NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

RENATO ANDRÉ DA SILVA VIOLA

BSc in Computer Science

ACCESSIBILITY IN MUSEUMS FOR THE BLIND OR VISUALLY IMPAIRED THROUGH SOUND

MASTER IN COMPUTER SCIENCE AND ENGINEERING

NOVA University Lisbon

Draft: *September 12, 2025*

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTMENT OF
COMPUTER SCIENCE

ACCESSIBILITY IN MUSEUMS FOR THE BLIND OR VISUALLY IMPAIRED THROUGH SOUND

RENATO ANDRÉ DA SILVA VIOLA

BSc in Computer Science

Adviser: Sofia Carmen Faria Maia Cavaco
Assistant Professor, NOVA University Lisbon

Co-adviser: Maria Armanda Simenta Rodrigues Grueau
Associate Professor, NOVA University Lisbon

MASTER IN COMPUTER SCIENCE AND ENGINEERING

NOVA University Lisbon

Draft: *September 12, 2025*

Abstract

In their mission as inclusive spaces of cultural participation and celebration, museums have taken considerable strides in overcoming their historical reliance on sight to adequately accommodate the needs of the Blind and Visually Impaired (BVI).

However, accessibility tends to be an afterthought rather than a priority, and most exhibitions have mainly remained inaccessible to a BVI audience. Meritable as they are, the most common accessibility methods often fail to balance exposure to information and proper artwork engagement, if even available. Finding their independence, mobility, and interpretative access conditioned, in Europe, low-vision people rarely attend these institutions despite enjoying and expressing the desire to experience visual art.

In this dissertation, we propose a remote and interactive approach to BVI-accessible visual art representation, in which spatial audio provides a feeling of immersion and simulates exploration. The devised system is symbiotically divided, with each part catering to a unique audience.

The system includes a 3-Dimensional (3D) soundscape editor for Windows, in which museum curators without prior sound design expertise may quickly develop immersive virtual auditory scenes representative of specific artworks. The generated environments are interactive via a mobile soundscape player, where BVI users autonomously explore their composition through assisted thumbstick movement, directional audio cues, and a vicinity scanning tool. The experience offers room for customization, incorporating controls for sensory load regulation.

Ultimately, our proposal aims to fulfill BVI visitors' informational and aesthetic needs regarding visual art access, promoting their independence while remaining cost-effective.

Keywords: Spatial Audio, Blind and Visually Impaired, Accessible Culture, HRTF, Soundscape

Resumo

Na sua missão como espaços inclusivos de participação e celebração cultural, os museus têm dado passadas consideráveis para ultrapassar a sua dependência histórica da visão e acomodar adequadamente as necessidades dos cegos e deficientes visuais.

No entanto, a acessibilidade tende a ser uma reflexão tardia e não uma prioridade, e a maioria das exposições tem permanecido inacessível a um público invisual e amblíope. Por muito meritórios que sejam, os métodos de acessibilidade mais comuns não conseguem muitas vezes equilibrar a exposição à informação e o envolvimento adequado com a obra de arte, se disponíveis sequer. Por verem a sua independência, mobilidade e acesso interpretativo condicionados, na Europa, as pessoas cegas e com baixa visão raramente frequentam estas instituições, apesar de apreciarem e manifestarem o desejo de experienciar a arte visual.

Nesta dissertação, propomos uma abordagem remota e interativa para a representação de arte visual acessível a pessoas com deficiência visual, em que o áudio espacial proporciona uma sensação de imersão e simula a exploração. O sistema concebido divide-se simbióticamente, em que cada parte se destina a um público único.

O sistema inclui um editor de paisagens sonoras em 3D para *Windows*, onde curadores de museus sem conhecimentos prévios de *design* de som podem rapidamente desenvolver cenas auditivas virtuais e imersivas, representativas de obras de arte específicas. Os ambientes gerados são interativos mediante um leitor de paisagens sonoras móvel, onde os utilizadores com deficiência visual exploram autonomamente a composição das cenas através de movimento assistido por um *thumbstick*, sinais direcionais de áudio e uma ferramenta de análise da vizinhança. A experiência oferece espaço para alguma personalização, incorporando controlos para regulação da carga sensorial.

Fundamentalmente, a nossa proposta visa satisfazer as necessidades informativas e estéticas dos visitantes de baixa visão no que respeita ao acesso às artes visuais, promovendo a sua independência e mantendo-se eficaz em custo.

Palavras-chave: Áudio Espacial, Cegos e Amblíopes, Cultura Acessível, HRTF, Paisagem Sonora

Contents

List of Figures

Acronyms

2D 2-Dimensional *(pp. 7, 15, 28, 30)*

3D 3-Dimensional *(pp. i, ii, v, 3, 4, 7, 8, 13–15, 17, 20, 22–30)*

AR Augmented Reality *(pp. 2, 18, 19)*

BVI Blind and Visually Impaired *(pp. i, 2–4, 13, 15–18, 20–22, 26–29)*

HRTF Head-Related-Transfer-Function *(pp. v, 8, 11–15, 18–22, 25, 26, 28, 30)*

ILD Interaural Level Difference *(pp. v, 10, 11)*

IPD Interaural Phase Difference *(p. 9)*

ITD Interaural Time Difference *(pp. v, 9–11)*

SAT Spatial Awareness Tool *(pp. 4, 14, 20, 22–24)*

VR Virtual Reality *(pp. 2, 24)*

1

Introduction

This chapter introduces the problems this dissertation addresses and establishes the motivation for exploring spatial audio as a tool to improve accessibility in art museums. It outlines the objectives of the research and the expected contributions and provides an overview of the proposed solution.

