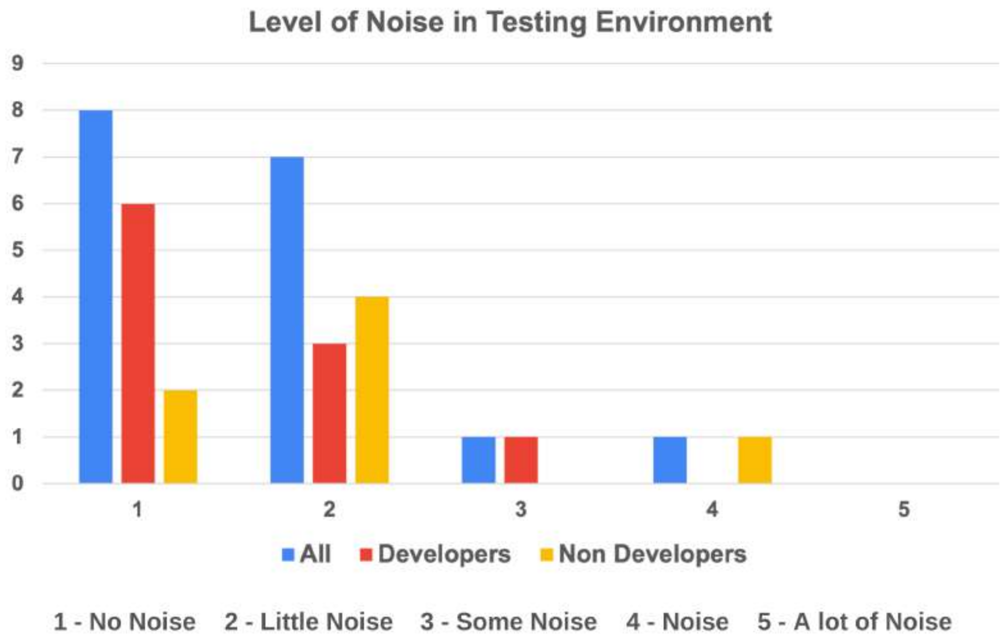


Figure 5.4: Distribution of participants by the reported level of noise of the testing environment.

5.1.2 Tasks

Before starting the tasks, the participants were given a small description of the scene they would be composing during the test and for which they would be creating an immersive audio clip. The imaginary soundscape consisted of a horse trotting in circles around the user, who would fire a shot into the air every time the horse passed by a specific position during its route.

The participants were given access to a folder with the two audio clips to be used during the test (*horse.wav* and *shot.wav*). All the ten tasks presented were designed to accomplish the creation of an immersive audio clip for the previously described soundscape, while exploring all the main features of the tool and getting feedback from the users' interaction with it.

Task 1: New Project & Environment Navigation

This introductory task consisted of creating a new project, by selecting one of the predefined sound environments (the default environment). Once the project was created, the user would test the Player's movements.

This task meant to evaluate how easily users can start a new project and how well they can navigate the environment in an immersive way. The Player is the representation of the user within the scene and as so it was important to determine the quality of its movement capacities.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Do users consider it a positive feature of the application to have a collection of pre-defined environments to choose from when creating a new project?**
- **What other type of generic audio environment do users consider to be important to include in Immerscape?**
- **Can users move the Player back and forward?**
- **Can users rotate the Player to change its movement's direction?**
- **What is the quality of the Player's movement capabilities?**
- **What other type of movement capacities should the Player have?**

Task 2: Scene Set Up

In the second task, the user was given a small explanation of the type of elements that can be used to compose the scene and what each element could be used for. Firstly, the participant was asked to switch to the up-view camera. Following that, the user was asked to populate the environment by positioning one audio object and four mark points in the scene so that it would have a similar layout to the one presented to the testers in the image provided (Figure 5.5). At last, the participant was asked to return to the player-view.

This task was meant to evaluate the user's capability to compose a scene while testing both the player camera mode and the up-view camera mode. Testing this functionality's usability was important because creating the scene's elements is one of the focal features of the tool.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users switch between camera modes?**
- **Do users consider that being able to switch between camera modes is a positive feature of Immerscape?**
- **Can users add a new Audio Object to the environment?**
- **Can users add new Points, given a specific layout?**
- **What other type of elements do users consider that should be available in Immerscape to compose scenes?**

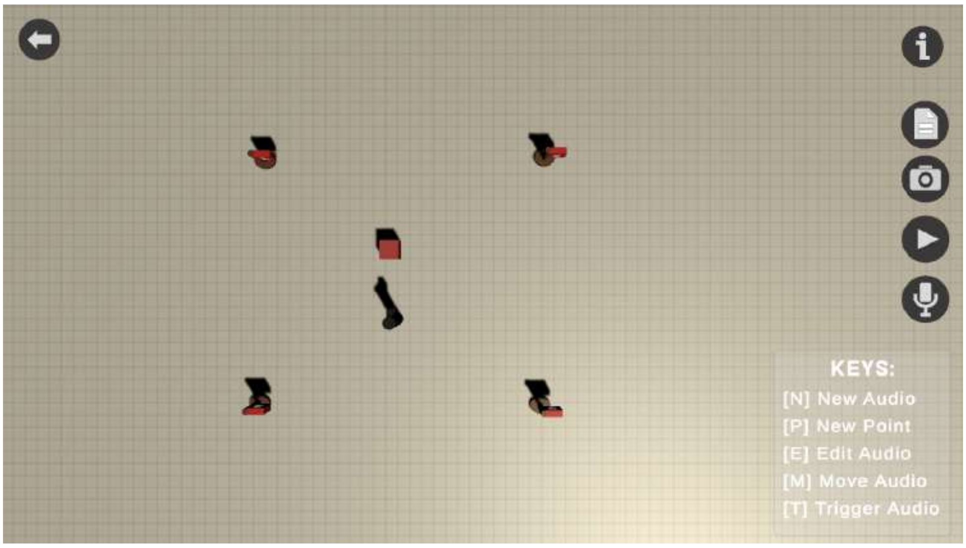


Figure 5.5: The image provided to the participants to demonstrate how the elements should be positioned within the working environment (Up-view camera mode). The elements include an audio object (square in the middle of the scene) and four mark points around it.

Task 3: Editing Panel

The third task required the user to open the editing panel of the Audio Object and edit the audio source. It was explained to the participant that when created, the Audio Object is empty, meaning it must be edited by the user to determine the properties of the sound it emits. The editing process included uploading an audio clip, as well as selecting that the sound should play in loop and on start. The player was also asked to test the altitude field, and given the liberty to experiment with other parameters. Lastly, the participant was asked to close the editing panel.

It was important to test this feature because the audio object must always be edited after it is created, in order for the sound to be included in the scene. Therefore, this is one of the most frequently used features of the Immerscape.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users open the Audio Object's editing panel?**
- **Can users upload the audio clip into the Audio Object?**
- **Can users configure the Audio Object to play on loop and on start?**
- **Are users able to experiment with the altitude parameter and then place the Audio Object back in its original position?**
- **How easy do users consider the editing of the Audio Object's properties?**
- **What other properties do users consider that they should be able to edit?**

Task 4: Testing the Audio

In this task, the user would play the scene to listen to the sound and navigate around the environment to explore the spatial dimensions of the sound being reproduced. Lastly, the participants were asked to stop the scene's reproduction.

This task tested how easy it was, for users, to reproduce the scene, to experiment the result of their actions, and assess the quality of sound immersion provided by Immerscape.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users reproduce the scene?**
- **Can users stop the scene's reproduction?**
- **Can users detect the changes in the direction of the sound when changing their relative position to the Audio Object (sound source)?**
- **Do users consider that being able to move the Player during the reproduction of the scene is a positive feature of Immerscape?**
- **How easy do users consider to test the result of reproduction, during the process of composing the scene?**

Task 5: Movement Panel

For the fifth task, the user was asked to open the Audio Object's movement panel in order to add movement to it. The user was instructed to add movement that used the four Points and a speed of 4. After adding the movement, the user would play out the scene to listen to the spatial effect of the sound moving. After pausing the scene, the user had to move the audio object back to its initial position. Lastly, the user was asked to close the movement panel.

The objective of this task was to determine the ease with which the user can apply movement to Audio Objects, as well as determine if users were capable of rearranging the objects' positions in the scene.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users open the Audio Object's movement panel?**
- **Are users able to define the desired movement route?**
- **Can users "pick up" the Audio Object and place it back in its original position?**
- **Do users consider that defining movement for an Audio Object using a maximum of 10 points is an important limitation of Immerscape?**
- **How easy is it to add movement to the Audio Object?**

- **What other aspects of the Audio Object's movement would the users like to be able to edit?**

Task 6: Trigger Panel

For the sixth task, the user had to create another audio object and edit it according to the instructions. The user should be able to easily create a second Audio Object. Users were also asked to open the Audio Object's trigger panel and add a trigger event to the second object so that it only played when triggered and not on the scene's start like the first one. The users were asked to close the trigger panel and test the resulting auditory scene by playing it. Lastly, when stopping the reproduction, the users had to reestablish the Audio Object's original position.

The objective of this task was to determine how easily the user can create a new Audio Object and open its panels, as well as evaluate how easy it is to define trigger events for a specific Audio Object. This task does a second iteration of things previously asked of the user. Therefore, it can be used to determine the user's capacity to retain information from previous tasks.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users create and edit a new Audio Object?**
- **Can users open the Audio Object's Trigger Panel?**
- **Can users select the Point that holds the trigger event?**
- **Can users select a specific trigger event?**
- **How easy do users consider adding a trigger event to the Audio Object?**
- **What other trigger events do users consider that should be available to choose from in Immerscape?**

Task 7: Record Immersive Audio

Once the users completed the first six tasks, they should have created a simple soundscape composed of two sounds. Assuming this was the final scene, the users were asked to record it and give it the name *audio-[number]*, where *[number]* was their testing identification number. Lastly, the users were asked to stop the recording process whenever they wanted.

This task aimed to test one of the focal features of the application: the recording of the composed immersive audio. This task was important to determine how easily a soundscape can be recorded by an inexperienced user.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users start the recording process?**

- **Can users stop the recording process?**

- **Do users consider that being able to listen to the audio being recorded is a positive feature of Immerscape?**

- **How easy do users consider recording an immersive auditory scene using Immerscape?**

Task 8: Save & Quit Project

For this task, the users were asked to leave the current project environment and save the progress made on the project before doing so. The users were asked to save the project and give it the name *project-[number]*, where *[number]* was their testing identification number.

The objective of this task was to test how easily projects can saved as well as to test the project environment's exit mechanisms.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **Can users save the progress made on a project?**

- **Do users know how to exit the project's environment?**

- **How important is it to users to be able to save the progress made on a project?**

Task 9: Access Recorded Audio

For this task, the user had to take advantage of the Recordings screen to access the previously recorded audio, and listen to it by clicking on its entry in the list of recordings. User feedback and suggestions on what features should compose this screen are highly important to the next stage of development.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **How important to the users is it to be able to access the recordings made through Immerscape's interface and be able to play them?**

- **How important to users is it to have the recordings ordered by descending date of creation?**

- **What other information/functionalities should be available, concerning the recordings?**

Task 10: Access Saved Project

Similarly to Task 9, the aim of Task 10 was to evaluate the usefulness of the features of the "Saved Projects" screen, as well as to test if the user would easily reopen an ongoing project.

During the execution of this task, the participants were encouraged to express any type of feedback. From this resulted the participants' answers to the following questions:

- **How important to the users is it to be able access and reopen saved projects?**
- **How important is it to the users to have the saved projects listed by update date (from most recently update to least recently updated)?**
- **What other information/functionalities should be available, concerning the saved projects?**

5.2 Results

As previously said, in order to validate the proposed solution, a series of user tests were performed on Immerscape's prototype. The referred tests were described in detail in section 5.1 (User Study) and subsection 5.1.2 (Tasks). Next, the feedback and the results obtained for each task, during the evaluation process, is presented.

All participants completed the ten proposed tasks successfully. From the tests originated, for each participant, an audio project and the corresponding recording of its acoustic environment.

Task 1: New Project & Environment Navigation

All participants were able to successfully complete Task 1. Despite differences in adaptation when asked to navigate the environment, all participants were able to create a new project, select the specified environment, move the Player back and forward and rotate the Player to change its direction.

Due to a Zoom constraint, in the remote control mode, it was not possible to test the Player's ability to look up and down. However, the researcher conducting the test took over to show how the Player can look up and down after asking the participants how they would do it. From this question, participants suggested, multiple times, a mouse hold and drag vertically to shift the Player's view.

All participants considered that having a collection of pre-defined sound environments, which the user can select to work on, is an advantage of the tool. When questioned about what other type of generic audio environment they would consider important to include, the participants gave some suggestions.

Developer participants suggested environments such a theatre, a coliseum, a stage environment, a stadium or a festival environment. Non-Developers suggested some of the same environments, namely the theatre, the stadium and the concert arena, but they

Figure 5.6: Distribution of participants (Developers/Non-Developers/All) by level of easiness at completing Task 1.

had other, and more specific, suggestions to make regarding the environments. One participant suggested including the sound of the Player's steps and other basic physical movements of the Player. Another interesting suggestion was to recreate the audio environment of a cinema setup.

Creating a new project and exploring the project environment selected with the Player was considered easy (29%) or extremely easy (59%) by most of the participants. Only 12% of all participants considered that completing Task 1 had medium difficulty.

However, if we distinguish between Developer participants and Non-Developers, the results show (Figure 5.6) that, generally, the Non-Developer participants considered the task harder to complete than the Developer participants. Only Non-Developer participants consider it to have Medium difficulty (12%), and while 30% of Developer participants considered it easy and 70% considered it very easy to complete, 29% of Non-Developer participants considered easy and 59% considered it very easy to complete.

Concerning the player's movement (results shown in Figure 5.7), the majority of all participants considered it to be satisfactory (35%) or very satisfactory (53%). Only 12% of all participants considered to be just somewhat satisfactory. Overall, the quality of the Player's movement capabilities is good or very good. However, both Developer and Non-Developer participants had suggestions on what other type of movement capacities the Player should have.

Concerning the participants' suggestions, 40% of the Developer participants considered that the player should be able to jump; 10% wished the player's vertical position

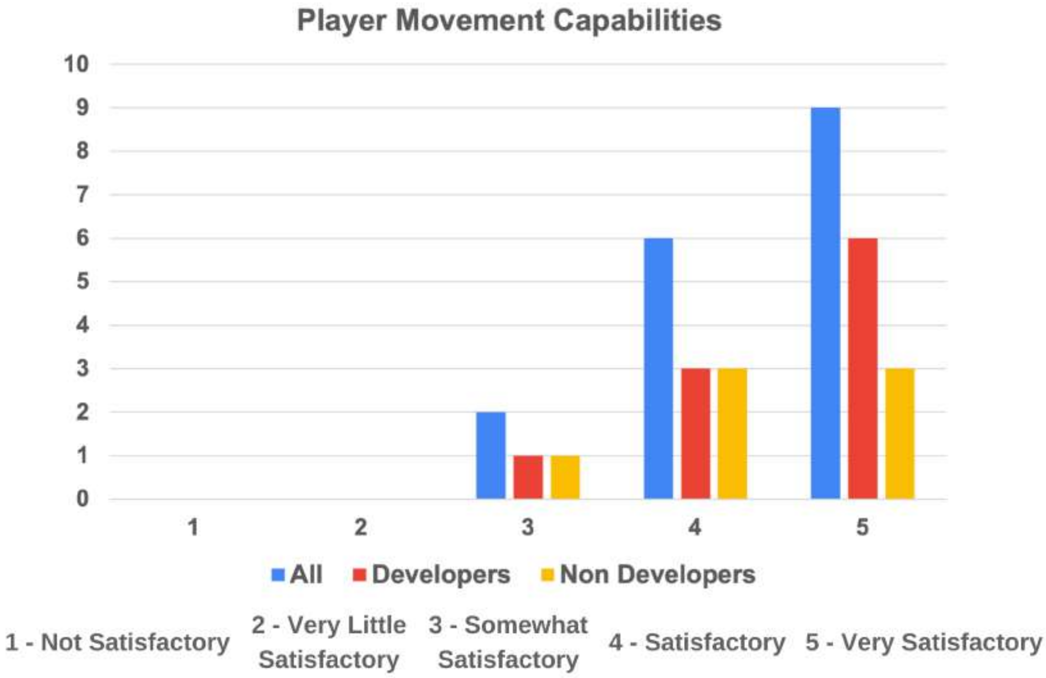


Figure 5.7: Results of the Player's movement capabilities evaluation.

could be adjustable; another 10% suggested that the player should be able to move laterally without having to rotate its camera; while other 40% of Developer participants considered that the current movement capacities of the player were enough.

One of the Non-Developer participants suggested having adjustable height for the Player, so the Player can represent an adult or a child perspective. 12% of Non-Developers suggested adding jumping to the Player's capabilities while 6% of the Non-Developer participants wished the could have a "running" mechanism (walk faster at times).

Task 2: Scene Set Up

All participants were capable of switching between the two camera modes easily. Additionally, all the participants considered being able to switch the camera modes, a positive feature of Immerscape. Two of the participants even commented on how the up-view mode is helpful when composing the scene.

Most participants quickly realized that they had to use the computer keys to create the elements. However, during the first creation attempt, some of the participants were not aware that the object would be created at the position indicated by the cursor. All participants were able to add a new Audio Object to the environment.

For subsequent elements, the Points, all participants were aware of the cursor's position before creating the element. Only one participant, a Non-Developer, and the oldest participant at 54 years old, had some difficulty placing the points as the figure in the testing guide indicated. However, this was not because the participant did not know how

Figure 5.8: Distribution of participants by level of easiness at completing Task 2.

to create the Points but rather because they had difficulty navigating the 3D environment from the Player's perspective.

Overall, 53% the participants considered Task 2 easy to complete, while only 29% of all participants considered it very easy to complete. However, out all participants, 12% considered it of Medium difficulty while 6% considered it difficult to complete. As Figure 5.8 shows, a Non-Developer participant, the same participant that had difficulty placing the Points, considered it difficult to complete.

When questioned about what other type of elements they would like to have to provide support in composing scenes, 30% of Developer participants suggested having elements that functioned as timed triggers of Audio Objects. The Non-Developer participants did not have suggestions of other elements, considering that the Audio Objects and Points were enough to compose the scenes.

Additionally, some participants wished they could also move the mark Points like they can move the Audio Objects. Two of the participants of Developer participants suggested a different Audio Object moving mechanism for the Player when in up-view mode. They wished that the interaction in this mode was less rooted in the player and suggested that it could be conceived as a third-person interaction (drag and drop independent of the Player's position). Some of the Developer participants also wished that it was possible to delete the elements.

Task 3: Editing Panel

All participants were able to open the Audio Object's Edition panel. They were all able to upload an audio clip from the one of the device's folder, configure the Audio Object to play on loop and at the scene's start. All participants experimented with the altitude parameter and figured out the two different ways they could place the Audio Object back in its original position.

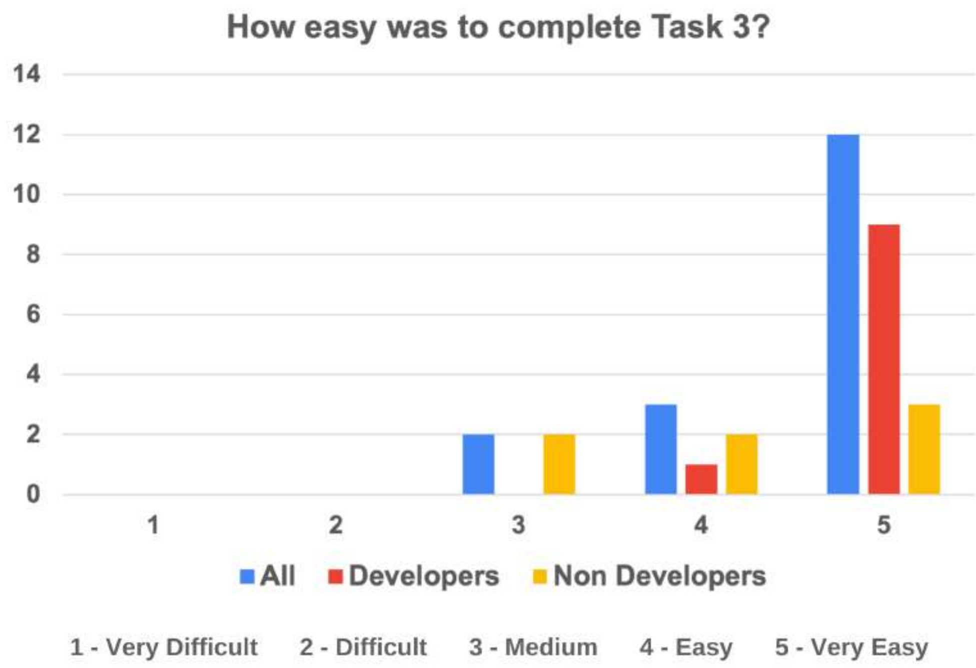


Figure 5.9: Distribution of participants by level of easiness at completing Task 3.

The task was considered very easy to complete by 70% of all participants; easy to complete by 18% of all participants and as having medium difficulty by 12% of all participants. The distribution of participants by level of easiness at completing Task 3 is presented in Figure 5.9.

However, when we look at the Developer participants' results, 90% of the Developer participants considered it very easy to complete, while only 42% of Non-Developer participants found it very easy to complete. Similarly, while the remaining 10% of Developer participants considered the task easy, the Non-Developers split evenly at 29% between considering the task easy and of medium difficulty.

Similarly, when asked to evaluate how easy it is to edit the Audio Object's properties (results shown in Figure 5.10), all of the Developer participants consider it very easy to do while the Non-Developer participants seemed to have experienced different levels of difficulty. However, overall, 88% of all participants considered the editing of the Audio Object's properties very easy.