In a colorful world filled with life and the most varied shapes and patterns, each of our five senses is vital in obtaining information about our surroundings [20]. Out of these, vision is the most dominant, as about 80% of what we learn about the world and the impressions we perceive are through our sight [11]. It is so valued that not only do 77% of people state that it is their most important sense, but they would also rather live shorter lives than longer ones without their sight [47, 17].

Unfortunately, sight is not a universal given. According to the International Agency for the Prevention of Blindness Vision Atlas [7], as of 2020, approximately 1.1 billion people worldwide were living with some form of vision loss. Over the last few decades, there has been a decrease in vision loss prevalence in proportion across the population. However, the absolute numbers have increased and are not dwindling anytime soon. On the contrary, vision loss is expected to grow across all categories, with projections indicating that by 2050, approximately 1.8 billion people worldwide will experience visual impairment [8], marking a 55% increase in vision loss.

If inclusivity and disability rights alone are not enough motivation to care for this community, the World Health Organization [35] states that everyone, if they live long enough, will experience some eye condition in their lifetime in need of proper care. As such, the concerns of the visually impaired community should in fact be everyone's concerns as well.

1.1 Motivation

Art is a universal form of human expression, whether cultural, creative, somewhere between, or something else entirely. From their very conception, museums have traditionally

been one of, if not the most prominent, ways to not only access such art but also celebrate and share it across generations. These institutions are mostly known for being the custodians of history, heritage, and artistic expression.

It is the very mission of a museum to be an inclusive and accessible space of cultural participation and a shared human experience. However, there are limitations to this statement. Exhibitions are historically visual, as most available art is specifically designed to be consumed that way. This poses a considerable barrier for blind and visually impaired people in terms of access to information and even hinders their mobility and independence within an exhibition [12, 26, 49, 50, 51].

Although the visually impaired are still not appropriately accommodated to this day, there have been strides in the right direction, one of which is in the form of legislation advocating for their right to participate in cultural life [22, 26, 29] – article 30 of the United Nations Convention on the Rights of Persons with Disabilities [34].

In recent times, museums too have made considerable efforts in adapting to the needs of the visually impaired aside from mobility and navigation, gradually becoming more inclusive and participatory spaces [4, 22, 26, 29, 37].

Though institutions are now more accessible than they ever were, most museums are still largely inaccessible to a BVI audience, as accessibility is more of an afterthought than a priority [10, 22, 37, 49]. As a result, low-vision people rarely attend these institutions - in Europe, only 5.5% of them do [50, 51]. The aforementioned visual centricity of exhibitions is a primary reason for this statistic, as it significantly limits not only their mobility but also their access to information, which in turn dramatically hinders their independence as well [12, 26, 49].

Such a low percentage of museum attendance is particularly demoralizing, since BVI people enjoy and express the desire to visit galleries and experience visual art [4, 10, 22, 24, 26]. However, they must be given the proper access to do so [22] and do not want to constantly rely on others [4, 24].

While the accessibility methods employed at museums are valuable, these tend to emphasize descriptive information over actual appreciation [24, 26, 29]. Good accessibility requires a fine-tuned balance of exposure to information and engagement with the artwork. Achieving this balance is especially important for visually impaired visitors, as it allows them to develop meaningful connections to art, going beyond the pure intellectual understanding of it [29]. Approaches combining music and soundscapes have shown promise but can be time and resource-intensive [24].

Spatial audio, haptics, Virtual Reality (VR), Augmented Reality (AR) and other technological advancements have produced increasingly immersive experiences, creating opportunities for remote artistic appreciation [13, 27, 43, 52]. People can experience art from the comfort of their homes or anyplace else thanks to digital platforms and virtual tours, which can dissolve physical constraints. This enables them to interact at their own pace and unshackle themselves from social tension despite the risk of less engagement than in person [14, 26].

The potential of virtual environments to provide a variety of interaction modalities and content has been thoroughly investigated, and sound-based approaches are especially important for blind and visually impaired people [43, 52]. By mimicking environmental cues, spatial audio improves spatial orientation and makes it easier for users to move freely and independently in virtual environments, helping them experience virtual worlds [43]. Frequently enhanced with 3D effects to improve spatial perception and experience, 3D audio improves navigation and immersive engagement in museum and gallery settings [52].

Primarily supporting the above-mentioned technologies, the smartphone is a suitable vehicle for accessibility as about 54% of the global population owns at least one [25], and it is rich in accessibility features among several others [5]. This study draws inspiration from several implementations of accessible technologies to develop a remote mobile and 3D sound-based solution that addresses BVI visitors' informational and aesthetic needs.

1.2 Problem Description & Objectives

As it was briefly alluded to in section 1.1, there are some limitations to the accessibility methods usually active at museums [10, 12, 22, 37]. Tours and workshops directed at a BVI audience are infrequent and inconsistent, usually must be reserved in advance, and are usually only available on specific dates or time slots. Though more common in museums, audio descriptions are primarily designed with normovisual people in mind, mainly focusing on interpretation and historical context, not accessibility. Braille-based brochures leave much relevant information aside, and braille proficiency is generally low [12].

The preferred form of interaction with artwork for BVI individuals seems to be the tactile approach since it allows them to feel the artwork up close and personal and sense its various features at a low level [24, 26]. However, high-level information about the piece is quite limited, and preservation efforts, intellectual access barriers, and the still prevalent visual centricity of exhibitions make these types of programs a rare occurrence among museums [26].

Drawing inspiration from the use of spatial audio in other research and accessible games to convey space and emotion [19, 31, 32], our approach focuses in assigning 3D sounds to specific elements of an artwork. The spatialized audio cues serve to aid BVI users in navigating and interpret the scene. Museum curators are responsible for generating these sound environments, ensuring these accurately represent the art pieces.

To facilitate this process, we propose a tool that allows museum curators to build a spatial audio environment representing art pieces. It is an intuitive and user-friendly 3D soundscape editor with support for immersive audio and interactivity, not requiring prior experience in sophisticated tools with a harsh learning curve. The generated environments are then made available via an Android mobile application, where BVI visitors can navigate a simplified top-down map-like view of the scene with contrasting elements, using virtual thumb joysticks to define direction and movement. One such joystick is implemented as a directional scanner for precise exploration. With a smartphone coupled with headphones,

BVI users "move" within the environment and explore immersive 3D sound, simulating the experience of physically approaching or moving away from different parts of the artwork, using basic and familiar controls.

1.3 Solution Overview and Contributions

This research is expected to yield the following contributions:

- An intuitive tool for creating and editing 3D auditory environments representing artwork. The tool will require no prior computer science or audio design expertise, and thus, can be used by museum curators to generate immersive auditory scenes. Curators may toggle between two modes within the application, the default one being development and the other emulating the scene's interactivity and navigation from a mobile end-user's perspective.
- A mobile application enabling BVI users to explore the generated artwork recreations through immersive audio and a directional scanning Spatial Awareness Tool (SAT), providing an experience that is both aesthetic and informational while promoting independence with simple controls. It will feature a graphical interface with large high-contrasting elements, to properly accommodate users with residual sight.
- A widely available approach to accessibility through the use of smartphones and headphones, allowing for both remote and on-site interaction with art.

1.4 Document Structure

There are four main chapters to this document:

- **Chapter 1 - Introduction:** Introduces the problems this dissertation addresses and establishes the motivation for exploring spatial audio as a tool to improve accessibility in art museums. It outlines the objectives of the research and the expected contributions and provides an overview of the proposed solution.
- **Chapter 2 - Background:** Presents the foundational concepts most relevant to understanding the work to be developed. It covers sound principles from its definition, perception, transmission, and spatial localization. Additionally, the definition and purpose of soundscapes are addressed, and most importantly, the role these play in accessibility for the blind and visually impaired.
- **Chapter 3 - Related Work:** Reviews relevant research and applications mainly related to immersive spatial audio and BVI accessibility, including technologies and approaches to address said accessibility. It is divided into sections, each focusing on a specific topic related to the dissertation's theme. Each section starts with a brief overview of related studies and projects displaying the current state of the

art. It is then followed by subsections where projects of particular relevance to this dissertation are explored in detail, from implementation to findings, and finally, how they relate to our work and what is to be learned from them.

- **Chapter 4 - Proposed Solution:** Provides a detailed overview of the proposed system and briefly exposes the solution's validation methodology. It also addresses the expected technological stack and the envisioned plan for the system's development. The work schedule is split into five distinct and concisely explained tasks mapped in a Gantt chart.

2

Background

This chapter presents the foundational concepts most relevant to understanding the work to be developed. It covers sound principles from its definition, perception, transmission, and spatial localization. Additionally, the definition and purpose of soundscapes are addressed, and most importantly, the role these play in accessibility for the blind and visually impaired.

2.1 What Is Sound And How It Is Perceived

As a physical phenomenon, sound is enabled by the vibration of a body with the properties of inertia and elasticity (which are attributes of nearly every object). Any vibration can produce sound if it meets the requirements for moving a body back and forth. The simplest of vibrations can be characterized by a sinusoid (Figure 2.1) and is the elementary unit for all possible vibrations.

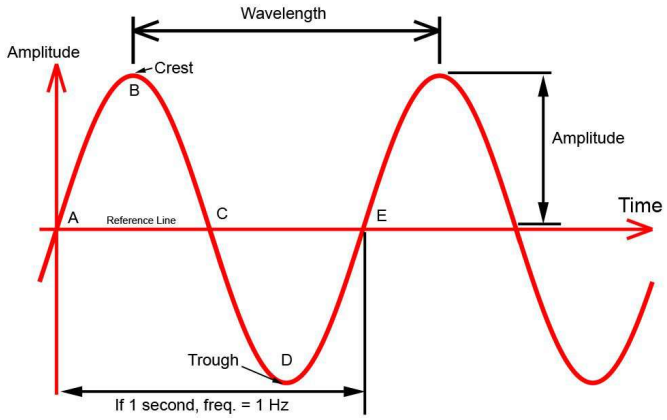

Figure 2.1: Graphical representation of a simple sine wave (Adapted from [3]). © 2010 - Bill Ansell.

Any vibration can be broken down into a composition of sine waves – a Fourier series, each uniquely identified by frequency, amplitude, and starting phase. Describing

a complex vibration by deriving the characteristics of its composing simple vibrations is named a Fourier analysis [53].