When queried on what other Audio Object's properties they would like to be able to edit, the participants presented a few suggestions. The most frequent suggestion was the editing of the Audio Object's color. The participants argued that it could be used to

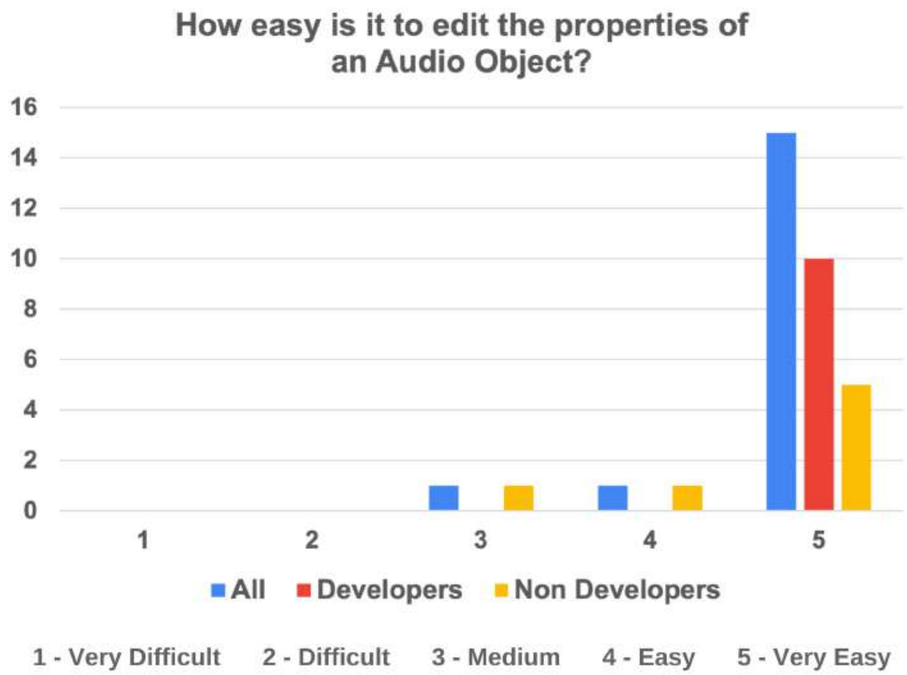


Figure 5.10: Results of users' evaluation on how easy it is to edit the Audio Object's properties.

distinguish the different audio sources without having to open one of its panel to know which one it is.

Developers suggested adding mechanisms to edit the clip's frequency and amplitude; a mechanism to determine that the clip will play a X amount of times; and a mechanism to be able to cut and trim the uploaded audio clip within Immerscape, as well as add sound effects (e.g. distortion) to the clips. The only suggestion made by Non-Developer participants was the editing of the Audio Object's color.

Task 4: Testing the Audio

All participants were easily able to play, experience the immersiveness of the audio playing and stop the reproduction of the scene. While the scene was playing, all participants were able to detect, to some degree, the different directions of sound while moving the player around the static audio object.

All participants tested moving the Player around the Audio Source and the effect that distance has on the reproduced audio. They all considered that being able to move the player while the scene was being reproduced was an advantage of Immerscape.

Overall, as shown in Figure 5.11 the large majority of participants (94%) considered Task 4 very easy to complete. Only one participant (Non-Developer) considered it easy to complete. This can be explained by the task's testing nature.

Developer participants considered it an intuitive, simple and clear task. Non-Developer participants commented on how easy it was to execute and how the play/stop button are

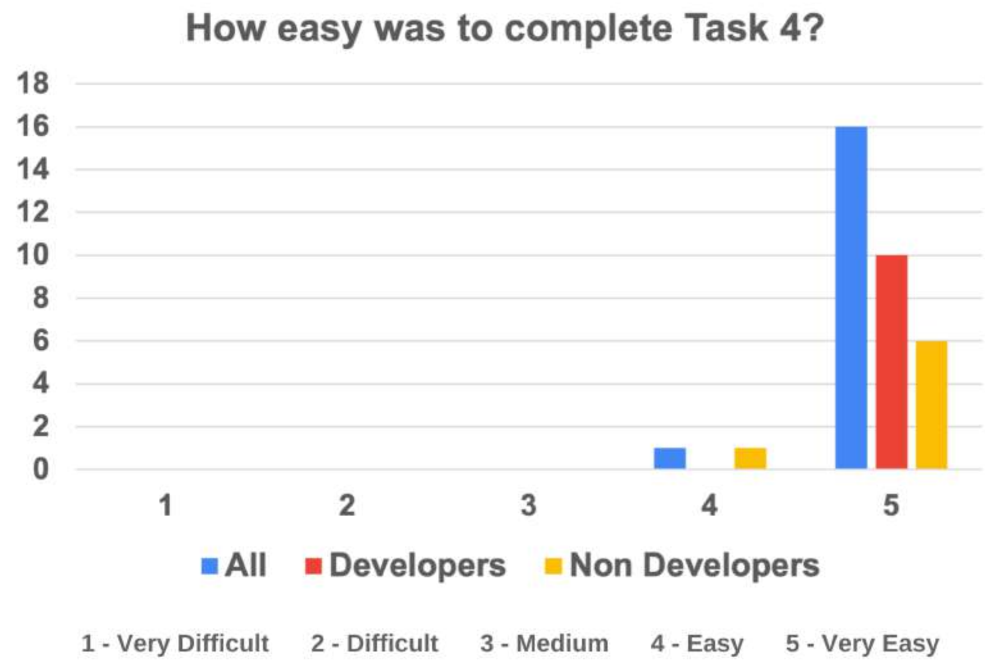


Figure 5.11: Distribution of participants by level of easiness at completing Task 4.

easily recognizable. One of the Non-Developer participants suggested having a keyboard key as a shortcut to start and stop the reproduction of the scene's audio (e.g. space key).

During the testing process, some participants reported on experiencing lag or cuts in the reproduction at times. This was mostly attributed to the testing conditions of the connection over Zoom as lag and cuts were not experienced locally. The scene's audio was recorded in a posterior task and sent to the participants so they could listen to it in full quality on their own devices.

Task 5: Movement Panel

Although all participants easily opened the audio object's movement panel, 41% of all participants struggled to understand how to define the Audio Object's movement, specifically why they had to indicate the number of points that describe the movement. In the end, all participants were able to define the movement route and apply it to the Audio Object.

All participants were able to "pick up" the Audio Object and place it back in its original position. At this point, some participants had already independently tried this functionality in previous tasks.

As shown in Figure 5.12, adding movement to an Audio Object was considered easy (35%) or very easy (59%) by the majority of all participants. Only one participant (Non-Developer) considered the adding movement as moderately difficult (medium).

When asked if they considered only having ten points to define the audio object's movement route a big limiting factor (Figure 5.13), 59% of all participants answered no,

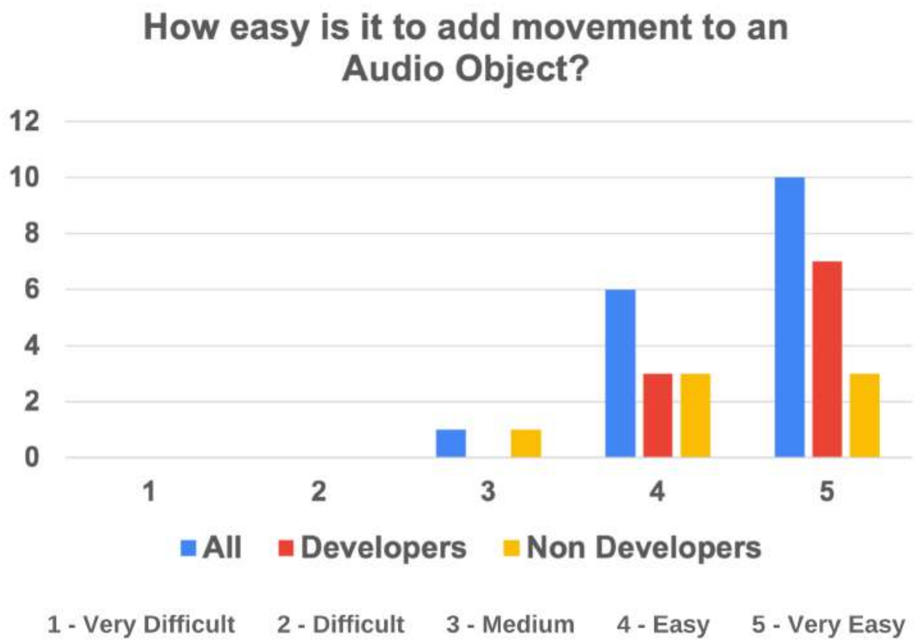


Figure 5.12: Results of users' evaluation on how easy it is to add movement to an Audio Object.

while 41% said yes. The Developer (40% Yes; 60% No) and Non-Developer (43% Yes; 57% No) results are very close to the overall results.

Overall, as presented in Figure 5.14, Task 5 was considered easy (47%) or very easy (53%) to complete by all participants. However, if we look into the Developer and Non-Developers results, we can see that while 80% of Developer participants considered it very easy to complete, only 14% of Non-Developers considered it very easy to complete. This shows that the task is not as easy for Non-Developers.

When asked what other aspects of the Audio Object's movement they would like to be able to edit, the Developer participants said that it should be possible to define vertical movement as well. Concerning the user's ability to move the audio object within the scene, two of the Developer participants considered that in up-view camera mode, it should be possible to move the audio objects using a drag-and-drop mechanism (a distinct mechanism from the "player picking up" mechanism that is used in player-camera mode).

Additionally, 43% of the Non-Developer participants also suggested that it should be possible to add vertical movement. Other Non-Developers suggested adding a mechanism to speed up the Audio Object's movement after passing by a certain Point or determine that the Audio Object stops for X seconds when it reaches a certain Point.

Task 6: Trigger Panel

All participants easily created a new audio object and edited it as they had previously done in Task 2 and 3, showing the quick learning support provided by the tool.

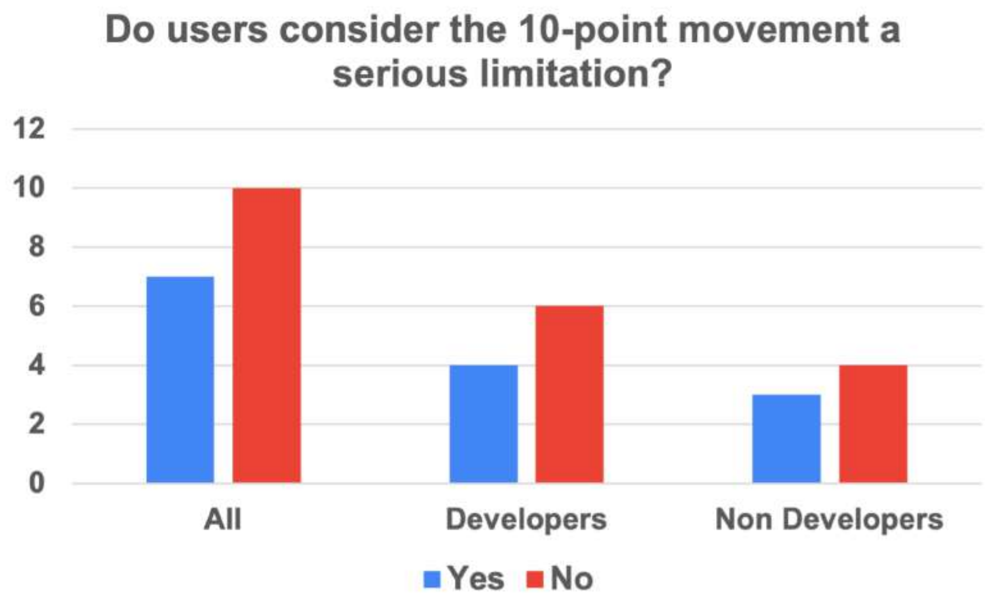


Figure 5.13: Results on whether or not users think that having a maximum of 10 points to define the Audio Object's movement is a serious limitation.

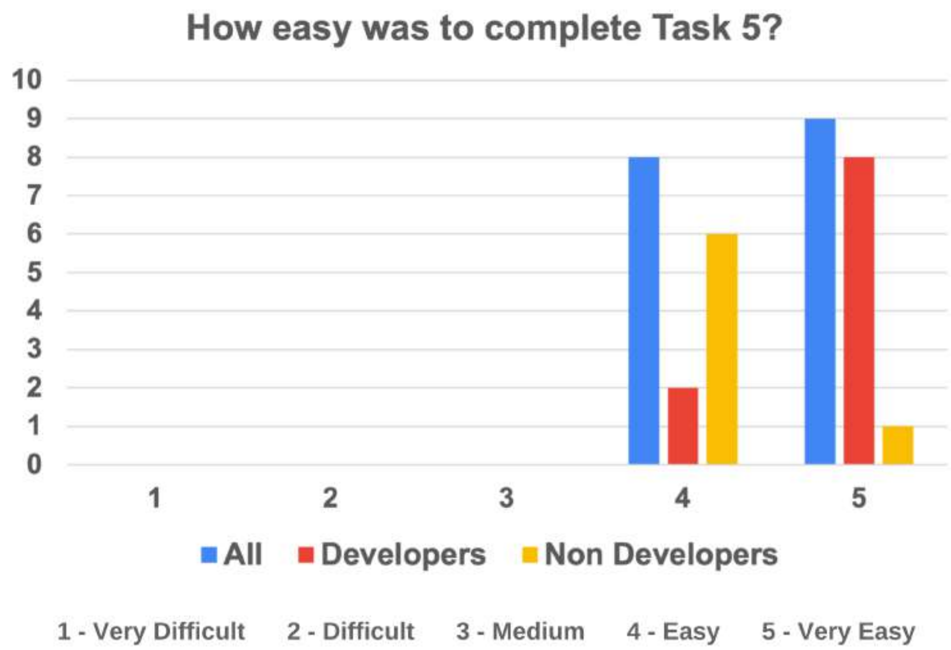


Figure 5.14: Distribution of participants by level of easiness at completing Task 5.

All participants were able to open the second Audio Object's trigger panel and define the trigger event by selecting a specific Point and by selecting the correct trigger event. After adding the trigger, all participants tested the result of their actions by playing the scene.

However, one of the participants informed that the result achieved, when reproducing the scene, was not the expected one. The participant explained that, although he experienced some spatial dimension of the sound, it was not as evident as he would have hoped. Considering the testing conditions (audio shared over Zoom), it was explained to the participant that once he completed Task 7, the researcher conducting the test would send the generated audio file from the scene to the participant. After Task 7 was completed, the participant experienced the audio generated by Immerscape locally on his device and stated to be satisfied with the result. We can conclude that the testing conditions were not ideal to listen to the immersive audio in full quality.

Figure 5.15: Distribution of participants by level of easiness at completing Task 6.

As can be inferred by the results presented in Figure 5.15, Task 6 was considered easy to complete by 35% of all participants and very easy to complete by 59% of all participants. Only one participant (Non-Developer) considered that it had a medium difficulty to complete. While 80% of the Developer participants considered it very easy to complete, only 29% of the Non-Developers participants considered it very easy to complete.

When queried about how easy they considered the mechanism of adding a trigger event to an Audio Object (results in Figure 5.16), the majority of all participants (76%) considered it very easy to do. However, while all Developers considered very easy to do,

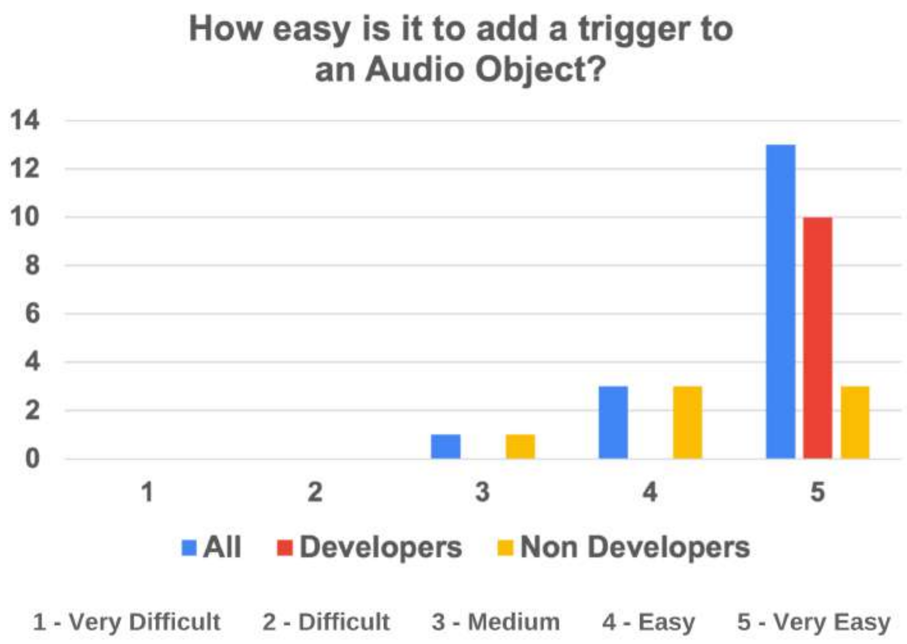


Figure 5.16: Results of users' evaluation on how easy it is to add a trigger event to an Audio Object.

the Non-Developer participants had divergent options. While 43% of Non-Developers considered adding a trigger event very easy to do, other 43% considered it simply easy and the remaining 14% considered that adding a trigger was of medium difficulty.

When questioned about what other trigger events they consider should be available to choose from in Immerscape, the participants had some suggestions. Developer participants suggested a "play on click" (already implemented), a "stop movement" event, a "start movement" event and triggers activated by timers. Besides also mentioning timed triggers, the Non-Developer participants suggested "triggers to change the Audio Object's movement" such as speeding up or slowing down its speed (as mentioned in the results of Task 5).

Task 7: Record Immersive Audio

Eventually, all participants were able to start the recording process. They were also all able to stop the recording process. All participants considered important to be able to listen to the scene's sound playing during the ongoing recording process.

Only one participant (Developer) struggled to understand how to record the scene, claiming that the microphone icon used for the recording button did not make it clear that it was meant to record the scene's audio and not the user's voice. When the participant was asked what symbol for the icon would make more sense to have instead of the microphone, a red square was suggested.

As Figure 5.17 shows, the recording process was considered very easy to accomplish by

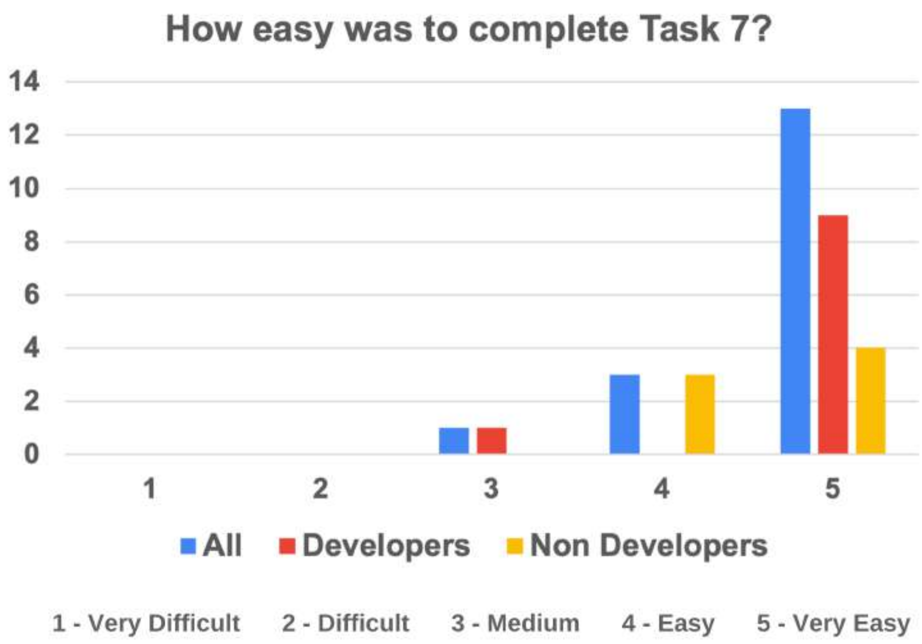


Figure 5.17: Distribution of participants by level of easiness at completing Task 7.

76% of all participants. Most Developer participants (90%) considered the task very easy. However, one Developer participant considered it of medium difficulty. This was the same participant that suggested an alternative icon for the recording button. Non-Developer participants considered the task either easy (47%) or very easy (57%) to accomplish.

Some Developer participants pointed out that should be some feedback about where the audio recorded is saved. A Non-Developer participant explained that they originally did not fully understand what "recording the audio" meant. As the previously mentioned Developer participant mentioned, they also thought that the recording button would be used to record new external audio that could be then added to the scene being composed.

Task 8: Save & Quit Project

All participants were able to save the progress made on the project they created. They all claimed that being able to save their project's progress was extremely important to them and a very helpful feature of the tool.

All participants were able to figure out how to exit the project's environment. One of the Developer participants pointed out that some explicit feedback should be provided, confirming that the project was saved before leaving the scene.

As can be inferred by the results shown in Figure 5.18, the majority of all participants (88%) considered Task 8 very easy to complete. All Developer participants as well as 91% of Non-Developers considered it very easy to complete.

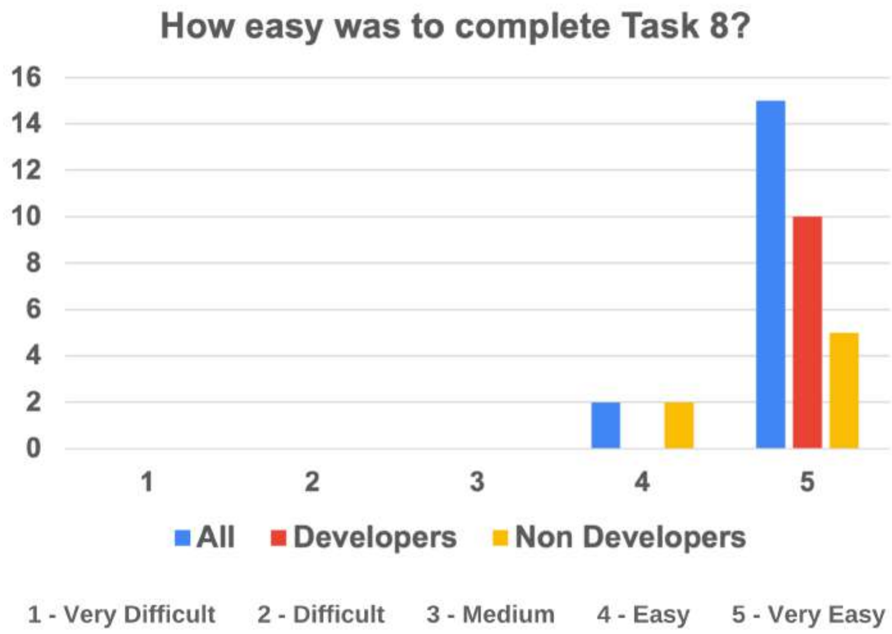


Figure 5.18: Distribution of participants by level of easiness at completing Task 8.