Without delving into the anatomical details, hearing starts once a sound wave vibrates our eardrum. After passing through the outer, middle, and inner ear, what reaches our auditory nervous system is no longer a mechanical vibration but a nervous impulse, which our brain now interprets. Perceptually, the changes in amplitude of a sine wave tend to be experienced as loudness, while changes in frequency are labeled as pitch [53].

2.2 Sound Propagation

As mentioned in section 2.1, for a sound to reach our ears or any other point, it must first travel through a medium with the properties of elasticity and inertia. That is to say that, for example, it can travel through solids, liquids, and gases but not through a vacuum [53]. The speed at which it propagates may vary with the temperature and density of the medium, which in air is around 343 meters per second (at 20º C) [39].

In the air, the very presence of its randomly moving molecules originates a static pressure, which, when disturbed by the vibrations of a sound source, leads to zones of alternating pressure (Figure 2.2 illustrates this) – where the molecules cluster more tightly is called an area of condensation. In contrast, a significant spread in molecule placement designates an area of rarefaction.

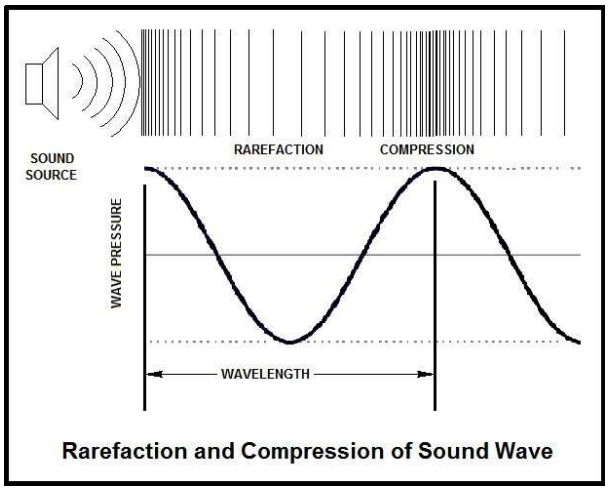

Figure 2.2: Representation of rarefaction and condensation zones (Adapted from [44]).

Sound waves propagate in all directions from a source (circularly in 2-Dimensional (2D), spherically in 3-Dimensional (3D)) and, while traveling, may come across several forms of interference, such as reflection, absorption, diffraction, and refraction. For example, an obstacle with a size similar to that of the sound's wavelength may produce an area past the object where wave magnitude is significantly reduced or even completely absent – a sound shadow. In addition, the intensity of a sound decreases quadratically with the distance to its source - the inverse square law [53].

2.3 Sound Localization

As sound has no intrinsic spatial dimensions, how we perceive spatial cues is a product of our auditory system's capability to process the physical properties of sound that correspond to spatial position [53]. There exist three spatial dimensions in which one can localize sound: the horizontal plane, commonly referred to as the azimuth, the vertical plane (elevation), and the distance (range) [39, 53]. These are properly illustrated in Figure 2.3.

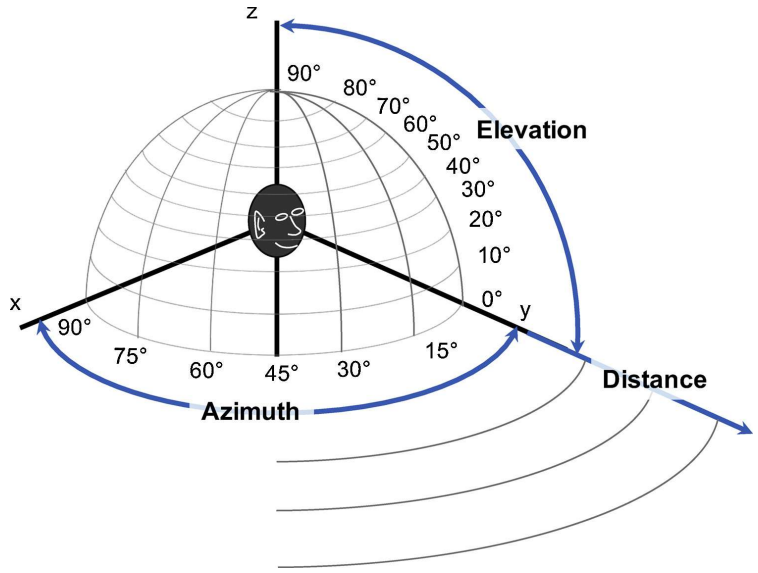

Figure 2.3: Polar coordinates used to locate a sound source in a 3D space centered on the listener (Adapted from [38]).

Most of our spatial perception heavily relies on our binaural hearing - our ability to interpret and locate what we hear from both ears. While monaural cues contribute to spatial hearing, namely for vertical localization and depth perception, the most important mechanisms for perceiving directional sound are binaural cues – used for localization in the horizontal plane, consisting of the interaural differences of time and intensity [41].

Via these differences in signal reception between both ears, time and level are key physical properties that enable sound localization along the azimuthal plane and will be further addressed in section 2.3.1. Another important property is the sound's spectral shape relevant for vertical localization, and it is impacted by interference phenomena such as reflection, diffraction, and absorption caused by a listener's physical features [39]. These alterations are captured by a Head-Related-Transfer-Function (HRTF), which we will delve into in section 2.3.3. Though not as much is known about perceiving the distance of a sound source, it is mainly influenced by the sound's loudness and early reflections from nearby surfaces [53].