Task 9: Access Recorded Audio

All participants were successful in finding their own recordings and playing them. All considered that having access to the recordings within the tool was important to some degree. Figure 5.19 shows how important the participants considered being able to access the recordings through Immerscape. Overall, the participants considered it either important or very important.

Additionally, the participants were queried on how important they considered having the recordings made ordered by descending date of creation (results shown in Figure 5.20). All the Developer participants, with one exception, considered that having the recordings ordered by creation date was very important. The Non-Developer participants had more divided opinions about it. 43% of Non-Developer participants considered it somehow important, 14% considered it important and the remaining 43% considered it very important.

As can be inferred by the results presented in Figure 5.21, the vast majority of all participants (88%) considered Task 9 very easy to complete. Only two Non-Developer participants considered it simply easy.

Developer participants suggested that having more buttons to control the reproduction mechanism of the audio clips (restart, play, pause), including the duration of the recorded clips, as well as a button to delete the recordings would be good additional features.

Non-Developers suggested adding the total duration of the clips' length to the interface as well as add a button to download the audio clips. It was explained to the

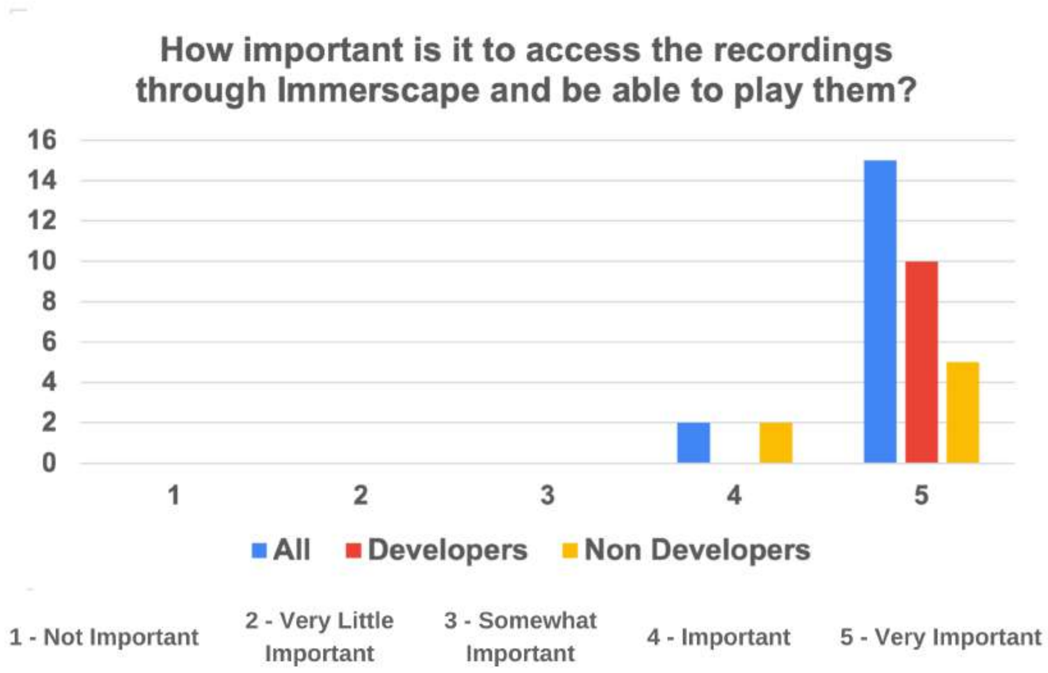


Figure 5.19: Results on how important users consider being able to access the recordings through Immerscape.

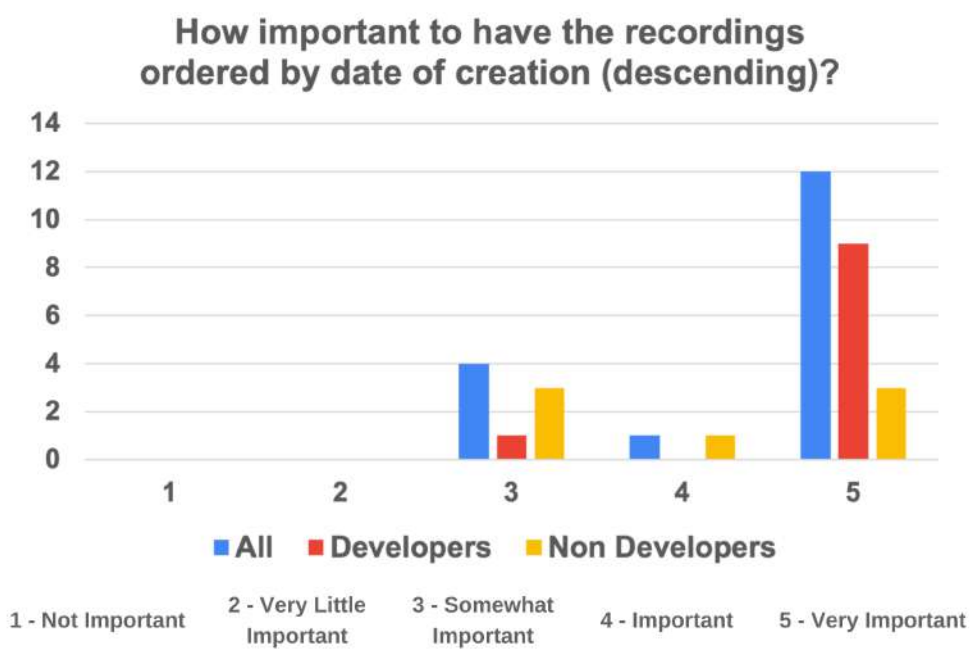


Figure 5.20: Results on how important users consider having the recordings ordered by creation date (descending).

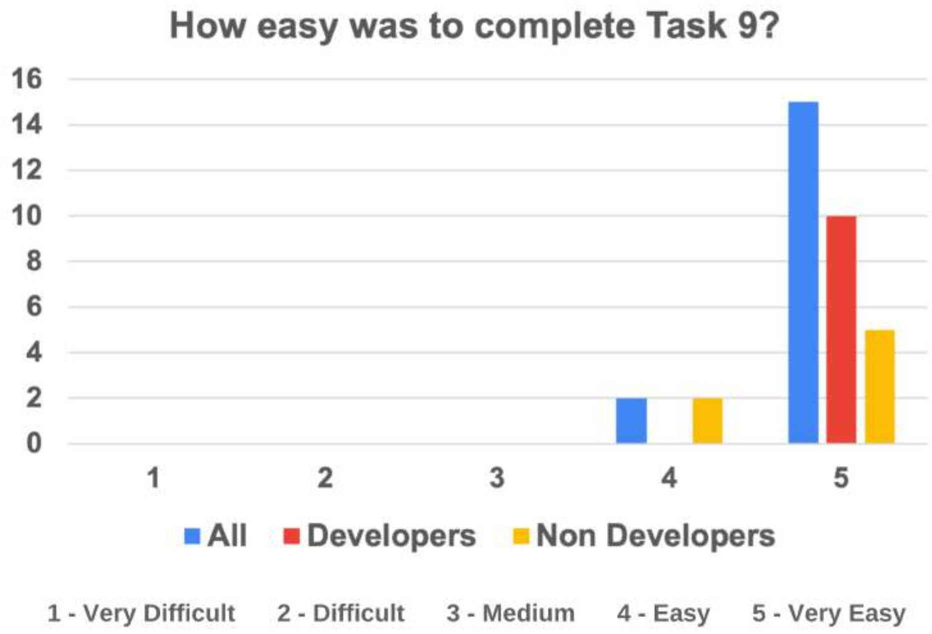


Figure 5.21: Distribution of participants by level of easiness at completing Task 9.

participant that suggested this that the audio clip is already saved in a folder and the participant argued that it could be useful to save it into a different folder.

Task 10: Access Saved Project

All participants considered to be extremely important to be able to reopen and edit the saved projects. When questioned about how important it was for them to have the saved projects listed by update date (from most recently update to least recently updated), 24% of all participants considered it important and 64% of all participants considered very important. Only 12% of all participants considered it somehow important (see Figure 5.22).

As shown in the results presented in Figure 5.23, with the exception of two Non-Developer participants, all the participants considered Task 10 very easy to complete. Those who did not, considered simply easy.

One Developer participant suggested that there should be other options for ordering the projects and two other Developer participants suggested that it should be possible to delete projects from Immerscape in the Saved Projects screen. The Non-Developers did not have any suggestions regarding the Saved Projects screen.

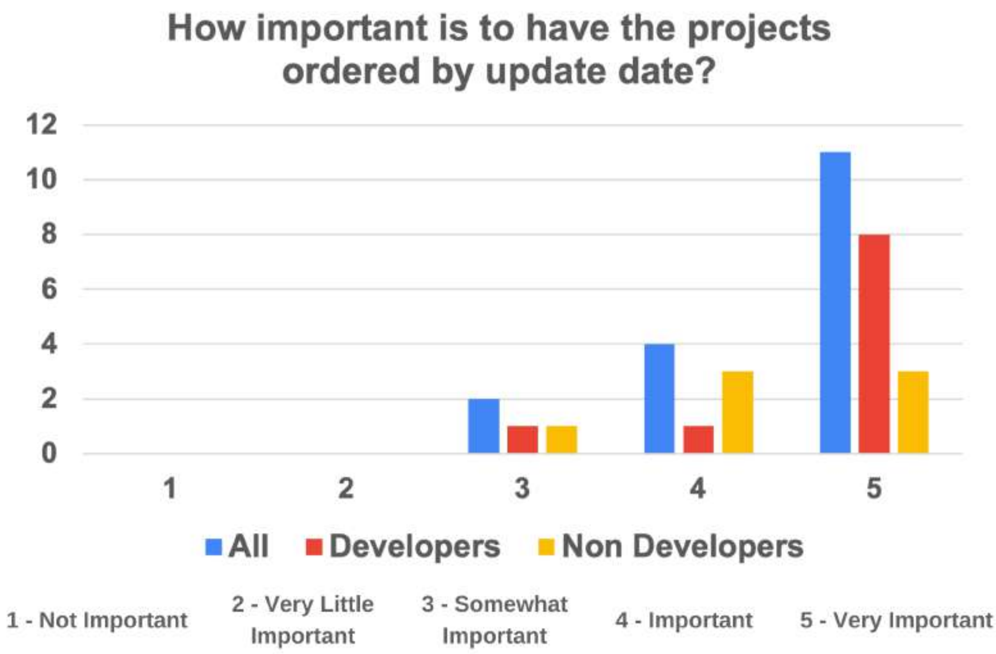


Figure 5.22: Results on how important users consider having the projects ordered by update date (descending).

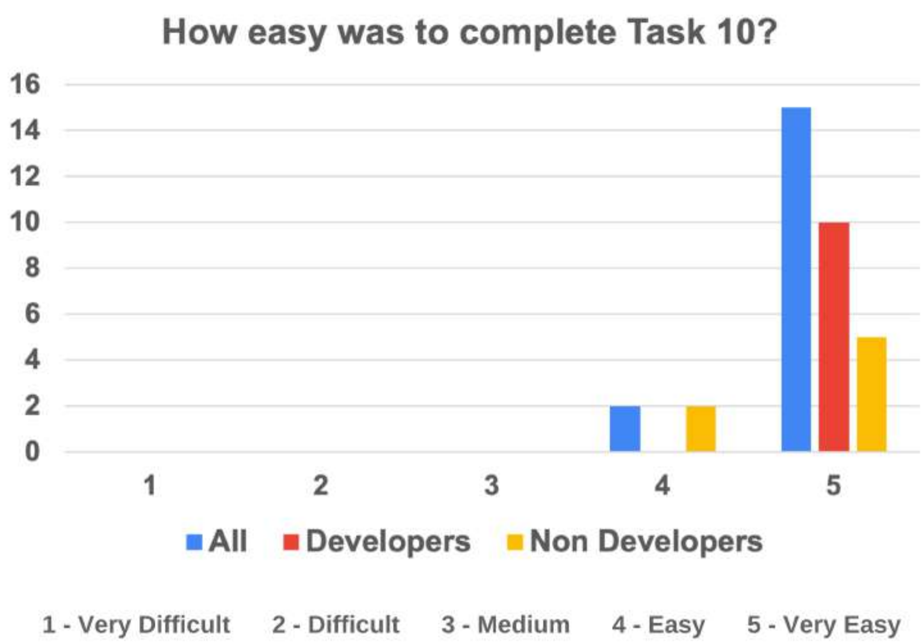


Figure 5.23: Distribution of participants by level of easiness at completing Task 10.

5.2.1 Results Overview

Upon completing all the tasks, the participants were given a form section to answer questions regarding the overall functioning of Immerscape and some additional space to global comments.

All participants reported being successful in completing the proposed tasks, by obtaining the recording of the composed auditory scene.

With the exception of one, the participants had never used a similar tool to generate immersive audio content. The participant that reported having used a similar application named MAX/MSP as the alternative. MAX/MSP is a visual programming language for music. The Non-Developer participant is a Musicology PhD student and he used it for a eight-channels composition.

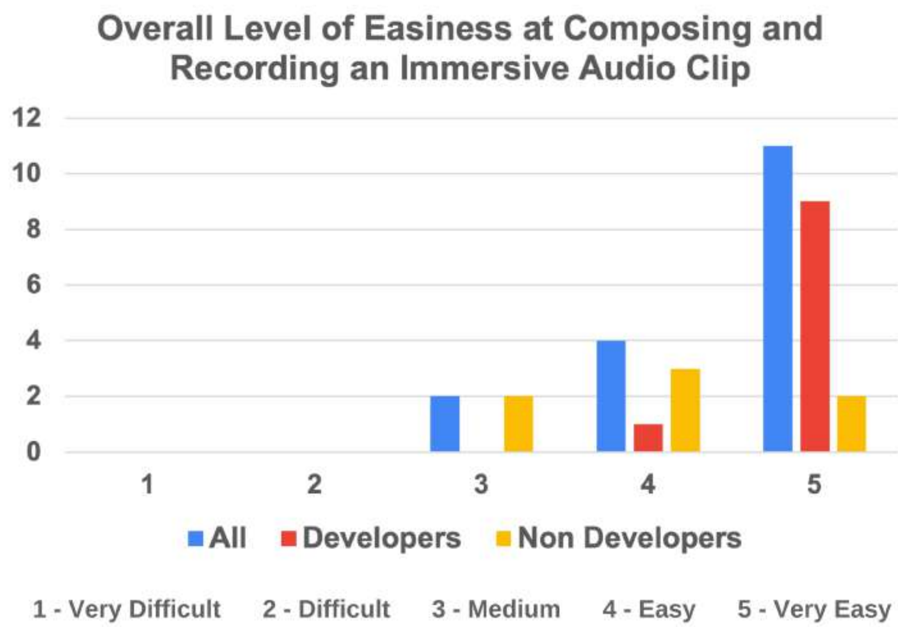


Figure 5.24: Overall level of easiness at composing and recording an immersive audio clip using Immerscape.

Globally, as presented in Figure 5.24, most participants considered it easy (24%) or very easy (64%) to create and record a soundscape using Immerscape. Only 12% of all participants considered that using Immerscape to generate immersive audio was somehow difficult (medium difficulty).

However, only looking into the Developer participants' results, it is observable that the large majority of Developer participants (90%) considered using Immerscape very easy, and 10%, simply easy.

On the other hand, the Non-Developer participants' results show that there were more divided opinions. Only 29% of all Non-Developer participants found Immerscape very easy to use; while 42% considered it easy and an additional 29% considered it of medium difficulty to use.

In order to further explore the different results between the two groups, we conducted a two-sample t-test. A two-sample t-test is a statistical test used to compare the means of two populations. Their results are used to determined if there is a significant difference between the mean of two samples that is unlikely to be due to sampling errors and random chance.

The t-tests are divided into two categories: paired and unpaired t-tests. For this particular case, an unpaired t-test was chosen, since it is used when comparing two independent or unrelated groups (Developers vs Non Developers). It is used to determine if there is a significant difference between the two. For each one of the ten tasks, we calculated the average easiness for both Developers and Non-Developers. The results of the unpaired t-test considered the difference between the two groups to be very statistically significant.

The contrast in the Developer vs Non-Developer results is noticeable in the different tasks' results and it can be explained by the differences in the participant groups. While the Developer participants present a narrow age range and similar education backgrounds, the Non-Developer participants represent more heterogeneous profiles. They represent a wider range of ages with different education backgrounds and different levels of experience working with software. It can be argued that the age and experience with simulated environments are the two biggest factors that influence the results. However, due to the small number of participants, we considered that further research should be done to determine that.

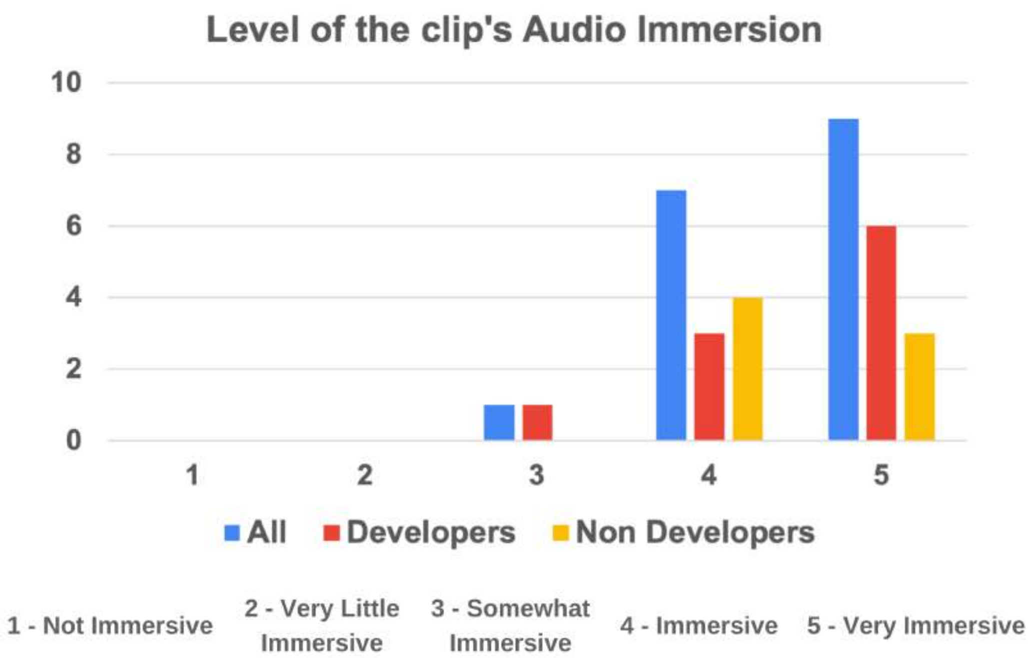


Figure 5.25: Evaluation of the generated audio clip's level of immersion.

Figure 5.26: Level of user satisfaction using Immerscape.

When questioned about the level of immersiveness and the quality of the audio clip generated by the tool, the results (presented in Figure 5.25) show, overall, a good or very good sense of immersion. According to the results, 53% of all participants consider the audio to have a very good sense of immersion, while 41% of all participants consider it good, and 6% of all participants consider it only somewhat immersive.

Additionally, the participants were queried on the level of satisfaction using Immerscape (results show in Figure 5.26). According to the results, 35% of all participants were satisfied using the tool's functionalities to create soundscape files whereas 65% of the participants claimed to be very satisfied while using the tool. 70% of all Developer participants reported being very satisfied while only 57% of all Non-Developer participants said the same.

When asked about how useful they considered Immerscape to help compose auditory scenes and generate immersive audio clips from them, the results (presented in Figure 5.27) show that the large majority of all participants (88%) recognizes Immerscape as very useful. While all Developers participants considered Immerscape very useful to generate immersive audio scenes, 29% of the Non-Developer participants considered it simply useful.

Functionalities Feedback

Overall, the Developer participants declared that the tool seems to have all of the essential functionalities and that its simplicity is appealing. They considered it easy to understand and use.

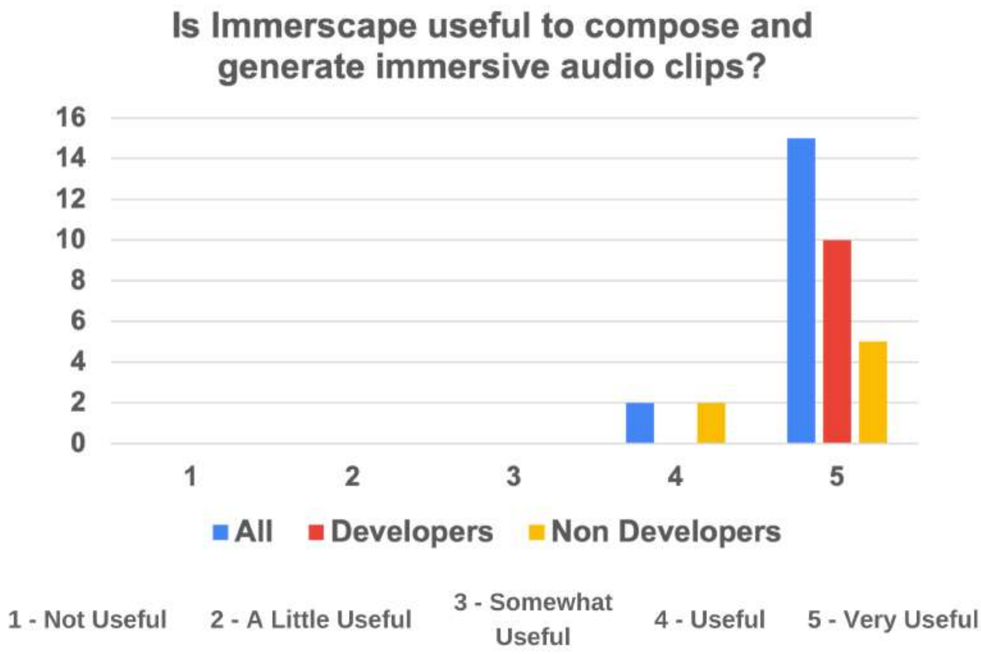


Figure 5.27: Evaluation of Immerscape usefulness to compose and generate immersive audio clips.

One of Non-Developers considered that they would need to experiment more with the tool and projects with more specific requisitions to point out missing functionalities. Another one suggested adding a microphone functionality to record external audio with Immerscape (create the clips that can be integrated into project with the tool). Besides that, Non-Developers considered the current functionalities to be enough to achieve the tool's goals.

Visual Component Feedback

Although the visual component was not the focus of this evaluation, Developer participants commented that it was satisfactory and intuitive. However, one of the Developer participants suggested that some cosmetic changes could make the audio object panels more aesthetically pleasing, as well as easier to read.

Non-Developers commented on how simple and functional the visual component of Immerscape is. One participant suggested that the Player Camera should be repositioned so the user could see the Player while navigating the environment, and another Non-Developer participant considered relevant to have a Portuguese version of the interface. Overall, Non-Developer participants were satisfied with the visual component, with the exception of the aforementioned improvement suggestions.

Audio Component Feedback

Developer participants commented on how good the sound quality was. They said it was a good immersive experience and considered that the parameters used in Immerscape to adjust the sound clips are the most common and sufficient.

One Non-Developer participant considered that according to the requirements of new scenes, new audio related functionalities should also be added to Immerscape. Some of these participants commented on how good the audio was. However, one of the Non-Developer participant wished the audio had been reproduced louder during the test. It is important to note that the audio reproduced over Zoom is lower than when reproduced locally.

Testing Limitations

After conducting the study with the 17 participants (10 developers; 7 non-developers) over Zoom, we became aware of some limitations of the testing process. Although the user tests over Zoom were the best compromise we found to facilitate Immerscape's evaluation during an ongoing pandemic, the process presented multiple challenges. Since some participants only had one computer screen, the process of switching between the control-remote Zoom window, the testing form, and the instructions guide, was a struggle at times. Additionally, some connection problems were observed by the participants, as well as a decrease in the volume of the audio reproduced over Zoom.

In addition to the technical difficulties experienced, other factors concerning the testing process could be improved. The length of the test was not ideal. Afterwards, we realized that it would have to been beneficial to somehow reduce the extension of the questions asked. The Developer group had a very narrow age range and, although the Non-Developer group had a wider range of ages, there were still large gaps between some of the participants ages (e.g. there are no participants with ages between 28 and 54). The small number of testing subjects in each group makes it hard to draw solid conclusions. For example, the 54 years old participant behaves as an outlier in some tasks. However, it is not clear if this is due to the participant's age or their level of experience with software, in particular, simulation or 3D environments.

5.3 Summary

This chapter presented and discussed the results and user feedback obtained during the evaluation phase of Immerscape's first working prototype.

It started by introducing the user study performed to validate the system's functionalities, assess the quality of the audio content generated, its immersiveness, as well as determine the ease of use and its acceptance by users. The study, conducted with a testing pool of 17 volunteers (divided into two groups, Developers and Non-Developers), consisted of the completion of 10 small related tasks.

The user study's protocol was presented (section 5.1) and the test's tasks were discussed individually (subsection 5.1.2), while explaining why each one of them was relevant to the evaluation process. Overall, this study aimed at answering the last research question of this dissertation, determining if Immerscape can help non-expert users create immersive audio representations of historical soundscapes.

The results gathered from the user study were then presented (section 5.2) and discussed, task by task. All participants completed the ten proposed tasks successfully. From the tests originated, for each participant, an audio project and the corresponding recording of its acoustic environment. The tasks' results showed that the tasks had different difficulty levels and that the two groups of participants revealed experiencing different difficulty levels while performing the same task.

Generally, Developer participants had an easier time using Immerscape than Non-Developer users. This could be explained by differences in their profiles. While the Developer participants present a narrow age range and similar education backgrounds, the Non-Developer participants represent more heterogeneous profiles. They represent a wider range of ages with different education backgrounds and different levels of experience working with software, especially simulation environments. It can be argued that the participants' age and their experience with simulated environments are the two biggest factors that influence the results. However, and due to the small number of participants, we considered that further research should be done to confirm that.

At last, we gave an overview of the results (subsection 5.2.1) of the system's first testing phase and we discussed some of the testing process limitations. The global results showed that, overall, users considered Immerscape easy to use. The users determined that the audio content generated had a good level of audio immersion and that Immerscape is very useful to compose auditory scenes and generate immersive audio files from them. In addition, all participants reported high levels of satisfaction while using Immerscape.

Despite the good results, the feedback from the user study showed that there are opportunities to improve Immerscape. The study also revealed some of the remote testing process's limitations, such as connection problems or the decreased audio quality in the reproduction over Zoom. This indicates that following user studies should be done in person and, preferably, with a larger numbers of participants.

6 Conclusion

This chapter overviews the developed work during the dissertation, discussing the main results of the proposed solution, its strengths and its weaknesses. It also briefly introduces the future work that should be done to improve the current version of the proposed solution.

Soundscapes are representations of acoustic environments considering a specific context and time. They evolve over time according to changes in social, cultural and political contexts, and constitute a unique way of experiencing auditory memories of a specific period in history. Considering that the soundscape of a place is ever-changing and it can provide us with important information about a location's context in time, the study and preservation of soundscapes, in particular historic soundscapes, has gained relevance in the cultural heritage protection and promotion process.

A good example of the aforementioned, is the PASEV project, whose main purpose is to preserve and promote Évora's rich cultural heritage, in particular, its historic soundscapes corresponding to the period between 1540 and 1910. For this purpose, the PASEV team aimed at providing audio content, related to the soundscapes of that period, to visitors, through all its interactive interfaces. However, making historical soundscapes available to the general public was a challenge, since there are no original audio captures and it is impossible to go back in time and record the acoustic experiences from centuries ago. For that reason, the closest we can get to experience an historic soundscape is by trying to recreate it. This can be achieved with the support of historic documentation and audio immersion technology.

The research on examples of applications that implement different audio immersion techniques was important to reflect on the current state of the art. Although some solutions to create immersive soundscapes are currently available, in this dissertation, we propose a software tool that can easily be used by non audio professionals to create immersive compositions of audio files, with the aim of providing realistic audio reenactments of historical events. The tool can easily be used by inexperienced users, from different backgrounds, to interactively create immersive soundscapes.

The proposed tool, Immerscape, emerges as a solution to answer the main research questions of this dissertation, concern with how we can create immersive audio to represent a historical soundscapes and how we can make that process of creation easily accessible to people with no background in sound mixing or engineering. Our solution confirms that we can design and develop software that helps non-expert users to create immersive audio representations of historical soundscapes.

Immerscape focuses on giving the users a set of simple-to-use functionalities so they can create spatially-aware immersive auditory experiences, using monaural or stereo audio clips. The audio clips can be uploaded to one of the available environments, adjusted and edited to compose an imaginary scene through sound. The immersive audio files generated are obtained by combining different current sound recordings, and by composing an immersive auditory scene that can be recorded by the user using the tool. The proposed technique gives the audio recording a spatial dimension, due to the use of HRTFs on each one of the audio clips uploaded into the 3D environment.

Looking to answer the last research question, the work was evaluated through a user study that focused on the tool's functionality and acceptance by users. The study showed that the tool is accessible to inexperienced users; it is easy to use and the audio produced in the process has a reported satisfactory quality, providing the desired audio immersion to the clips, when playing the composed scene through headphones. The results also show that, although the tool performed well with the users, and had overall a good acceptance, there is room for continuous improvement (e.g. by making the player's capabilities more intuitive). There were also some recognizable differences between Developer and non-Developer users. However, with the current data, it was not possible to determine which of the participants' profiles' characteristics is responsible for the different results.

The evaluation process had some recognizable limitations, such as the small number of participants per group and lack of diversity in the participants' groups profiles (e.g. the narrow age range or lack of older participants). Another influential factor in the testing process was the quality of audio reproduction and the remote control connection over Zoom.

In spite of the limitations of the presented user study, its results confirm that the tool's current version includes all the essential working functionalities and that users considered it useful to create immersive audio clips. However, the user study also pointed out some of the functionalities that can be expanded or improved to offer the users a wider range of more convenient options when composing their scenes.

Elaboration of a Master's dissertation during the COVID-19 Pandemic

It is impossible to ignore the impact COVID-19 had, and still has, to this day, in the world, and how it forced us to find new ways to work. This dissertation was entirely developed while the world was still dealing with a pandemic. Undeniably, the pandemic affected all our personal lives, our routines, and the way we work. It also made us reconsider our habits, our priorities and shifted the way most businesses operated.

As aforementioned in the exposure of this dissertation's motivation, one of the additional benefits of the approach presented is that the immersive audio generated can be experienced independently of the user's location. The audio can be provided through digital platforms and experienced remotely. During the COVID-19 pandemic, restrictions were implemented on cultural sites, museums, and many other places. At home, the users are able to use their own devices to connect with the outside world when that contact is limited. The immersive audio generated, paired with visual components, can give users an opportunity to remotely explore audiovisual content related to the sites which, for one reason or another, they cannot visit.

Fortunately, I was able to complete all my work in the safety of my home. These were not the ideal conditions, but it showed everyone involved the importance of adaptation and the new forms of communication that current technologies allow. The evaluation process was conducted remotely, not only for safety reasons, but also for the convenience of the participants. However, after the remote testing experience, and the limitations detected during the process, we determined that it would be best if the following evaluations could be done in person, even if it is logistically harder to organize and it requires some cautionary measures.

6.1 Future Work

While developing the system and during user testing sessions some ideas for future research directions and overall improvements to Immerscape emerged or were suggested by the participants. To plan the future work, we took the feedback provided by the study's participants and the observed interaction notes, to redesign some aspects of the tool and implement some of the suggestions given. The process of empathizing with the participants' evaluation, and the analysis of the first testing phase results restarts the iterative development process (more about in subsection 2.6.3).

Taking into consideration the first iteration's feedback, some goals for the following iteration, focused on improving the current prototype, were defined. Some of the improvements that should be addressed in following iterations are:

- **Project Creation:** Add more generic environments. The most common suggestions by the participants were a stadium environment and a theater room environment. Ideally, these environment should be designed and produced in collaboration with 3D modelling experts and acoustic designers/sound engineers.

- **Environment Navigation:** The most common suggestion was to add the jumping ability. It would have to be determined the best key to trigger the jumping behavior in the player.

- **Camera Switch:** While in the Up-view mode, moving the Audio Objects should use a different approach than the currently implemented. While currently the movement of objects is rooted and controlled by the Player, participants in the user study suggested that in the Up-view mode it should be independent from the Player. Participants suggested a third person drag-and-drop mechanism to move the elements while in the Up-view camera mode.

- **Create Audio Object:** It should be possible to delete any Audio Object.
- **Point:** It should be possible to move or delete any Point. Similarly to what is already available for the Audio Object, it should also be possible to determine the vertical position of Point.
- **Edit Audio Object:** The participants indicated there should be a way to distinguish the Audio Objects, and so a mechanism to edit the Audio Object's color should be implemented. On hover, it should also be shown a label with the Audio Object's identifier so the user can determine which one it is without having to open one of its panels.
- **Audio Object Movement:** Concerning the Audio Object's movement, the participants considered that vertical movement should be possible (using the vertical Points previous mentioned).
- **Audio Object Triggers:** Concerning the Audio Object's triggers, and taking into consideration the types of triggers users have suggested, we determined that the Trigger Panel should suffer a redesign process. Instead of one section with all the available triggers, the panel should have a carousel that shows the available triggers according to the categories: "Activated on Entry", "Activated on Click", "Timed Triggers" and "Movement Triggers".
- **Play the Scene's Audio:** Users suggested the addition of a keyboard shortcut to play the audio, as well as be able to stay in the Up-view mode while the audio plays (not switch into the Player mode automatically). It might be useful to keep both options available (let the user choose if it prefers the automatic switch or not). This could be a topic subject to research on the following iteration.
- **Record Audio:** Similarly to what was discussed for the Play mechanism, it might be useful to have a keyboard shortcut to open the recording window and not have the automatic switch into the Player mode when the recording process begins. In addition, there should be some feedback telling the users which folder the audio file was saved into at the end of the recording process. In the Recordings screen, the participants claimed that they should be able to delete and download audios.
- **Save Project:** It might be interesting to implement an automatic saving feature to the save system. A participant also pointed out that when the user chooses to quit without saving, there is a dialogue missing to double check if the user really intends

to leave without saving the progress. Additionally, in the Saved Projects screen, there should be an option to delete saved projects.

- **Help:** While implementing the new features and redesigning some of the aforementioned mechanisms, the help information should also be updated.

Once the changes that take into consideration the initial user feedback are implemented, a new evaluation study should be done. The new testing process should aim to provide a more complete evaluation of the tool's functionalities and its usability. It should target a different range of user profiles, including a wider range of older users and users with different level of experience with simulation/game environments.

The evaluation process described in chapter 5 should also be modified. The test itself should be designed to be shorter but more focused on specific tasks. Ideally, the user study should be performed in person to eliminate any of the connection and Zoom limitations.

Bibliography

[1] "2D spatial audio in a molecular navigator/editor for blind and visually impaired users". English. In: *DAFx 2017 - Proceedings of the 20th International Conference on Digital Audio Effects*. University of Edinburgh, Sept. 2017, pp. 360–366 (cit. on pp. 32, 33).

[2] AcousticNature. *What is a Soundscape? (Definition and Science of Hearing)*. 2020. URL: https://acousticnature.com/journal/what-is-a-soundscape (visited on 01/22/2021) (cit. on p. 8).

[3] G. D. A. Alan Dix Janet Finlay and R. Beale. *Human-Computer Interaction*. 3rd ed. Pearson Education Limited, 2004 (cit. on pp. 23, 25, 26, 48).

[4] Apple. *Overview of binaural panning in Logic Pro*. URL: https://support.apple.com/pt-pt/guide/logicpro/lgcp8e81655e/mac (visited on 01/26/2021) (cit. on p. 20).

[5] C. Ardito et al. "Enriching Archaeological Parks with Contextual Sounds and Mobile Technology". In: *ACM Trans. Comput.-Hum. Interact.* 19.4 (Dec. 2012). ISSN: 1073-0516. DOI: 10.1145/2395131.2395136. URL: https://doi.org/10.1145/2395131.2395136 (cit. on p. 2).

[6] L. Barclay. "Augmenting Urban Space with Environmental Soundscapes and Mobile Technologies". In: *Soundscape: The Journal of Acoustic Ecology* 16 (2017). URL: https://research-repository.griffith.edu.au/handle/10072/374289 (cit. on p. 2).

[7] K. Brandenburg, S. Brix, and T. Sporer. "Wave field synthesis: New possibilities for large-scale immersive sound reinforcement". In: *Proceedings of the 18th International Congress on Acoustics* 1 (2004). URL: https://www.icacommission.org/Proceedings/ICA2004Kyoto/pdf/Mo5.E.1.pdf (cit. on p. 21).

[8] Britannica. *Psychophysics*. URL: https://www.britannica.com/science/psychophysics (visited on 01/22/2021) (cit. on p. 12).

[9] U. of California. *HRTF Data - The CIPIC Interface Laboratory Home Page*. URL: https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/ (visited on 02/13/2021) (cit. on p. 19).

[10] M. Casu, M. Koutsomichalis, and A. Valle. "Imaginary Soundscapes: The SoDA Project". In: *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*. AM '14. Aalborg, Denmark: Association for Computing Machinery, 2014. ISBN: 9781450330329. DOI: 10.1145/2636879.2636885. URL: https://doi.org/10.1145/2636879.2636885 (cit. on p. 35).

[11] COGNITUS. *Immersive Audio*. URL: http://cognitus-h2020.eu/index.php/immersive-audio/ (visited on 01/30/2021) (cit. on p. 20).

[12] B. Costerto. *WA systematic review of the most appropriate methods of achieving spatially enhanced audio for headphone use*. Ed. by Pinecone Research Labs. May 2013. URL: https://pineconeresearchlabs.wordpress.com/2013/05/01/pr01/ (cit. on p. 16).

[13] Dolby. *Dolby Atmos - Dolby*. retrieved Oct/2021. 2021. URL: https://www.dolby.com/technologies/dolby-atmos/ (visited on 10/20/2021) (cit. on p. 35).

[14] A. Ferreira. "Extending the Auditory Atlas of the City of Évora". MA thesis. NOVA School of Science and Technology, NOVA University Lisbon, 2021 (cit. on pp. 4, 36–38).

[15] N. Fonseca. "3D particle systems for audio applications". In: *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx-13). Maynooth, Ireland: National University of Ireland*. 2013 (cit. on p. 35).

[16] I. D. Foundation. *Augmented Reality*. URL: https://www.interaction-design.org/literature/topics/augmented-reality (visited on 01/28/2021) (cit. on p. 22).

[17] M. Gerzon and A. at Oxford. *What is Ambisonics?* URL: https://intothesoundfield.music.ox.ac.uk/what-is-ambisonics (visited on 01/26/2021) (cit. on p. 21).

[18] D. Gomes. "Immersive Audio for User-Generated Content". MA thesis. NOVA School of Science and Technology, NOVA University Lisbon, 2018 (cit. on p. 30).

[19] Google. *Augmented Reality*. URL: https://arvr.google.com/ar/ (visited on 01/28/2021) (cit. on p. 22).

[20] *Image adapted from brainHQ*. URL: https://www.brainhq.com/brain-resources/cool-brain-facts-myths/how-hearing-works/ (cit. on p. 9).

[21] *Image adapted from Wikipedia Commons*. URL: https://upload.wikimedia.org/wikipedia/commons/9/9a/Sine_wave_amplitude_wavelength.svg (cit. on p. 10).

[22] *Image adapted from Wikipedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Phase_shift.svg (cit. on p. 12).

[23] *Image from Concertgebouw Brugge Website.* URL: https://www.concertgebouw.be/en/club-surround-wave-field-synthesis-1 (cit. on p. 22).

[24] IRCAM. *LISTEN HRTF DATABASE.* URL: http://recherche.ircam.fr/equipes/salles/listen/ (visited on 02/13/2021) (cit. on p. 19).

[25] Y. R. Jenny Preece Helen Sharp. *Interaction Design: Beyond Human-Computer Interaction.* 4th ed. Wiley, 2015 (cit. on pp. 23, 25).

[26] J. S. Johannes Goebel Todd Vos and A. Ascani. *Wave Field Synthesis.* URL: https://empac.rpi.edu/program/research/wave-field-synthesis (visited on 01/27/2021) (cit. on p. 21).

[27] K. Jorgensen. "Left in the dark: playing computer games with the sound turned off". In: *Pac-Man to pop music: Interactive audio in games and new media.* University of Waterloo, Canada: Ashgate Publishing Group, 2008, pp. 163–176. ISBN: 978-0-7546-6211-2 (cit. on pp. 2, 3).

[28] F. Kaghat et al. "A new audio augmented reality interaction and adaptation model for museum visits". In: *Computers & Electrical Engineering* 84 (June 2020), p. 106606. DOI: 10.1016/j.compeleceng.2020.106606 (cit. on p. 34).

[29] G. Kearney. *Binaural audio for virtual and augmented reality.* URL: https://www.york.ac.uk/research/impact/binaural-sound/ (visited on 01/28/2021) (cit. on pp. 22, 43).

[30] L-Acoustics. *ISA Studio - L-Acoustics.* retrieved Sept/2021. Oct. 2021. URL: https://www.l-acoustics.com/products/l-isa-studio/# (cit. on p. 35).

[31] M. M. Lab. *HRTF Measurements of a KEMAR Dummy-Head Microphone.* URL: https://sound.media.mit.edu/resources/KEMAR.html (visited on 02/13/2021) (cit. on p. 19).

[32] I. Menezes. "Uma Estrutura para gerir plataformas do domínio das Paisagens Sonoras Históricas". MA thesis. NOVA School of Science and Technology, NOVA University Lisbon, 2021 (cit. on pp. 4, 38).

[33] meSch. *meSch - About.* URL: https://www.mesch-project.eu/about/ (visited on 02/10/2021) (cit. on p. 31).

[34] T. Miyashita et al. "An Augmented Reality museum guide". In: *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality.* 2008, pp. 103–106. DOI: 10.1109/ISMAR.2008.4637334 (cit. on p. 34).

[35] N. Montan. *AAR - An Audio Augmented Reality System.* Department of Microeletronics and Information Technology, KHT, Royal Institute of Technology, Stockholm, 2002. URL: https://people.kth.se/~maguire/DEGREE-PROJECT-REPORTS/030409-Nils-Montan.pdf (cit. on pp. 33, 34).

[36] *Neumann KU 100 Dummy Head Mic Image.* URL: https://www.proaudioeurope.com/microphones/stereo-mics/neumann-ku-100-dummy-head-binaural-stereo-microphone.html (cit. on p. 16).

[37] J. Nielsen. *10 Usability Heuristics for User Interface Design.* URL: https://www.nngroup.com/articles/ten-usability-heuristics/ (visited on 09/24/2021) (cit. on p. 25).

[38] J. K. Nigel Bevana and J. Maissela. "What is Usability?" In: *Proceedings of the 4th International Conference on HCI.* Stuttgart: National Physical Laboratory and University College Cork, 1991 (cit. on p. 24).

[39] OpenStax. *Speed of Sound, Frequency, and Wavelength.* URL: https://openstax.org/books/college-physics/pages/17-2-speed-of-sound-frequency-and-wavelength (visited on 02/22/2021) (cit. on p. 10).

[40] S. Particles. *Overview - Sound Particles.* retrieved Oct/2021. 2021. URL: https://www.soundparticles.com/products/soundparticles/overview (visited on 10/01/2021) (cit. on p. 35).

[41] PASEV. *Historical Soundscapes Meeting Évora 2019 - PASEV.* URL: https://pasev.hcommons.org/project/ (visited on 01/21/2021) (cit. on p. 9).

[42] PASEV. *Project - PASEV.* URL: https://pasev.hcommons.org/project/ (visited on 01/21/2021) (cit. on pp. 3, 4).

[43] D. Petrelli. *meSch - How does the form of the interactive object affect the visiting experience?* 2019. URL: https://www.mesch-project.eu/how-does-the-form-of-the-interactive-object-affect-the-visiting-experience/ (visited on 02/10/2021) (cit. on p. 31).

[44] B. Ponte. "Exploring soundscapes and physical spaces with a mobile augmented-reality application". MA thesis. NOVA School of Science and Technology, NOVA University Lisbon, 2021 (cit. on pp. 4, 38).

[45] V. P. C. Project. *Experience | Virtual Paul's Cross Project.* URL: https://vpcp.chass.ncsu.edu/experience/ (visited on 01/20/2021) (cit. on p. 29).