2.3.1 Interaural Differences

The differences between the signals our two ears receive are coined as interaural differences. As briefly mentioned in section 2.3, they are our auditory system's primary mechanisms for localizing sound along the azimuthal plane. Figure 2.4 illustrates them clearly.

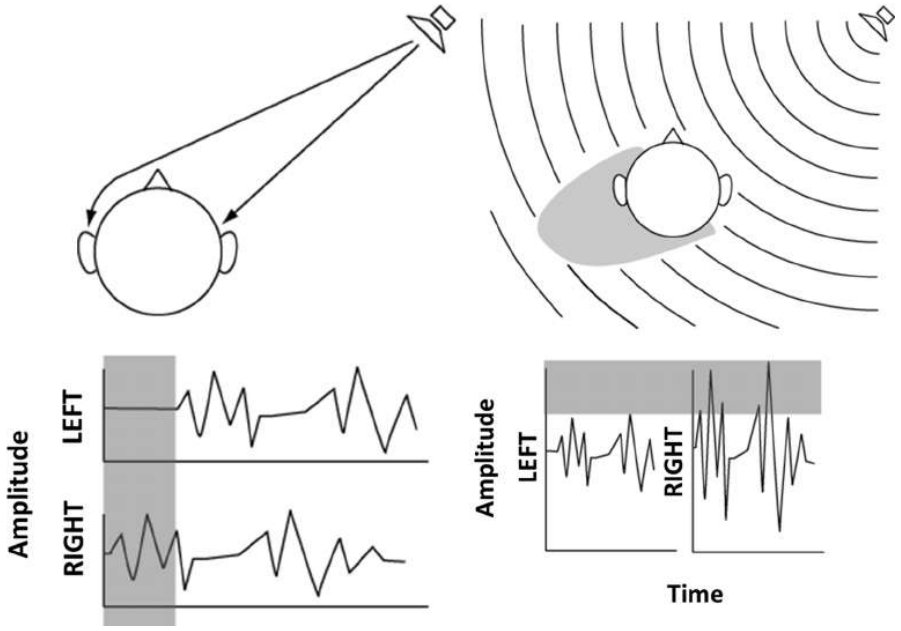

Figure 2.4: Interaural Time Difference and Interaural Level Difference, respectively (Adapted from [55]).

These mechanisms depend on the sound signal's nature and can be influenced by environmental cues that introduce conflicting information [41]. One may classify interaural differences into two primary categories:

- **Interaural Time Difference:** The Interaural Time Difference (ITD) refers to the delay in sound arriving in one ear compared to the other.

  Binaural delay is the designation of the maximum time delay between the ears, and it enables the resolution of a source in the direction of the ear that first heard it – the precedence effect. The sound source's angle of incidence affects the additional distance that the wave must travel to the farthest ear, thus impacting the binaural delay. The brain usually localizes the sound towards the earliest source for similar sound sources in different locations.

  When considering sinusoidal signals, ITDs may be expressed as Interaural Phase Differences and are fundamental for locating low-frequency sounds where the wavelength is large enough for phase differences to be noticeable. For higher frequency sounds, the interpretation for both ITD and Interaural Phase Difference (IPD) becomes ambiguous [53], and one can no longer tell which ear is leading or lagging [39, 41].

- **Interaural Level Difference:** The Interaural Level Difference (ILD) pertains to the difference in intensity of the same sound between the two ears.

  The intensity of the stimulus at the ear closer to the source is slightly more significant due to proximity (as explained by the inverse square law, mentioned at the end of section 2.2) [41, 53]. However, the extra distance the wave travels to the farther ear is negligible and the intensity differences minimal [41].

  The primary contributor to ILDs is not proximity but the attenuation caused by the head's sound shadow for high-frequency sounds. A higher frequency implies a shorter wavelength and a greater sound shadow, thus a more noticeable difference in level [53]. The same cannot be said for low frequencies, at which the head is not a decent barrier to sound [41].

These mechanisms complement each other in providing fundamental cues for localizing sound sources in the azimuthal plane. However, these are not without limitations, some of which we will address in subsection 2.3.2.

2.3.2 The Cone Of Confusion

Though the interaural differences are crucial and effective in locating sound along the horizontal plane, one exception is the cone of confusion [39, 53]. When a listener's head is stationary, detecting whether a sound is coming from the front or behind, or even its elevation, is far from obvious. The difficulty arises because some locations are producing the same interaural differences [39, 53]. The plane in which these ambiguous points lie is dubbed the mid-sagittal plane [53], and it forms the so-called cone of confusion (see Figure 2.5), its axis being the interauricular line [39].

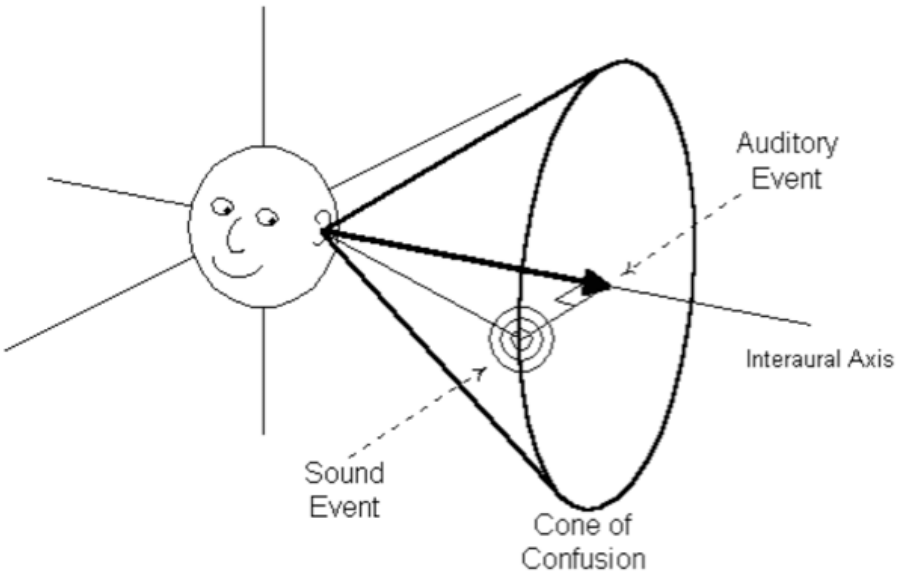

Figure 2.5: Cone of confusion, in which ITDs and ILDs are indistinguishable (Adapted from [30]).

Any sound deriving from this cone's circumference produces neither time nor level differences, while any sound coming from elsewhere inside the cone has at least one other point mirroring its interaural differences [39]. Every sound source location has its cone of confusion describing the other possible locations producing the same interaural differences [53]. The most intuitive way to reduce this ambiguity is our natural head movements [39, 41, 53]. Simply leaning our head or rotating it tears down the original cone of confusion by altering the ITD and ILD values, introducing additional binaural and spectral cues – enhancing localization.

If head movement is not an option, our auditory system relies on the spectral cues derived from the HRTFs [39, 53] (we will address them in some detail in section 2.3.3), which are not only critical in determining the vertical direction of sound but also in distinguishing between the ambiguous sounds along the azimuthal plane [39].

2.3.3 Head Related Transfer Function

As explained in section 2.3.2, due to the cone of confusion, interaural differences alone are not enough to adequately determine the direction of a sound, especially along the vertical plane. The path of a sound to our ears is rarely uneventful, as our very anatomy induces significant changes in its spectrum. Our head, torso, and, more importantly, the structure of the pinna act as sound shadows in the sound's trajectory, which are particularly accentuated for high-frequency sounds (since the wavelength is closer to the small size of the pinna) [53]. Furthermore, the listed physical features interfere with an incoming sound wave through reflection, diffraction, and absorption, attenuating or delaying specific frequencies composing it [39, 41].

These effects, caused by the filtering action of the pinna and body, are known as spectral cues [53] and are determined monaurally [41]. Depending on both the position of the sound's source as well as its angle of incidence [41, 53], such cues uniquely shape a sound's spectrum at the eardrum level. They are captured by what is called a Head Related Transfer Function (HRTF) [53].

An HRTF is a descriptor for the transformations a sound undertakes on its route to a listener's tympanum due to the hearer's physical features. For complex and high-frequency sounds, it is a highly competent mechanism for estimating the vertical position of a sound. However, it performs poorly for non-complex and low-frequency sounds (likely due to the wavelength exceeding the size of the pinna) [39]. Figure 2.6 shows how it may be measured.

Vertical localization is generally less accurate than its horizontal counterpart. However, coupling spectral cues with the previously mentioned interaural differences is a good approach for resolving the ambiguities within the cone of confusion without the need for head movement [53].

Just as individuals greatly vary in their morphology, so does the shape and size of their pinna and thus, the filtering effect their ears impose on incoming soundwaves.

Figure 2.6: HRTF measurements in an anechoic chamber (Adapted from [28]).

Consequently, each person has a unique set of HRTFs learned throughout their lives [39, 41]. Additionally, research has found that amplifications or attenuations in particular regions of the frequency spectrum seem to correspond to specific sound source positions [41].

From a practical perspective, generalizing the spectral characteristics that compose HRTFs across a vast demographic simplifies their implementation process while ensuring reasonable accuracy for many listeners. However, to no surprise, an individual's best HRTFs are their own [41], providing the most accurate spatial perception. Nonetheless, there have been studies in which subjects are fed audio signals through HRTFs other than their own [39, 41] – to find their ability to localize sound significantly hampered [41]. After some time, the participants appear to begin learning the HRTFs they have been given and eventually regain normal vertical localization [39]. Some people are known to localize audio better than others, so their HRTFs tend to be considered more applicable for general use [41].

Delivering audio stimuli over headphones is convenient and establishes a controlled environment[53]. However, certain headphone types induce a feeling of the sound emanating from inside the head, rather than the natural tridimensional externalization (azimuth, elevation, and range) we are accustomed to [41, 53]. In short, headphones fail to deliver on the nuances captured by HRTFs since they emit sound directly to our ear, while a real outside source would first be affected by our physical features. Fortunately, simulating externalization is relatively intuitive: recreate a sound with all the spectral complexity of a real-world one by applying HRTF-based filtering and only then transmit it [53].

2.4 Soundscape

By R.M. Schafer's [45] definition, a soundscape englobes the various acoustic elements within an auditory environment in its totality and mainly focuses on how this environment

is humanly interpreted [2, 9, 16]. While a soundscape can be a physical phenomenon such as an acoustic environment, the emphasis on human perception distinguishes this term from being just that, as it can also be thought of as a perceptual concept [2, 9]. Its perceptual quality allows listeners to attribute meaning to their surroundings, immersing themselves personally.