[46] M. Risoud et al. "Sound source localization". In: *European Annals of Otorhinolaryngology, Head and Neck Diseases* 135.4 (2018), pp. 259–264. ISSN: 1879-7296. DOI: https://doi.org/10.1016/j.anorl.2018.04.009. URL: https://www.sciencedirect.com/science/article/pii/S187972961830067X (cit. on p. 17).

[47] A. Rodrigues and N. Correia. "Using Technology in Digital Humanities for Learning and Knowledge Dissemination". In: *Revista EducaOnline* 15.2 (2021), pp. 27–44 (cit. on p. 9).

[48] J. Rosário, A. Rodrigues, and N. Correia. "A responsive platform for the Auditory Atlas of Évora". In: *Paisagens sonoras históricas : Anatomia dos sons nas cidades. Évora : Publicações do Cidehus.* (2021) (cit. on p. 9).

[49] J. Rosário. “Uma plataforma responsiva para o Atlas Auditivo de Évora”. MA thesis. NOVA School of Science and Technology, NOVA University Lisbon, 2019 (cit. on pp. 4, 38).

[50] F. Rumsey. *Spatial Audio*. Focal Press, 2001. ISBN: 0-240-51623-0 (cit. on pp. 8, 14–16, 18–21).

[51] K. Salo, M. Bauters, and T. Mikkonen. “User Generated Soundscapes Activating Museum Visitors”. In: *Proceedings of the Symposium on Applied Computing*. SAC ’17. Marrakech, Morocco: Association for Computing Machinery, 2017, pp. 220–227. ISBN: 9781450344869. DOI: 10.1145/3019612.3019691. URL: https://doi.org/10.1145/3019612.3019691 (cit. on p. 35).

[52] A. Saxena and A. Y. Ng. “Learning sound location from a single microphone”. In: *2009 IEEE International Conference on Robotics and Automation*. 2009, pp. 1737–1742. DOI: 10.1109/ROBOT.2009.5152861 (cit. on p. 14).

[53] R. M. Schafer. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books, 1994 (cit. on p. 8).

[54] B. Shneiderman et al. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 6th ed. Pearson, 2016 (cit. on pp. 24–26).

[55] A. Singh. *Image adapted from ResearchGate*. URL: https://www.researchgate.net/figure/Pressure-propagation-of-a-sound-wave-through-air_fig2_334784649 (cit. on p. 13).

[56] N. G. Society. *Hearing With Your Brain | National Geographic Society*. URL: https://www.nationalgeographic.org/media/hearing-your-brain/ (visited on 02/19/2021) (cit. on p. 2).

[57] M. Southworth. “The Sonic Environment of Cities”. In: *Environment and Behavior* 1.1 (1969), pp. 49–70 (cit. on p. 8).

[58] Statista. *Number of smartphone users worldwide from 2016 to 2021(in billions)*. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ (visited on 01/28/2021) (cit. on p. 2).

[59] A. R. Studios. *Spatial Audio*. URL: https://www.abbeyroad.com/spatial-audio/ (visited on 01/30/2021) (cit. on p. 20).

[60] S. S. Studios. *Mixing: Panning*. URL: https://signaturesound.com/mixing-panning/ (visited on 01/27/2021) (cit. on p. 20).

[61] L. Sun. *Image adapted from ResearchGate*. URL: https://www.researchgate.net/figure/Interaural-Time-Difference-ITD-and-Interaural-Level-Difference-ILD_fig1_276472968 (cit. on p. 18).

[62] V. Sundareswaran et al. "3D audio augmented reality: implementation and experiments". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* 2003, pp. 296–297. DOI: 10.1109/ISMAR.2003.1240728 (cit. on p. 34).

[63] U. Technologies. *Unity Real-Time Development Platform | 3D, 2D VR & AR Engine.* URL: https://unity.com/ (visited on 02/10/2021) (cit. on p. 42).

[64] UNESCO. *Historic Centre of Évora.* URL: https://whc.unesco.org/en/list/361/ (visited on 01/19/2021) (cit. on p. 1).

[65] M. Vecco. "A definition of cultural heritage: From the tangible to the intangible". In: *Journal of Cultural Heritage* 11.3 (2010), pp. 321–324. ISSN: 1296-2074. DOI: https://doi.org/10.1016/j.culher.2010.01.006. URL: http://www.sciencedirect.com/science/article/pii/S1296207410000361 (cit. on pp. 1, 9).

[66] J. Vilar. "Sistema para Criação e Exploração de Narrativas Interativas e Jogos". MA thesis. NOVA School of Science and Technology, NOVA University Lisbon, 2022 (cit. on pp. 4, 38).

[67] J. Wall. *Recovering Lost Acoustic Spaces: St. Paul's Cathedral and Paul's Churchyard in 1622.* 2014. URL: https://www.digitalstudies.org/articles/10.16995/dscn.58/ (visited on 01/20/2021) (cit. on pp. 5, 8, 28, 29).

[68] U. of York. *SADIE | Spatial Audio For Domestic Interactive Entertainment.* URL: https://www.york.ac.uk/sadie-project/about.html (visited on 02/10/2021) (cit. on pp. 19, 43).

[69] W. A. Yost. *Fundamentals of Hearing - An Introduction.* Fourth. Academic Press, 2000. ISBN: 0-12-775695-7 (cit. on pp. 8–11, 13, 14, 16–19, 21, 46).

A

Storyboards

The written document with information provided by the historians part of the PASEV team was translated into storyboards that illustrate the scenes for better understanding. The storyboards, created using Canva, aimed to summarize all the information in a visual form and support the composition of the scenes.

PASEV STORYBOARDS

HISTORICAL SCENES RECREATION STORYBOARDS

SCENE 1 THE KING'S ACLAMATION CEREMONY

SCENE 2 THE KING'S FUNERAL CEREMONY

SCENE 3 BANDSTAND INAUGURATION

SCENE 4 THE FIRST OF DECEMBER

ÉVORA HISTORICAL SOUNDSCAPE

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

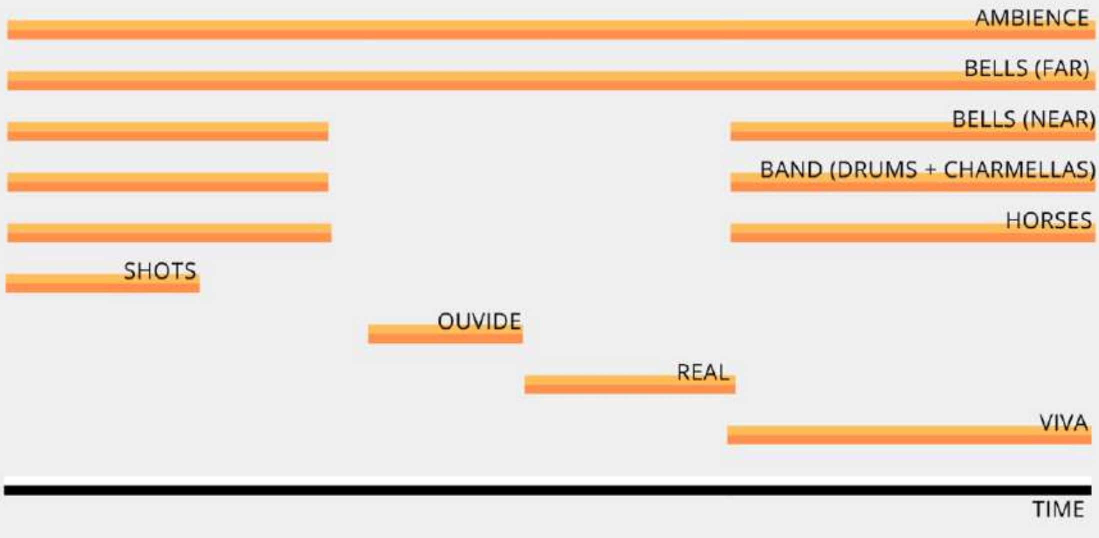
SCENE 1 THE KING'S ACLAMATION CEREMONY

ORIGINAL TITLE CERIMÓNIA DA ACLAMAÇÃO DO REI D. AFONSO VI (AUTO DO LEVANTAMENTO)

DATE 23 NOV, 1656

SOUNDS DRUMS, CHARMELLAS, HORSES, BELLS, SHOTS, VOICES

DESCRIPTION FESTIVE PARADE THROUGHOUT THE CITY OF ÉVORA; INCLUDED A BAND, THE CITY OFFICIALS RIDING ON HORSES; BELLS ECHOING FROM ALL THE CHURCHES; SHOTS FIRED INTO THE AIR; AND A LARGE CROWD OF PEOPLE THAT GATHERED TO CELEBRATE THE NEW KING.



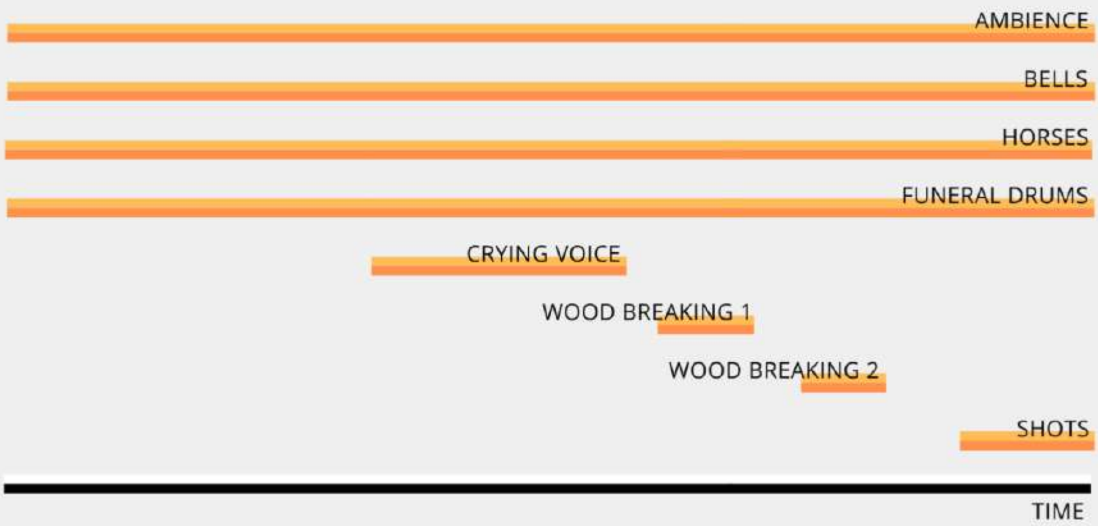
SCENE 2 THE KING'S FUNERAL CEREMONY

ORIGINAL TITLE CERIMÓNIA DA QUEBRA DOS ESCUDOS PELA MORTE DE D. JOÃO V

DATE 26 AUG, 1750

SOUNDS FUNERAL DRUMS, HORSES, BELLS, WOOD BREAKING, SHOTS, CRYING

DESCRIPTION SOMBER FUNERAL PARADE THROUGHTOUT THE CITY OF ÉVORA; ACCOMPANIED BY FUNERAL DRUMS AND THE OFFICIALS ON HORSES AND BELLS ECHOING; "QUEBRA ESCUDOS" BY SOUND WOOD BREAKING; SHOTS FIRED AND THE MOURNING VOICE ANNOUNCING THE KING'S DEATH .



SCENE 3 BANDSTAND INAUGURATION

ORIGINAL TITLE INAUGURAÇÃO DO CORETO NO PASSEIO PÚBLICO

DATE 20 MAI, 1888

SOUNDS FIREWORKS, CLAPS, PEOPLE OUTSIDE, ORCHESTRA MUSIC

DESCRIPTION CONCERT WITH LIVE AUDIENCE IN A PUBLIC PARK TO CELEBRATE THE INAUGURATION A BANDSTAND. EXCERPT OF THE CONCERT (GUILHERME TELL PIECE); PEOPLE CHEERING AT THE END.

SCENE 4 THE FIRST OF DECEMBER

ORIGINAL TITLE FESTEJOS DO 1º DE DEZEMBRO

DATE 01 DEC, 1881

SOUNDS FIREWORKS, CHANTS, BELLS, 1ST DEC. ANTHEM

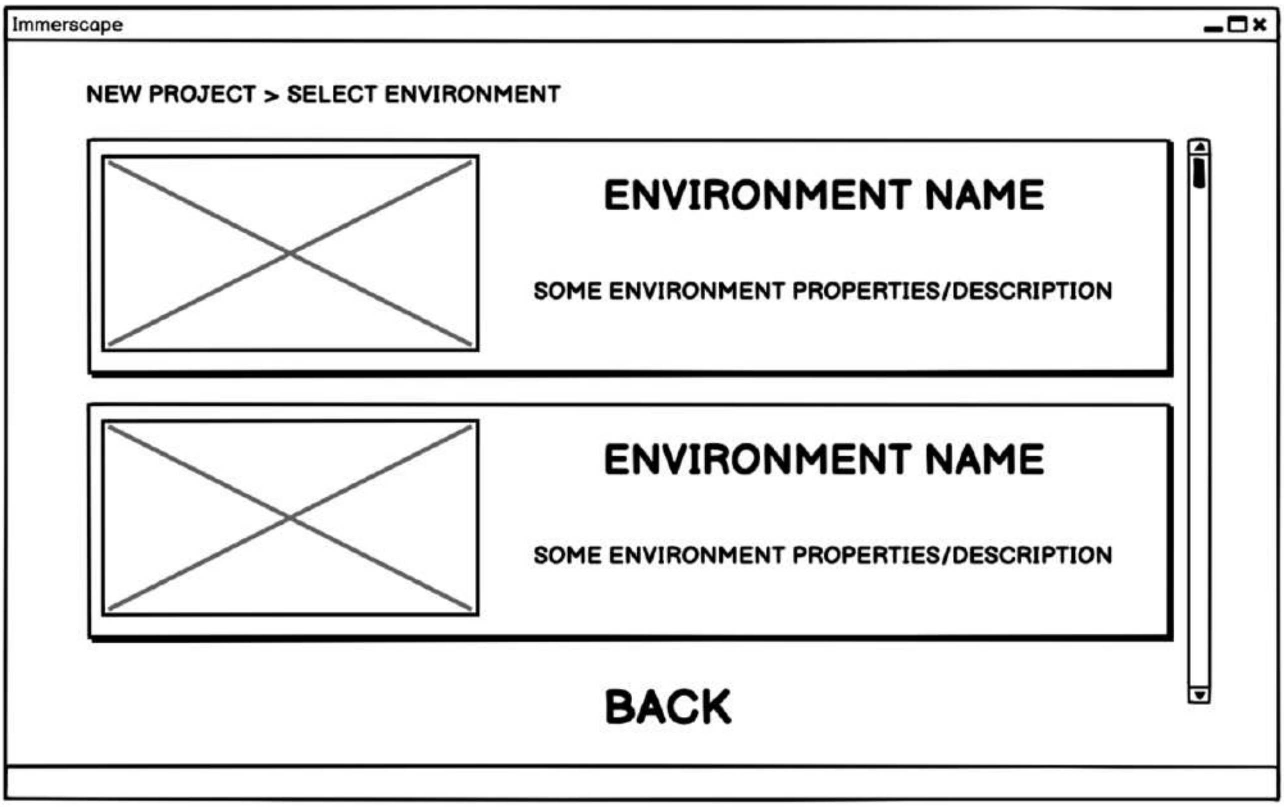
DESCRIPTION PARADE THROUGHT THE CITY; THE BAND PLAYS THE 1ST OF DECEMBER ANTHEM AND PEOPLE CELEBRATE WITH FIREWORKS AND "VIVA PORTUGAL" CHANTS OUTSIDE.

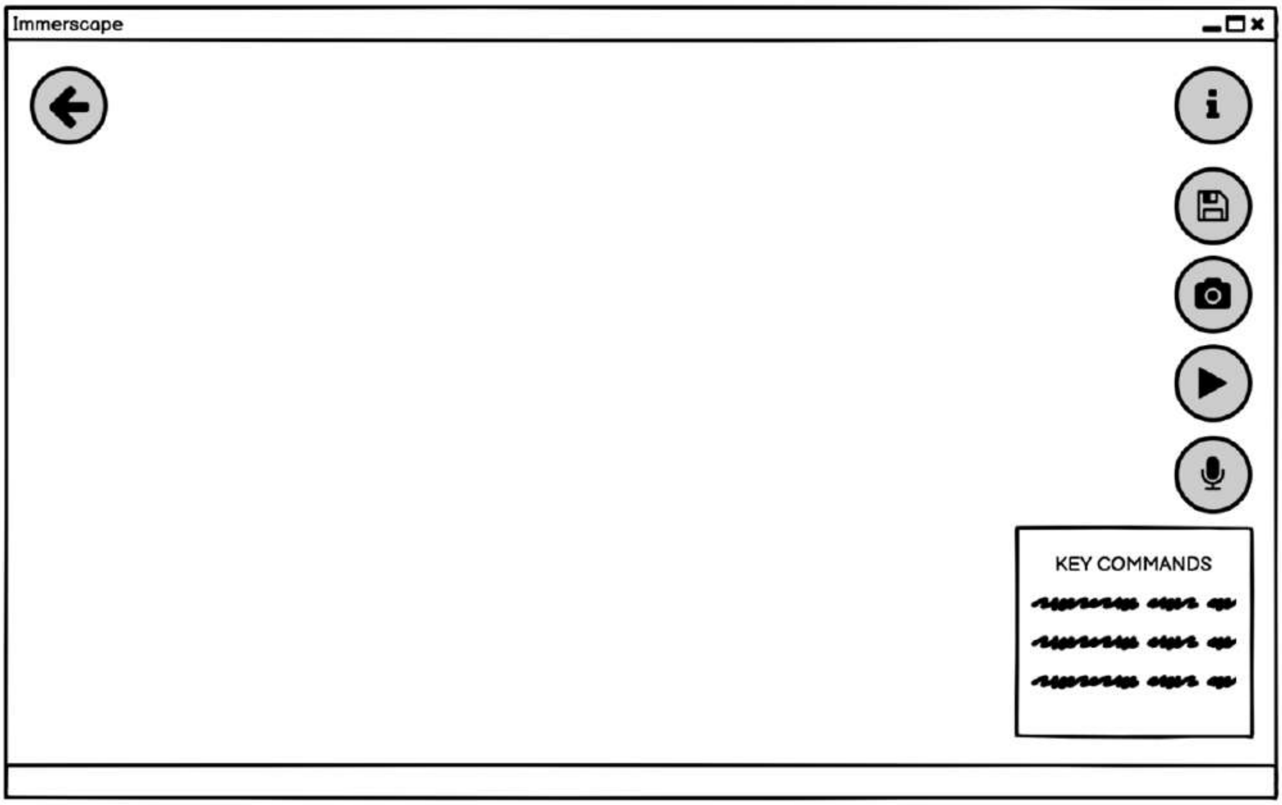
B

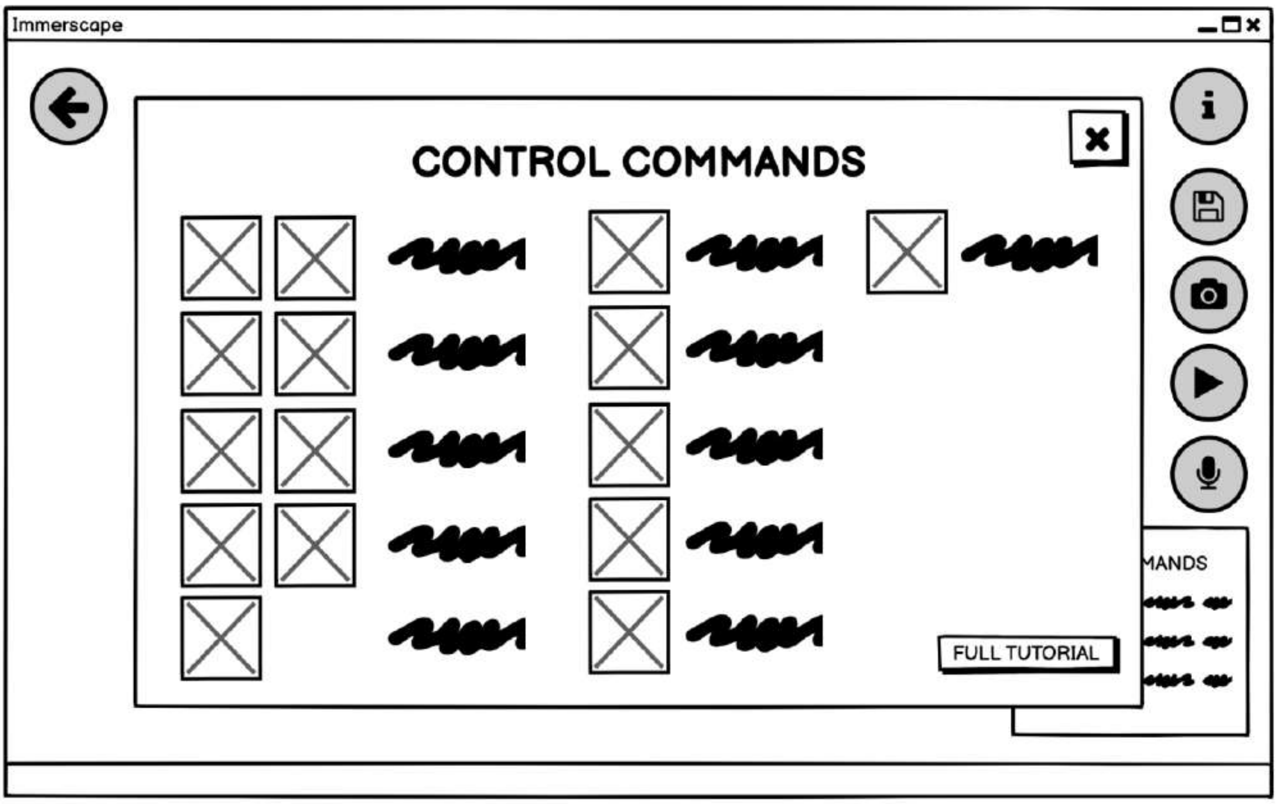
Immerscape Mockups

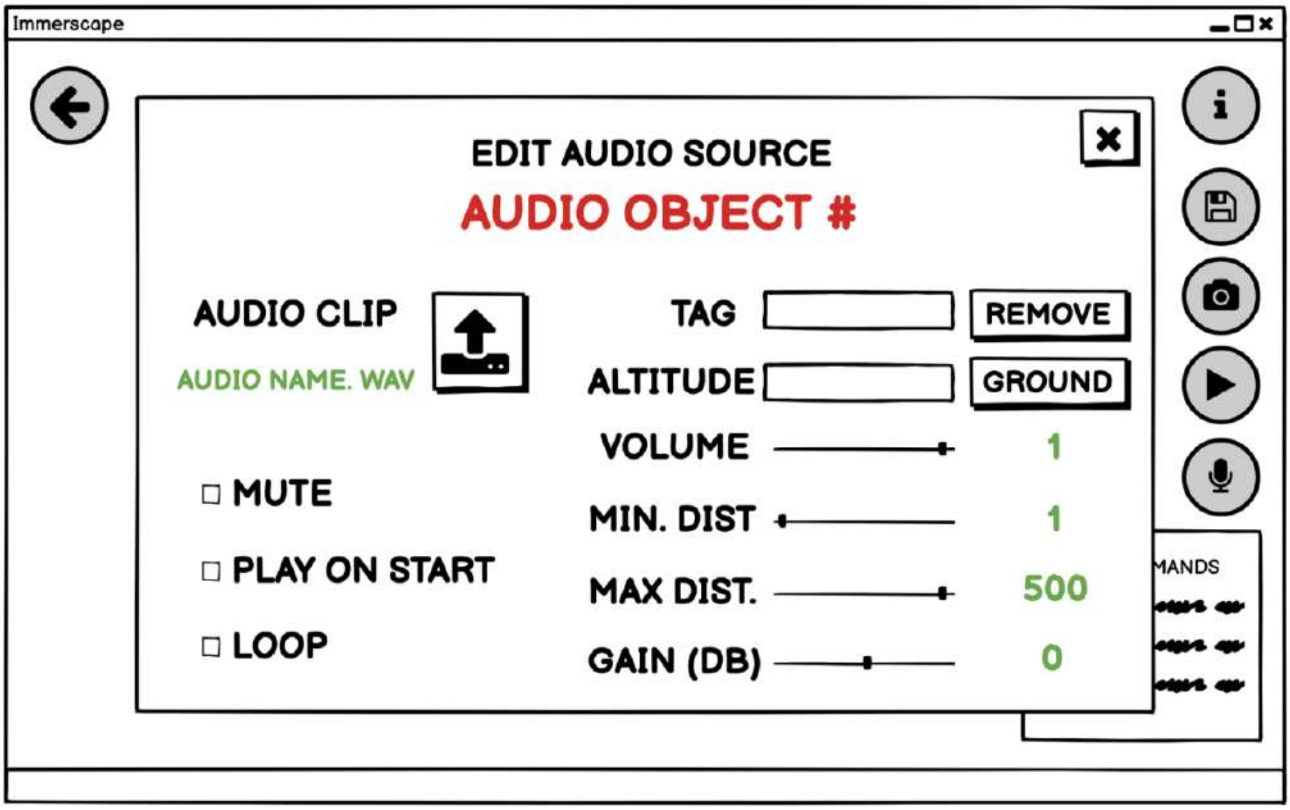
During the design of Immerscape, after the requirements analysis and the identification of the tool's main features, there were made progressive attempts at designing a simple interface for Immerscape's first prototype. The following images are the last version of the prototype's mock-ups (created using Balsamiq Cloud).