Interacting with a soundscape yields distinct results for different individuals, as people's preferences and expectations greatly vary [16], especially according to the space and context of the environment [9]. Among other socially relevant values, soundscapes provide a sense of place and atmosphere, express cultural and historical values, and promote human connection to nature by enhancing the aestheticism of the experience.

The fidelity of a soundscape is mainly evaluated on how different sound sources are perceived and the level of noise present [16]. So, too, is its general quality evaluated on several descriptors, some of which being noise annoyance, pleasantness, and quietness [2].

In the context of this dissertation, we are particularly interested in the role of the soundscape as an accessibility tool for Blind and Visually Impaired (BVI) individuals. Since their reliance on vision is limited, if any, they are not only more sensitive to auditory cues but can also draw more information from acoustic signals than sighted people [42]. For these individuals, a high-fidelity soundscape is as functional as it is aesthetic since, from its acoustical information, they can derive information regarding the morphology and movement of objects. A fine example is rainfall, as it contours environmental elements, enhancing spatial awareness and connection to the environment [42].

2.5 Discussion

This chapter aims to provide a structured understanding of sound, building up to sound localization, the most critical aspect of this dissertation. We start by laying the groundwork through concise overviews of the most fundamental concepts regarding sound, such as what it is and how humans perceive it(section 2.1), and how it travels (section 2.2).

We then detail sound localization (section 2.3), introducing the three spatial dimensions where sound may be localized and exploring two techniques on which spatial audio heavily relies. Section 2.3.1 covers the interaural differences of time and level, which are essential for the binaural determination of a sound source's direction in the horizontal plane. Additionally, section 2.3.3 explains the notion of HRTF, its relevance for the monaural perception of 3D sound through spectral cues, and how it complements the interaural differences by resolving the cone of confusion (addressed in section 2.3.2).

The chapter closes on a less technical note. Section 2.4 focuses on the soundscape as a means to evoke meaning and a sense of atmosphere within an auditory environment, briefly touching on how it is useful for BVI individuals.

Chapter 3 presents some studies related to our work where most, if not all, of the principles addressed were put into practice. In the research, we see the vast applications of sound within cultural settings and elsewhere, mainly studying its effectiveness in

enhancing accessibility and engagement. 3D audio, interactive soundscapes and sound-based Spatial Awareness Tool (SAT)s are examples of such uses.

Informed by the implementation and findings of related studies, we incorporate the presented concepts into this thesis, with spatial audio standing at the core of our proposal. Employing the concepts of interaural differences and HRTFs, we intend to non-visually convey the spatial characteristics of artworks through BVI-accessible interaction and exploration. In addition, we look to evoke emotional and aesthetic engagement beyond factual description, leveraging soundscape theory to this end.

3

Related Work

This chapter reviews relevant research and applications mainly related to immersive spatial audio and BVI accessibility, including technologies and approaches to address said accessibility. It is divided into sections, each focusing on a specific topic related to the dissertation's theme. Each section starts with a brief overview of related studies and projects displaying the current state of the art. It is then followed by subsections where projects of particular relevance to this dissertation are explored in detail, from implementation to findings, and finally, how they relate to our work and what is to be learned from them.

3.1 Immersive Audio Experiences in Cultural Environments

Audio is a widely used vehicle for delivering immersive experiences in cultural environments. The following two studies exemplify this statement and are centered around spatial audio since it is a central theme of this dissertation.

Focused on recreating the city of Évora's culturally rich historical soundscapes, Ferreira [19] created Immerscape, a tool aimed at non-expert users for generating 3D audio scenes, utilizing HRTFs to spatialize sound. We cover Immerscape in further detail in subsection 3.4.1. Kabisch et al. [23] used motion tracking, image analysis, and sonification alongside real-time directional sound to integrate panoramic visual landscapes with spatialized audio, presenting such research in an interactive art exhibit.

3.1.1 Eyes-Free Art

Looking to enhance the accessibility of visual art to BVI individuals and go beyond the shortcomings of the typical audio descriptions or guides, Rector et al. [37] designed Eyes-Free Art. Eyes-Free Art was a novel approach to sonically interacting with 2D art, aiming to be both aesthetically stimulating and engaging.

It is a carefully crafted proxemic audio interface that, mirroring the conventional intuition of visual proxemic interfaces, renders more detail as proximity to the piece increases. In order to define such proximity, a Microsoft Kinect device was used to track

the user's position and movements to determine the distance to the painting and if it is facing toward it. The audio interpretation a user hears varies according to the proxemic zone in which it finds itself, of which there are four distinct and equally sized ones (Figure 3.1). Upon entering any zone, the user is verbally alerted to where it is and may continue moving between zones, spending as much time as he/she wants in each.

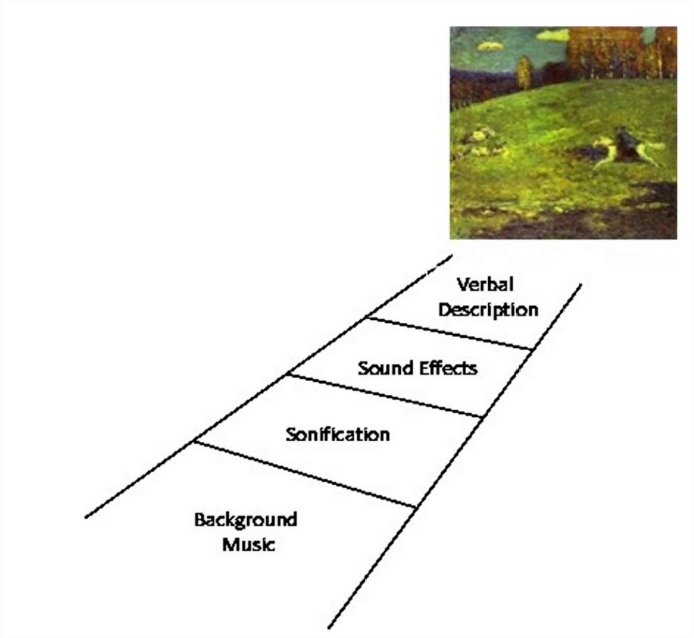

Figure 3.1: The four zones of the Eyes-Free Art proxemical interface. Image is *The Blue Rider* from Wassily Kandinsky (Reproduced from [37]). © 2017 Copyright held by the owner/author(s).