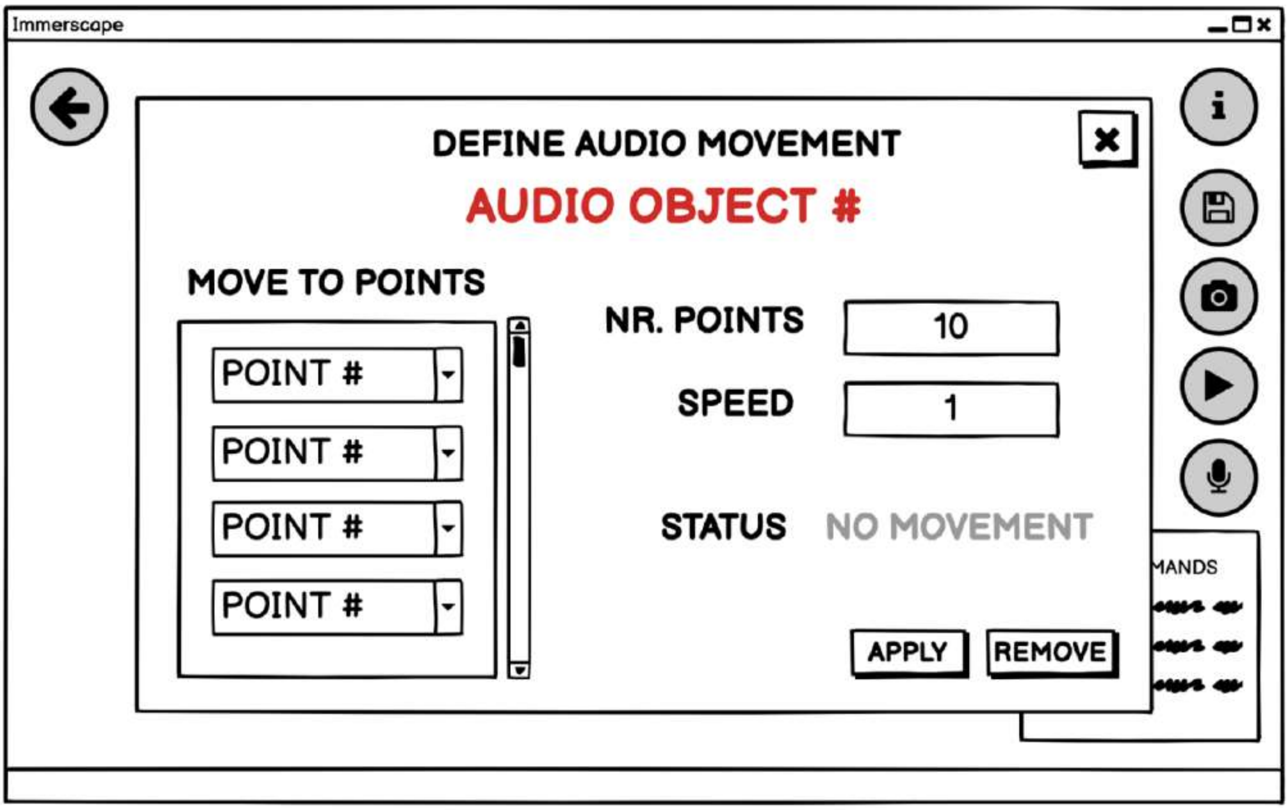

Immerscape
IMMERSCAPE
NEW PROJECTS
SAVED PROJECTS
RECORDINGS
QUIT

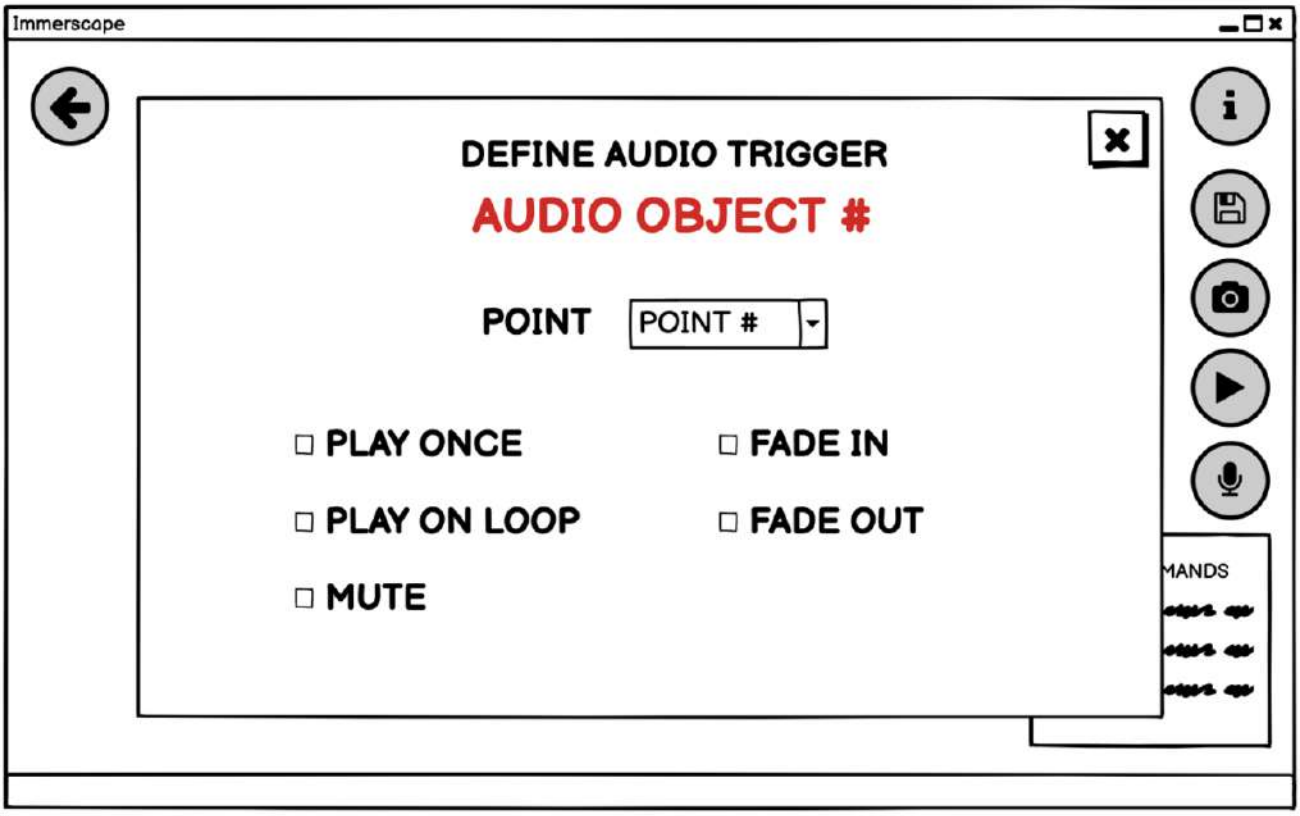

Immerscape
NEW PROJECT > SELECT ENVIRONMENT
ENVIRONMENT NAME
SOME ENVIRONMENT PROPERTIES/DESCRIPTION
ENVIRONMENT NAME
SOME ENVIRONMENT PROPERTIES/DESCRIPTION
BACK

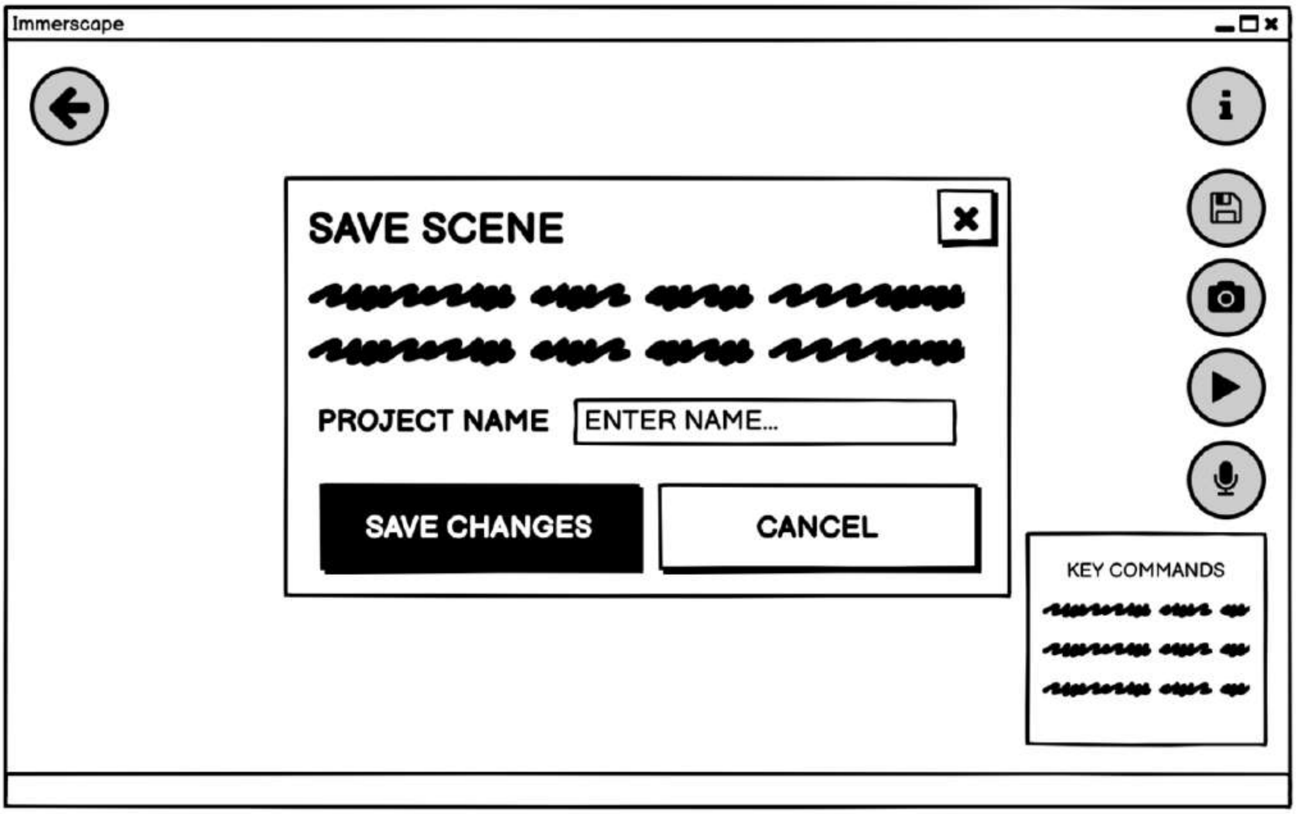

Immerscape
KEY COMMANDS

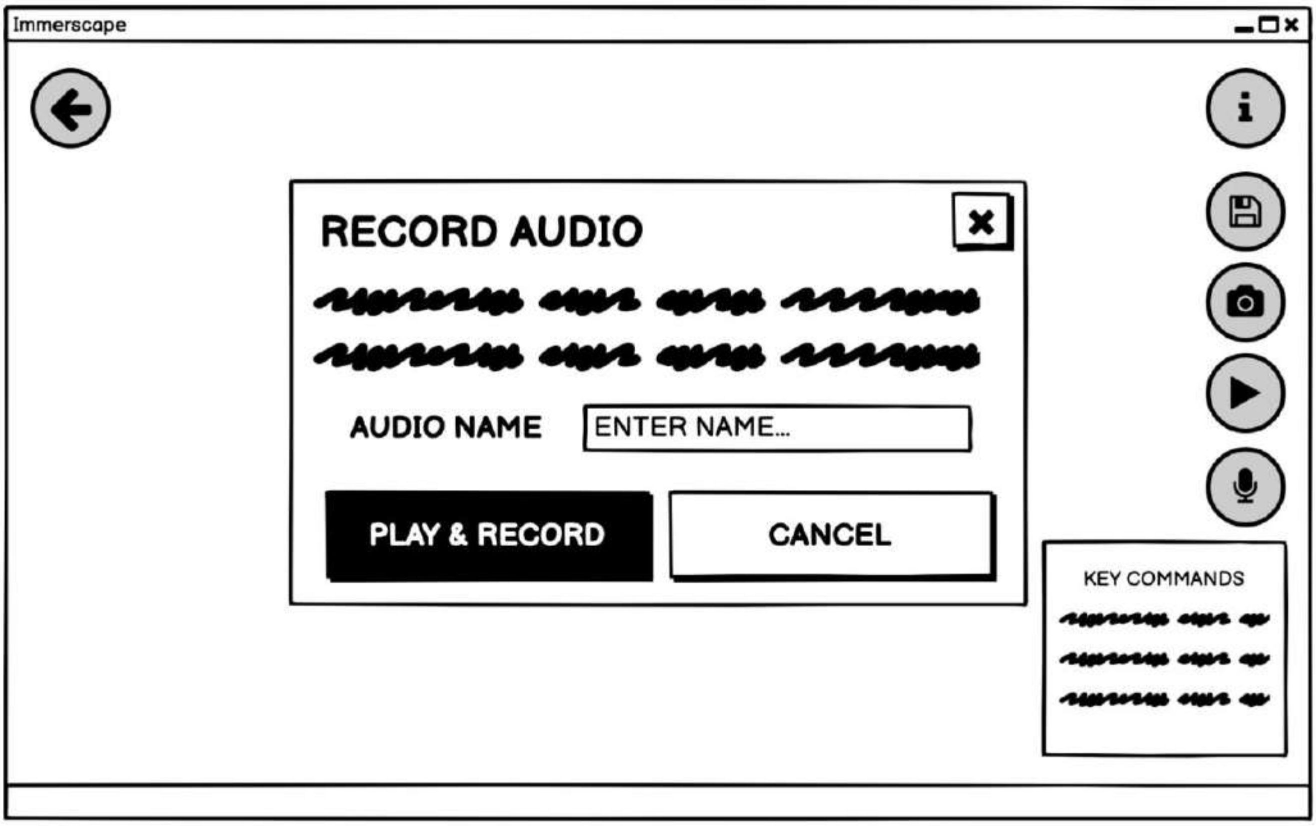

Immerscape
CONTROL COMMANDS
FULL TUTORIAL
MANDS

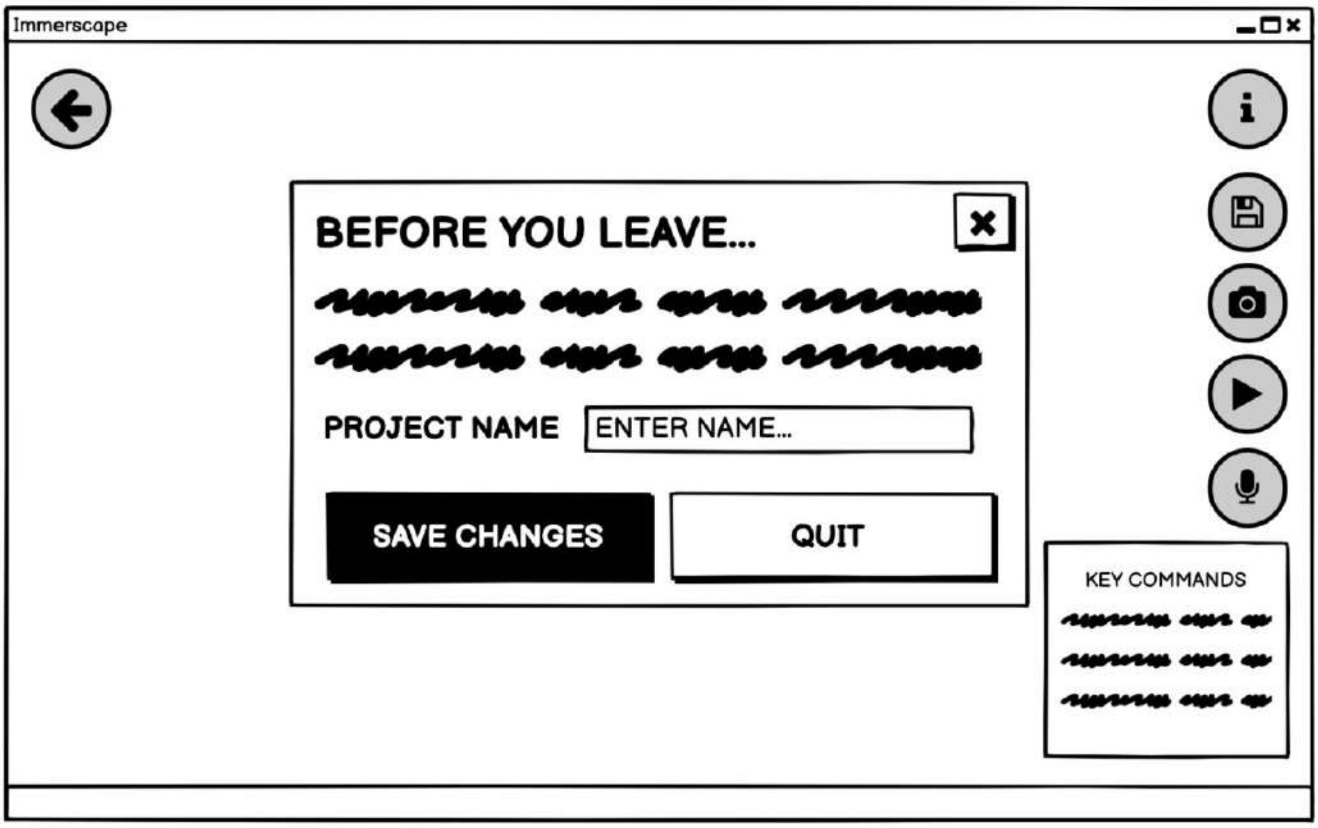

Immerscape
EDIT AUDIO SOURCE
AUDIO OBJECT #
AUDIO CLIP
AUDIO NAME. WAV
TAG
REMOVE
ALTITUDE
GROUND
VOLUME
1
MUTE
MIN. DIST
1
PLAY ON START
MAX DIST.
500
LOOP
GAIN (DB)
0
MANDS

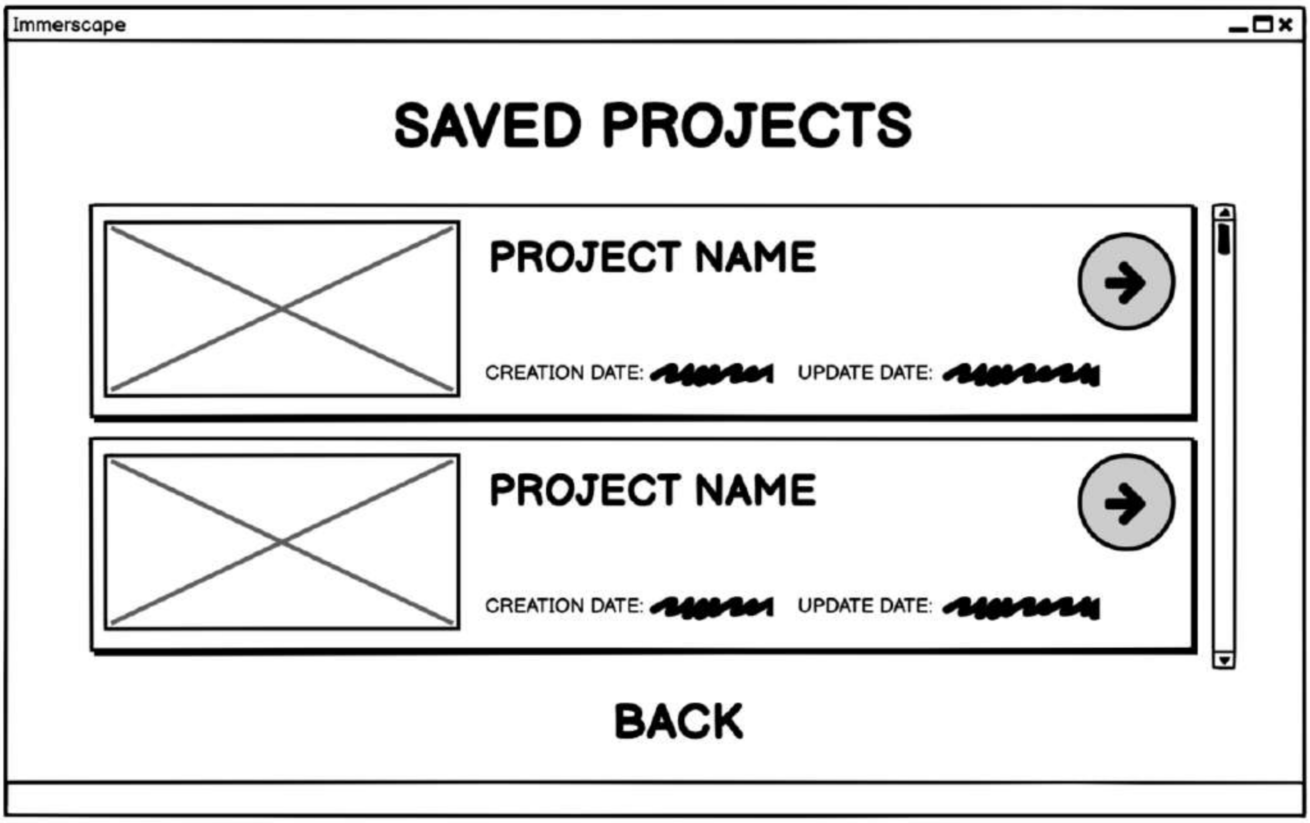

Immerscape
DEFINE AUDIO MOVEMENT
AUDIO OBJECT #
MOVE TO POINTS
POINT #
POINT #
POINT #
POINT #
NR. POINTS
10
SPEED
1
STATUS
NO MOVEMENT
APPLY
REMOVE
MANDS

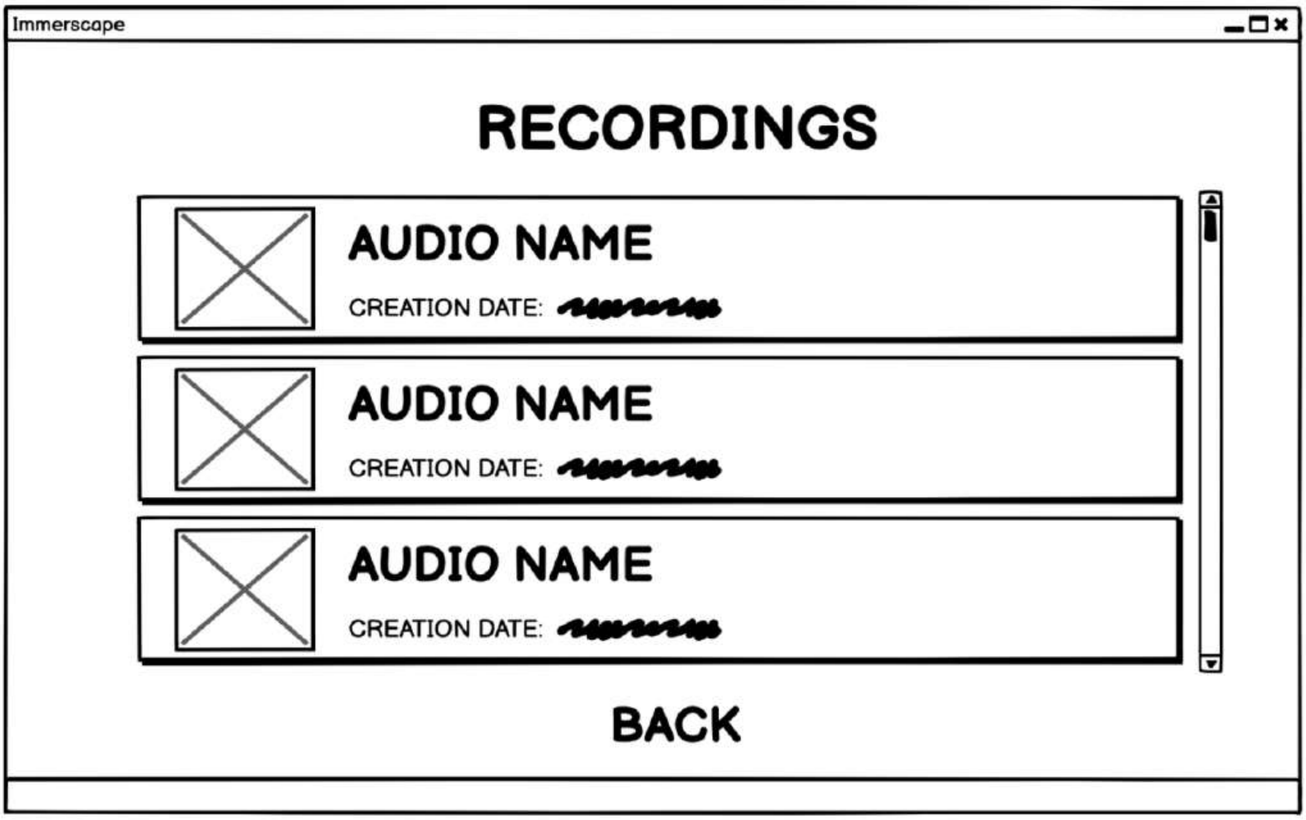

Immerscape
DEFINE AUDIO TRIGGER
AUDIO OBJECT #
POINT
POINT #
PLAY ONCE
FADE IN
PLAY ON LOOP
FADE OUT
MUTE
MANDS


Immerscape
SAVE SCENE
PROJECT NAME
ENTER NAME...
SAVE CHANGES
CANCEL
KEY COMMANDS


Immerscape
RECORD AUDIO
AUDIO NAME
ENTER NAME...
PLAY & RECORD
CANCEL
KEY COMMANDS


Immerscape
BEFORE YOU LEAVE...
PROJECT NAME
ENTER NAME...
SAVE CHANGES
QUIT
KEY COMMANDS


Immerscape
SAVED PROJECTS
PROJECT NAME
CREATION DATE:
UPDATE DATE:
PROJECT NAME
CREATION DATE:
UPDATE DATE:
BACK


Immerscape
RECORDINGS
AUDIO NAME
CREATION DATE:
AUDIO NAME
CREATION DATE:
AUDIO NAME
CREATION DATE:
BACK

C

User Tests

In this appendix, it is presented the documentation required for the user study. The first document refers to the participant's test guide. The guide included a consent of participation form and the test instructions for each of the tasks the participants had to complete.

Additionally, it is also presented the form that participants had to complete during the testing process. The presented form registered the results and the participants' feedback. It is important to highlight that in addition to the questions answered in this document, other feedback was registered during the individual testing sessions. The subjective questions users answered were analyzed individually to understand the point of view of test subjects, particularly the difficulties they found during the process.

Criação de Áudio Imersivo no contexto de Paisagens Sonoras Históricas

Investigadores: Carolina Ferreira, Armanda Rodrigues, Sofia Cavaco

Informação e Consentimento do Participante no Estudo

O presente projeto de investigação corresponde ao desenvolvimento de uma plataforma de geração de áudio espacial com o objetivo de auxiliar o projeto PASEV na sua missão de preservar e promover paisagens sonoras históricas. A particularidade desta plataforma prende-se na imersividade auditiva conferida ao ficheiro gerado pelo utilizador, a partir do conteúdo áudio carregado para a mesma. Neste estudo pretendemos testar as funcionalidades da plataforma assim como analisar a facilidade com que um utilizador não-especializado consegue construir cenas sonoras e gerar ficheiros áudio representativos das mesmas.

O presente estudo será realizado com recurso a um computador via Zoom e com apoio de auscultadores. Será realizado num ambiente com algum ruído de fundo, de forma a simular ambientes de uso real da aplicação.

A sua participação é voluntária. Não receberá nenhum benefício pessoal ou monetário por participar neste estudo, mas permitirá aos investigadores recolherem dados que valorizam a experiência. O anonimato dos participantes será mantido durante a análise dos dados e a publicação/apresentação dos resultados:

(1) O participante receberá uma referência numérica e o seu nome não será divulgado.
(2) Os dados serão guardados segundo esta referência numérica e não segundo o nome do participante.
(3) Só os membros da equipa de investigação terão acesso em detalhe os dados recolhidos.
(4) Os ficheiros e formulários de resposta serão guardados em locais seguros, acessíveis apenas a investigadores autorizados.
(5) A sessão de teste será gravada para referência durante a análise, publicação, e apresentação dos resultados. A gravação será posteriormente apagada.

Se tiver alguma questão sobre este estudo, não hesite em perguntar agora ou em qualquer outro momento usando os seguintes contactos:

Professora: Sofia Cavaco
Departamento Informática
Faculdade Ciências e Tecnologia, UNL
scavaco@fct.unl.pt

Professora: Armanda Rodrigues
Departamento Informática
Faculdade Ciências e Tecnologia, UNL
a.rodrigues@fct.unl.pt

Estudante: Carolina Ferreira
Departamento Informática
Faculdade Ciências e Tecnologia, UNL
crd.ferreira@campus.fct.unl.pt

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVALINCS LABORATORY FOR COMPUTER SCIENCE AND INFORMATICS

ÉVORA HISTORICAL SOUNDSCAPE

Criação de Áudio Imersivo no contexto de Paisagens Sonoras Históricas

Investigadores: Carolina Ferreira, Armanda Rodrigues, Sofia Cavaco

Compreendo a natureza deste estudo e concordo em participar. Dou permissão ao investigador principal e seus associados de apresentarem os resultados deste trabalho por escrito ou de forma oral em aulas ou apresentações académicas de forma a permitir o avanço da ciência, sem nenhuma autorização adicional minha, desde que a minha identidade não seja revelada nem divulgada.

Assinatura do participante:

____

Data: _____________/____________________/______________

Criação de Áudio Imersivo no contexto de Paisagens Sonoras Históricas

Investigadores: Carolina Ferreira, Armanda Rodrigues, Sofia Cavaco

Instruções de Teste

O presente guião corresponde a um conjunto de 10 tarefas para testar as principais funcionalidades da aplicação de geração de áudio, Immerscape. As tarefas serão desempenhadas no computador e guiadas via Zoom com recurso a auscultadores.

Neste documento são apresentadas as instruções para desempenhar cada uma das tarefas. Adicionalmente, será fornecido um formulário online no qual se encontram perguntas relativas às tarefas e espaço para responder às mesmas. Neste conjunto de testes serão avaliadas as principais funcionalidades da aplicação assim como determinada a facilidade com que é possível criar uma cena sonora e gerar áudio 3D representativo da mesma.

Por favor, leia cada tarefa até ao fim antes de a iniciar e procure esclarecer quaisquer dúvidas, caso estas existam. As tarefas devem ser realizadas sequencialmente e após a realização de cada uma, deve ser preenchida a seção do formulário correspondente à mesma. No final da execução das tarefas, terá uma seção aberta dedicada a comentários que queira tecer.

Cena Exemplo de Teste:

Vamos criar uma cena simples para testar algumas das principais funcionalidades da ferramenta áudio. Imaginemos um cenário em que estamos num espaço aberto e temos cavalos que se movimentam à nossa volta num movimento circular. Sempre que os cavalos passam à nossa direita, é disparado um tiro para o ar no centro do percurso do cavalo.

Link Formulário de Teste: https://forms.gle/ufC9j5j8rBsMqXhj9 (preencha a parte referente à informação do participante e ambiente sonoro)

Tarefa 1

Será apresentado o ecrã principal da aplicação. Deve criar um novo projeto e selecionar como ambiente sonoro o *template* do ambiente “default”. Neste ponto deve conseguir encontrar um ambiente vazio e deve ter a perspetiva do *player*. O *player* será o foco do teste uma vez que este representa a nossa posição imersiva na cena, ou seja, é a representação dos nossos olhos e ouvidos na cena. Explore o ambiente movimentando o *player* para a frente e para trás. Rode a posição do *player* para alterar a direção do movimento. Olhe para cima e para baixo.

Preencha a seção do formulário referente à Tarefa 1.

Tarefa 2

O próximo objetivo será popular a cena. Existem dois tipos de elementos que podem ser utilizados para criar as cenas sonoras. O principal elemento é um objeto áudio (representado por um cubo vermelho) que pode ser configurado como uma fonte sonora. O segundo elemento é um ponto de marcação (representado com uma bandeira numerada). Estes pontos podem ser utilizados para "desenhar" rotas de movimento para os objetos áudio assim como para desencadear terminados eventos quando os objetos áudios passam pelos mesmos.

Utilize a funcionalidade de trocar a câmara para ter uma vista aérea da cena. O objeto preto em cena é o *player*. No modo de vista aérea, aponte a posição na superfície onde quer colocar o objeto e crie um objeto áudio (audio object 1) em frente ao seu *player*. Usando um método idêntico, adicione quatro pontos de marcação formando um quadrado em volta do objeto áudio, semelhante ao esquema da seguinte figura 1. Volte a trocar a câmara para o modo da visão do *player*.

Preencha a seção do formulário referente à Tarefa 2.

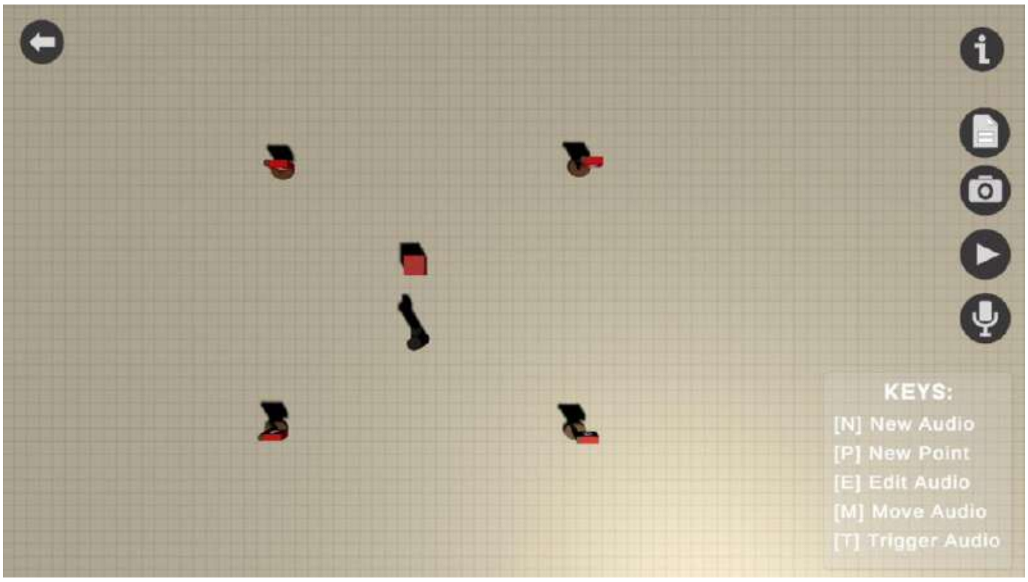


Figura 1 - Esquema da disposição dos pontos de marcação em relação à fonte sonora

Tarefa 3

A fonte sonora que criou deve estar à sua frente. Quando são criadas, as fontes sonoras são "vazias", ou seja, precisam de ser editadas para definir todas as caraterísticas do som que emitem. Abra o painel de edição do objeto áudio que criou apontando para o mesmo. Comece por fazer carregamento do clipe áudio "horses.wav" (localizado na pasta "testing audios" na secretária). Certifique-se que o clipe da fonte sonora será tocado no início da cena e que o clipe será tocado em *loop*. Experimente o campo da altitude com o valor 2 para elevar a posição da fonte sonora. Volte a posicionar o objeto no chão. Feche o painel de edição do áudio.

Preencha a seção do formulário referente à Tarefa 3.

Tarefa 4

Uma vez editada a fonte sonora e uma vez determinada que esta deve tocar no início da cena, selecionando o modo de reprodução da cena, deverá ser possível ouvir o áudio emitido pela mesma. Reproduza a cena. Neste momento a sua cena será composta por apenas um único som. Enquanto no modo de reprodução, utilize o *player* para se descolar na cena, posicionando-se alternadamente à direita, à esquerda, à frente e atrás da fonte sonora; aproximando-se e afastando-se da mesma para determinar o seu efeito espacial. Pare a reprodução da cena.

Preencha a seção do formulário referente à Tarefa 4.

Tarefa 5

O efeito espacial do som pode também ser obtido se, em vez de movimentar o *player*, optarmos por ter um objeto áudio que se move seguindo um padrão ou "rota". Para definir esse padrão deve abrir o painel de movimento do objeto áudio. Queremos que o objeto áudio se movimente percorrendo os quatro pontos criados de forma circular.

No painel de movimento deve preencher o número de pontos que definem o movimento (neste caso serão 4) assim como determinar a ordem dos pontos que o objeto deve percorrer. Utilize uma velocidade igual a 4. Aplique o movimento e feche o painel. Reproduza a cena para ver se obteve o movimento esperado. Pare a reprodução. Quando parar a reprodução, clique com o cursor e segure o objeto áudio para o descolar para o interior do quadrado novamente.

Preencha a seção do formulário referente à Tarefa 5.

Tarefa 6

À semelhança do que fez anteriormente na Tarefa 3, crie um outro objeto áudio (audio object 2). Edite o novo objeto áudio apenas fazendo carregamento do clipe áudio "shot.wav" (localizado na pasta "testing áudios", na secretária) e determinando que o objeto estará a uma altitude igual a 3. Feche o painel de edição. Queremos que a nova fonte áudio (audio object 2) apenas seja reproduzida quando o ponto de marcação (target) à sua direita é ativado. Um ponto de marcação é ativado por exemplo quando um objeto áudio passa por este.

Para determinar em que fonte sonora é aplicada o *trigger*, ou seja, o evento, devemos escolher o objeto áudio e abrir o seu painel de *trigger*. No painel de *trigger,* deve selecionar que ponto de marcação aciona o evento e qual será o evento. Para o audio object 2, escolha o ponto de marcação à sua direita. Selecione o evento em que o clipe é reproduzido **apenas uma vez** por completo quando o *trigger* é acionado.

Reproduza a cena para determinar se satisfaz as suas expectativas. Pare a reprodução e volte a posicionar o audio object 1 no centro do quadrado.

Preencha a seção do formulário referente à Tarefa 6.

Tarefa 7

Assumimos que esta é a cena final. Após a última reprodução da cena, determinámos que estamos satisfeitos com o resultado sonoro. O próximo passo é gravar a cena. Quando pedida a introdução de um nome para a gravação introduza "audio-[número]" em que [número] será o número que lhe será dito no momento do teste e que corresponde ao seu identificador. Quando estiver satisfeito com a sua cena, pare de gravação.

Preencha a seção do formulário referente à Tarefa 7.

Tarefa 8

Deixe o ambiente do projeto. Quando lhe for perguntado se pretende guardar alterações, deve optar por guardar, dando ao projeto o nome "project-[número]" em que [número] será o seu número de participante (que lhe será dito no momento do teste).

Preencha a seção do formulário referente à Tarefa 8.

Tarefa 9

Uma vez de volta ao ecrã principal, deve optar por ver todas as gravações realizadas. Clique na sua gravação para ouvir o resultado da mesma. Volte para o ecrã principal.

Preencha a seção do formulário referente à Tarefa 9.

Tarefa 10

Finalmente, de volta ao ecrã principal, deve optar por ver todos os projetos salvos. Deve encontrar o seu projeto no topo da lista de projetos salvos e voltar a abri-lo. Volte ao ecrã principal.

Preencha a seção do formulário referente à Tarefa 10.

Terminou todas as tarefas. Preencha a última seção do formulário referente às observações globais acerca da plataforma Immerscape. **Muito Obrigada pela sua colaboração!**

Formulário de Teste: Immerscape

O presente questionário corresponde a uma série de perguntas sobre 10 tarefas e visa em avaliar as funcionalidades da aplicação de geração de áudio, Immerscape. O protótipo da aplicação é produzido como resultado do projeto de investigação na FCT NOVA e integrado no projeto PASEV.

Ao longo do formulário serão dadas instruções de como proceder à realização das tarefas, a serem desempenhadas por ordem, no computador e com recurso a auscultadores. As tarefas serão acompanhadas via Zoom. Neste conjunto de testes serão avaliadas as principais funcionalidades da aplicação assim como determinada a facilidade com que é possível criar uma cena sonora e gerar áudio 3D representativo da mesma.

Por favor, leia cada tarefa até ao fim antes de a iniciar e procure esclarecer quaisquer dúvidas, caso estas existam. As tarefas devem ser realizadas sequencialmente e após a realização de cada uma, deve ser preenchida a seção do formulário correspondente à mesma. No final da execução das tarefas, terá uma seção aberta dedicada a comentários que queira tecer.

Este questionário complementa os documentos "Informação e Consentimento do Participante no Estudo" e "Instruções de Teste". Leia atentamente os referidos documentos antes de prosseguir neste questionário. Caso tenha alguma dúvida, deve questionar no momento ou posteriormente usando os contactos apresentados no documento "Informação e Consentimento do Participante no Estudo".

*Obrigatório

1. Email *

____

2. Concordo com a informação apresentada e aceito participar nos testes que me foram apresentados? *

Marcar apenas uma oval.

- Sim
- Não

Informação do Participante

Nesta secção são apresentadas questões que visam a caraterizar o indivíduo.

3. Número de Participante *

Deverá pedir ao investigador o seu número de participante.

____

4. Idade *

____

5. Sexo *

Marcar apenas uma oval.

- Feminino
- Masculino
- Prefiro não responder
- Outra: ______________________

6. Nivel de Escolaridade *

Selecione um dos níveis de escolaridade. Caso não se encontre nas opções indicadas, introduza o seu nível de escolaridade.

Marcar apenas uma oval.

- 2º ciclo do ensino básico
- 3º ciclo do ensino básico
- Ensino Secundário
- Licenciatura
- Mestrado
- Doutoramento
- Outra: ______________________

7. Tem algum problema de audição? *

Se sim, qual? Caso contrário, escreva "Não" na caixa de texto.

____

8. Conhece alguma ferramenta de criação de áudio? *

Se sim, por favor explicite qual/quais e para que fins a(s) usou. Caso contrário, escreva "Não" na caixa de texto.

____

Ambiente Sonoro

A seção destina-se a apurar as condições em que foi realizado o teste.

9. Indique o tipo de auricular/auscultador utilizou para ouvir o áudio: *

Os auriculares, também chamados de headphones "on-ear" (na orelha), assentam diretamente na cavidade do ouvido. São habitualmente mais pequenos que os auscultadores. Os auscultadores são também chamados de headphones "over-ear" (sobre a orelha). Estes tendem a ser os maiores em tamanho e, por cobrirem toda a orelha, oferecem normalmente uma excelente qualidade de áudio.

Marcar apenas uma oval.

- Auricular com fios
- Auricular wireless
- Auricular com fios & com noise canceling
- Auricular wireless & com noise canceling
- Auscultadores com fios
- Auscultadores wireless
- Auscultadores com fios & noise canceling
- Auscultadores wireless & com noise canceling

10. Identifique o modelo dos auriculares/auscultadores que utilizou para ouvir o áudio: *

Se não souber o modelo do seu dispositivo, escreva "Não sei" na caixa de texto.

____

11. Qual o nível de ruído na posição em que se encontra? *

Classifique o nível de ruído em sua volta: 1 - Muito pouco ruído; 2 - Pouco ruído; 3 - Algum ruído mas não predominante; 4 - Ruído; 5 - Muito ruído

Marcar apenas uma oval.

	1	2	3	4	5	
Muito pouco ruído	○	○	○	○	○	Muito ruído

12. Observações sobre o ambiente sonoro: *

Descreva qual foi a influência da sua posição no ambiente e o impacto do ruído na sua capacidade de ouvir.

____

Tarefa 1

Será apresentado o ecrã principal da aplicação. Deve criar um novo projeto e selecionar como ambiente sonoro o template do ambiente "default". Neste ponto deve conseguir encontrar um ambiente vazio e deve ter a perspetiva do player. O player será o foco do teste uma vez que este representa a nossa posição imersiva na cena, ou seja, é a representação dos nossos olhos e ouvidos na cena.

Explore o ambiente movimentando o player para a frente e para trás. Rode a posição do player para alterar a direção do movimento. Olhe para cima e para baixo.

13. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 1: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

14. Considera um ponto positivo da ferramenta ter uma coleção de ambientes pré-definidos para escolha quando à criação de um novo projeto? *

Marcar apenas uma oval.

◯ Sim
◯ Não

15. Que tipo de ambientes sonoros genéricos pré-definidos considera que seria importante incluir como opções na ferramenta? *

16. Conseguiu movimentar o player para a frente e para trás? *

Marcar apenas uma oval.

◯ Sim
◯ Não

17. Conseguiu rodar a posição do player? *

Marcar apenas uma oval.

◯ Sim
◯ Não

18. Como qualifica a capacidade de movimentação do player? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco Satisfatória	◯	◯	◯	◯	◯	Muito Satisfatória

19. Que outro tipo de capacidade de movimento acha que o player deveria ter? *

20. Comentários relativos à Tarefa 1

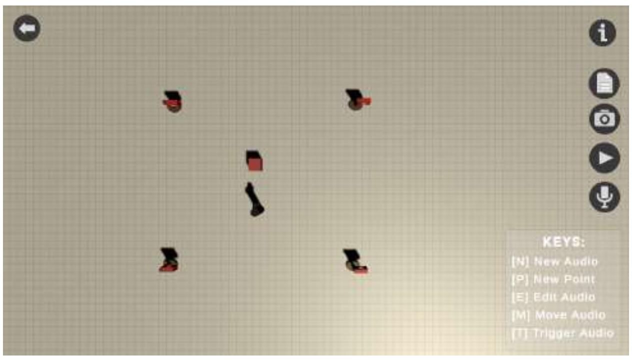
Tarefa 2

Os detalhes acerca da cena que vamos criar estão explicitados na seção "Cena Exemplo de Teste" no documento "Instruções de Teste".

O próximo objetivo será popular a cena. Existem dois tipos de elementos que podem ser utilizados para criar as cenas sonoras. O principal elemento é um objeto áudio (representado por um cubo vermelho) que pode ser configurado como uma fonte sonora. O segundo elemento é um ponto de marcação (representado com uma bandeira numerada). Estes pontos podem ser utilizados para "desenhar" rotas de movimento para os objetos áudio assim como para desencadear terminados eventos quando os objetos áudios passam pelos mesmos.

Utilize a funcionalidade de trocar a câmara para ter uma vista aérea da cena. O objeto preto em cena é o player. No modo de vista aérea, aponte a posição na superfície onde quer colocar o objeto e crie um objeto áudio (audio object 1) em frente ao seu player. Usando um método idêntico, adicione quatro pontos de marcação formando um quadrado em volta do objeto áudio, semelhante ao esquema da seguinte figura 1. Volte a trocar a câmara para o modo da visão do player.

Figura 1: Exemplo do posicionamento dos elementos em cena



21. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 2: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

22. Conseguiu alternar entre os modos da câmara? *

Marcar apenas uma oval.

◯ Sim
◯ Não

23. Considera um ponto positivo da ferramenta conseguir alternar o modo da câmara? *

Marcar apenas uma oval.

◯ Sim
◯ Não

24. Conseguiu adicionar o objeto áudio à cena em frente do seu player? *

Marcar apenas uma oval.

◯ Sim
◯ Não

25. Conseguiu adicionar os pontos de marcação numa disposição semelhante à da Figura 1? *

Marcar apenas uma oval.

◯ Sim
◯ Não

26. Que outro tipo de elementos considera que poderiam ser importantes adicionar durante a construção das cenas? *

Para além de fontes sonoras e marcas de movimentação/trigger, que outros elementos gostaria de ter disponível? Ou alternativamente, que funcionalidades adicionais gostaria que os elementos disponíveis tivessem?

27. Comentários relativos à Tarefa 2

Tarefa 3

A fonte sonora que criou deve estar à sua frente. Quando são criadas, as fontes sonoras são "vazias", ou seja, precisam de ser editadas para definir todas as caraterísticas do som que emitem. Abra o painel de edição do objeto áudio que criou apontando para o mesmo.

Comece por fazer carregamento do clipe áudio "horses.wav" (localizado na pasta "testing audios" na secretária). Certifique-se que o clipe da fonte sonora será tocado no início da cena e que o clipe será tocado em loop. Experimente o campo da altitude com o valor 2 para elevar a posição da fonte sonora. Volte a posicionar o objeto no chão. Feche o painel de edição do áudio.

28. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 3: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

29. Conseguiu abrir o painel de edição do objeto áudio? *

Marcar apenas uma oval.

◯ Sim
◯ Não

30. Conseguiu fazer o carregamento do ficheiro de áudio? *

Marcar apenas uma oval.

◯ Sim
◯ Não

31. Selecionou a opção "Play On Start" para assegurar que o áudio tocaria no início da cena? *

Marcar apenas uma oval.

◯ Sim
◯ Não

32. Selecionou a opção "Loop" para assegurar que o áudio tocaria repetidamente durante a reprodução da cena? *

Marcar apenas uma oval.

◯ Sim
◯ Não

33. Experimentou a altitude e conseguiu posicionar o objeto áudio novamente no chão? *

Marcar apenas uma oval.

◯ Sim
◯ Não

34. Classifique a facilidade em editar as propriedades do objeto áudio: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

35. De acordo com aquilo que viu no painel de edição do objeto áudio, que outras caraterísticas do objeto áudio gostaria de conseguir editar? *

36. Comentários relativos à Tarefa 3

Tarefa 4

Uma vez editada a fonte sonora e uma vez determinada que esta deve tocar no início da cena, selecionando o modo de reprodução da cena, deverá ser possível ouvir o áudio emitido pela mesma. Reproduza a cena. Neste momento a sua cena será composta por apenas um único som.

Enquanto no modo de reprodução, utilize o player para se descolar na cena, posicionando-se alternadamente à direita, à esquerda, à frente e atrás da fonte sonora; aproximando-se e afastando-se da mesma para determinar o seu efeito espacial. Pare a reprodução da cena.

37. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 4: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

38. Conseguiu reproduzir a cena e ouvir o áudio carregado? *

Marcar apenas uma oval.

◯ Sim
◯ Não

39. Conseguiu parar a reprodução da cena? *

Marcar apenas uma oval.

◯ Sim
◯ Não

40. Conseguiu detetar diferenças na direção do som ao alternar a sua posição em relação à fonte sonora? *

Marcar apenas uma oval.

◯ Sim
◯ Não

41. Considera que conseguir movimentar o player durante a reprodução da cena é um ponto positivo da ferramenta áudio? *

Marcar apenas uma oval.

- Sim
- Não

42. Classifique a facilidade em testar a reprodução da cena quando se está a compor a mesma: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

43. Comentários relativos à Tarefa 4

Tarefa 5

O efeito espacial do som pode também ser obtido se, em vez de movimentar o player, optarmos por ter um objeto áudio que se move seguindo um padrão ou "rota". Para definir esse padrão deve abrir o painel de movimento do objeto áudio. Queremos que o objeto áudio se movimente de forma a percorrer os quatro pontos criados de forma circular.

No painel de movimento deve preencher o número de pontos que definem o movimento assim como determinar a ordem dos pontos que o objeto deve percorrer. Utilize uma velocidade igual a 4. Aplique o movimento e feche o painel. Reproduza a cena para ver se obteve o movimento esperado. Pare a reprodução. Quando parar a reprodução, clique com o cursor e segure o objeto áudio para o descolar para o interior do quadrado novamente.

44. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 5: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

45. Conseguiu abrir o painel de movimentação do objeto áudio? *

Marcar apenas uma oval.

- Sim
- Não

46. Quando experimentou a reprodução da cena, obteve o movimento que esperava do objeto áudio? *

Marcar apenas uma oval.

- Sim
- Não

47. Após terminar a reprodução da cena, conseguiu pegar no objeto áudio com o cursor e move-lo para uma posição próxima da sua posição inicial? *

Marcar apenas uma oval.

- Sim
- Não

48. Considera uma limitação grave apenas poder usar no máximo 10 pontos para definir o movimento de um objeto áudio?

Marcar apenas uma oval.

- Sim
- Não

49. Classifique a facilidade em adicionar movimento ao objeto áudio: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

50. De acordo com aquilo que viu no painel de movimentação do objeto áudio, que outros aspetos do movimento gostaria de conseguir editar? *

51. Comentários relativos à Tarefa 5

Tarefa 6

À semelhança do que fez anteriormente na Tarefa 3, crie um outro objeto áudio (audio object 2). Edite o novo objeto áudio apenas fazendo carregamento do clipe áudio "shot.wav" (localizado na pasta "testing áudios", na secretária) e determinando que o objeto estará a uma altitude igual a 3. Feche o painel de edição. Queremos que a nova fonte áudio (audio object 2) apenas seja reproduzida quando o ponto de marcação 2 (target 2) é ativado. Um ponto de marcação é ativado por exemplo quando um objeto áudio passa por este.

Para determinar em que fonte sonora é aplicada o trigger, ou seja, o evento, devemos escolher o objeto áudio e abrir o seu painel de trigger. No painel de trigger, deve selecionar que ponto de marcação aciona o evento e qual será o evento. Para o audio object 2, escolha o ponto de marcação à sua direita. Selecione o evento em que o clipe seja reproduzido apenas uma vez por completo quando o trigger é acionado.

Reproduza a cena para determinar se satisfaz as suas expectativas. Pare a reprodução e volte a posicionar o audio object 1 no centro do quadrado.

52. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 6: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

53. Conseguiu realizar a edição da nova fonte de áudio? *

Marcar apenas uma oval.

- Sim
- Não

54. Conseguiu abrir o panel de trigger do objeto áudio? *

Marcar apenas uma oval.

- Sim
- Não

55. Selecionou o ponto de marcação à sua direita como sendo o ponto de trigger para o evento? *

Marcar apenas uma oval.

- Sim
- Não

56. Selecionou o "Play Once Shot" como sendo o evento para o trigger? *

Marcar apenas uma oval.

- Sim
- Não

57. Quando experimentou a reprodução da cena obteve o resultado que esperava? Conseguiu ouvir o áudio relativo ao ficheiro "shot.wav"? *

Marcar apenas uma oval.

- Sim
- Não

58. Classifique a facilidade em adicionar um trigger ao objeto áudio: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

59. De acordo com aquilo que viu no painel de trigger do objeto áudio, que outros eventos gostaria de ter disponíveis? *

60. Comentários relativos à Tarefa 6

Tarefa 7

Assumimos que esta é a cena final. Após a última reprodução da cena, determinámos que estamos satisfeitos com o resultado sonoro. O próximo passo é gravar a cena. Quando pedida a introdução de um nome para a gravação introduza "audio-[número]" em que [número] será o número que lhe será dito no momento do teste e que corresponde ao seu identificador de participante. Quando estiver satisfeito com a sua cena, pare de gravação.

61. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 7: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

62. Conseguiu iniciar a gravação da cena? *

Marcar apenas uma oval.

- Sim
- Não

63. Conseguiu terminar a gravação da cena? *

Marcar apenas uma oval.

- Sim
- Não

64. Considera que conseguir ouvir a reprodução da cena durante a gravação é um ponto positivo da ferramenta áudio? *

Marcar apenas uma oval.

- Sim
- Não

65. Classifique a facilidade em gravar uma cena: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

66. Comentários relativos à Tarefa 7

Tarefa 8

Deixe o ambiente do projeto salvando o seu progresso. Quando lhe for perguntado se pretende guardar alterações, deve optar por guardar, dando ao projeto o nome "project-[número]" em que [número] será o seu número de participante (que lhe será dito no momento do teste).

67. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 8: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

68. Quão importante é para si poder guardar o progresso de um projeto? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco importante						Muito importante

69. Comentários relativos à Tarefa 8

Tarefa 9

Uma vez de volta ao ecrã principal, deve optar por ver todas as gravações realizadas. Clique na sua gravação para ouvir o resultado da mesma. Volte para o ecrã principal.

70. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 9: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil						Muito Fácil

71. Quão importante é para si poder aceder às gravações feitas através da ferramenta áudio e reproduzi-las? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco importante	◯	◯	◯	◯	◯	Muito importante

72. Quão importante é para si ter as gravações feitas ordenadas por data decrescente de criação? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco importante	◯	◯	◯	◯	◯	Muito importante

73. Considerando aquilo que viu no ecrã das gravações, que outra informação/funcionalidades gostaria de ter relativo ao mesmo? *

74. Comentários relativos à Tarefa 9

Tarefa 10

Finalmente, de volta ao ecrã principal, deve optar por ver todos os projetos salvos. Deve encontrar o seu projeto no topo da lista de projetos salvos visto que é o mais recente e abri-lo outra vez. Volte ao ecrã principal.

75. Qual o grau de dificuldade que sentiu ao realizar a Tarefa 10: *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

76. Quão importante é para si poder aceder aos projetos salvos através da ferramenta áudio? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco importante	◯	◯	◯	◯	◯	Muito importante

77. Quão importante é para si ter a listagem de projetos salvos ordenada por data de atualização? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco importante	◯	◯	◯	◯	◯	Muito importante

78. Considerando aquilo que viu no ecrã dos projetos salvos, que outra informação/funcionalidades gostaria de ter relativo ao mesmo? *

79. Comentários relativos à Tarefa 10

Últimas Considerações

Este espaço é reservado para comentários adicionais que queira prestar, quer das tarefas executadas. Utilize este espaço para considerações e/ou sugestões que queria fazer. A seguir encontram-se tópicos sugeridos para possíveis comentários que possa querer tecer.

80. Foi bem sucedido/a na realização das tarefas conseguindo uma gravação da cena que compôs? *

Marcar apenas uma oval.

◯ Sim
◯ Não

81. Qual o grau de dificuldade na criação de áudio espacial utilizando a ferramenta? *

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Difícil	◯	◯	◯	◯	◯	Muito Fácil

82. Classifique a imersividade do áudio gravado durante o teste: *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco Imersivo	◯	◯	◯	◯	◯	Muito Imersivo

83. Classifique a satisfação ao usar a ferramenta: *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco Satisfeito/a	◯	◯	◯	◯	◯	Muito Satisfeito/a

84. Considera que esta pode ser uma ferramenta útil para quem queira compor e gerar clipes de áudio imersivo? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco Útil	◯	◯	◯	◯	◯	Muito Útil

85. Alguma vez utilizou uma aplicação semelhante? Se sim, qual? *

86. Comentários sobre funcionalidades

Que outras funcionalidades consideraria importante incluir, etc.

87. Comentários sobre a componente visual

Apresentação da aplicação, interface, layout dos painéis, etc.

88. Comentários sobre a componente auditiva

Qualidade sonora, manipulação de parâmetros, imersão auditiva, etc.

89. Outros comentários/sugestões

Aproveite este campo para deixar outras opiniões ou sugestões que considere relevantes.

I

Supporting Technologies

This annex lists (in alphabetic order) other software tools and resources that helped the development of this dissertation. It is explained their use and presented the version used.

Audacity Free, open source, cross-platform audio software (Version 3.0.0.0). Used to edit, cut, and normalize some of the audio clips used during the historical scenes building process and during the testing phase.
Available at: https://www.audacityteam.org/

Balsamiq Cloud Balsamiq Wireframes (Version cloud-1484. Editor Version: 3.7.4) is a fast and approachable low-fidelity wireframing software. Used to sketch the interface for Immerscape's first prototype (Appendix B).
Available at: https://balsamiq.cloud/

Canva It is a graphic design platform, used to create media graphics, documents and other visual content. The application includes free templates and free graphic assets for users to use. It was used to create many graphic UI elements for Immerscape, to create the storyboards (Appendix A), and create or edit many of the images present throughout this document.
Available at: https://www.canva.com/

GarageBand (2020 Version 10.4.2). It is a line of digital audio workstations for macOS, iPadOS, and iOS devices that allows users to create music or podcasts. Used to edit, cut, and normalize some of the audio clips used during the historical scenes building process and during the testing phase.
Available for MacOs at: https://www.apple.com/mac/garageband/

GitHub Desktop (2017 Version 2.9.4 x64). GitHub Desktop is an application that enables you to interact with GitHub using a GUI instead of the command line or a web browser. Used for version control and backup during the development process.
Available at: https://desktop.github.com/

Google Forms It is a survey administration software included as part of the free, web-based Google Docs Editors suite offered by Google. It was used to create the user study testing form (Appendix C).
Available at: https://www.google.com/forms/about/

Microsoft Excel (2021 Version 16.54 Student Pack). It is a spreadsheet developed by Microsoft. It was used in the context of this dissertation to analyze the results of the user study and create the graphs presented in chapter 5.
Available at: https://www.microsoft.com/en-us/microsoft-365/excel

Microsoft PowerPoint (2021 Version 16.54 Student Pack). It is a slide presentation program currently owned by Microsoft. It was used to create the icons for Immerscape's interface.
Available at: https://www.microsoft.com/en-us/microsoft-365/powerpoint

NAudio (Version 2.0.0). NAudio is an open source Audio and MIDI library for .NET written by Mark Heath. It was tested to use to build the Audio Recorder component but ultimately not used in the prototype.
Available at: https://github.com/naudio/NAudio

Resonance Audio for Unity (Version 1.2.1). It is a multi-platform spatial audio SDK, delivering high fidelity at scale. This powerful spatial audio technology is critical to realistic experiences for AR, VR, gaming, and video. It was used in Unity to create immersive audio.
Available at: https://github.com/resonance-audio/resonance-audio-unity-sdk/releases

SavWav (2012) It is a C# script to save an AudioClip as a .wav file in Unity3D. It was used as the basis to develop the Audio Recorder Component. Provided by Calvin Rien on GitHub.
Available at: https://gist.github.com/darktable/2317063

TextMeshPro Package for Unity (Version 2.1.1, July 2020). TextMeshPro is the ultimate text solution for Unity. It is the perfect replacement for Unity's UI Text and the legacy Text Mesh. It was used to build most of the user interface elements.
Available at: https://docs.unity3d.com/Manual/com.unity.textmeshpro.html

Unity (Version 2019 4.19f1). Unity is a cross-platform game engine developed by Unity Technologies. It was used as the main development environment, using its 3D features.
Available at: https://unity3d.com/get-unity/download

Unity ProBuilder Package (Version 4.4.0, August 2020). ProBuilder is a unique hybrid of 3D modeling and level design tools, optimized for building simple geometry but capable of detailed editing too. ProBuilder was used to build the simple geometry shapes that represent some of the pre-defined 3D environments.
Available at: https://unity.com/features/probuilder

Unity ProGrids Package (Version 3.0.3 preview, March 2020). Advanced grid and object snapping for the scene view. ProGrids is a dynamically placed grid, available on all 3 axes and at any position in the scene. In addition, ProGrids is completely customizable. Used to make the ground element of the environments.
Available at: https://docs.unity3d.com/2020.1/Documentation/Manual/progrids.html

Unity Recorder Package (Version 2.5.5, March 2021). The Unity AOV Recorder allows you to capture specific render pass data, including audio. The Recorder can only be used in the Unity Editor. It does not work in standalone Unity Players or builds. This package was used for recording the immersive audio generated during the development of the initial solution.
Available at: https://docs.unity3d.com/Packages/com.unity.recorder@2.0

Unity Standalone File Browser (Version 1.2). It is simple wrapper for native file dialogs on Windows/Mac/Linux. It was used in the development of Immerscape for the Audio Object's editing panel.
Available at: https://github.com/gkngkc/UnityStandaloneFileBrowser

Visual Studio Code Version: 1.56.2 (Node.js: 14.16.0). It is a source-code editor made by Microsoft. It was used to create and edit the C# scripts that compose Immerscape's functionalities and components' behavior.
Available at: https://code.visualstudio.com/download

Zoom (Version 5.8.0). Zoom is a secure, reliable video platform powers all of your communication needs, including meetings, chat, phone, webinars, and online events. Zoom was used for many meetings throughout the development process and during the testing phase, using its remote control and audio sharing features.
Available at: https://zoom.us/download