The furthest zone consists only of background music that sets the piece's mood. It is followed by a sonification area that communicates the painting's chromatic diversity through musical features. The second-closest zone, sound effects, highlights the painting's literal aspects, such as the type of objects and their spatial correlation. The final and most detailed zone consists of a manually curated verbal description.

Some initial interviews were conducted, from which Rector et al. noted the importance of using commodity technology (promoting control and independence) and including both the literal and subjective aspects of a painting. A final evaluation with 13 BVI participants attested to the success of this implementation, as patrons felt immersed and had a rich experience interpreting the artwork.

Eyes-Free Art resonates with the current dissertation in several ways. Most importantly, it integrates zones of differing detail according to proximity, allowing users to explore at their own pace and at their desired level of detail. This closely aligns with our goal of promoting independence and interactivity. Furthermore, it also addresses mapping visual elements to sound, though the current work focuses on spatial audio rather than sonification.

3.1.2 Audio-augmented museum experiences with gaze tracking

Aiming to enrich the perception of landscape and genre paintings, Yang et al. [52] track a visitor's gaze and spatialize sounds for drawn objects and scenes within the paintings. The system personalizes the audio output based on the user's gaze; the system amplifies the sounds directed at the viewer's focal point, attenuating the rest.

Gaze and pose tracking required an eye tracker and a laptop connected to a backpack. Additionally, headphones were used for spatial audio playback (Figure 3.2). Sound propagation was dynamically simulated according to a user's gaze and pose via the Google Resonance Audio SDK[1]. At the same time, the Unity[2] game engine was employed to model the room and map the various sound sources.

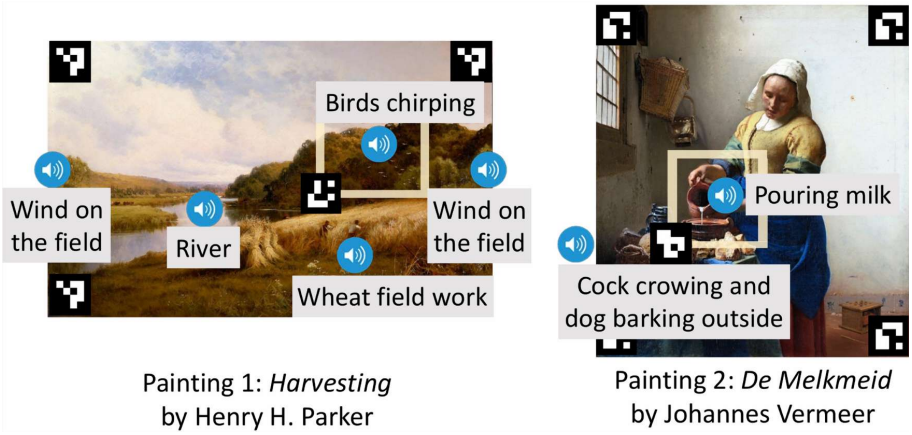

Figure 3.2: Two of the paintings used in the application. Blue audio icons represent virtual sounds accurately spatialized relative to the user (b). (Adapted from [52]). © 2019 Copyright held by the owner/author(s).

A user study with 14 young adults revealed some challenges regarding the consistency of eye-tracking and differences in preference across individuals, such as the amplification of sound and the smoothness of its adjustment. Overall, the experience was still positively received, as it helped most users focus on areas of interest, some even feeling guided by their gaze.

While gaze tracking is a mostly intuitive approach to dynamic audio spatialization and interactivity with artwork, it is not the most appropriate technique in our work, focusing on BVI users. Nonetheless, this study provides valuable insights into our own. Akin to our proposal's proximity-based 3D audio, it narrates a painting's visual elements by spatially embedding sounds to specific points of interest and dynamically adjusting their intensities. Furthermore, there is relevance in learning from the challenges highlighted by Yang et al., namely in ensuring smooth audio transitions, responsiveness, and accommodating some degree of personal preference through personalization.

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

3.2 BVI Accessibility in Multisensory Experiences

Multisensory experiences have been shown to have the potential to improve accessibility and independence for the blind and visually impaired by providing alternative ways of perceiving essentially visual content.

Li [27] developed an audio-only inclusive prototype for navigating Augmented Reality (AR) content without relying on visual cues, incorporating spatialized audio to provide intuitive feedback on object proximity and spatial relationships. Meanwhile, Cavazos Quero et al. [12] implemented a touch-sensitive multimodal guide providing localized audio descriptions based on touch, which promoted independence in both the exploration and interpretation of artworks. Banf and Blanz [6] presented a system using touchscreens that allow a visually impaired user to explore an image and receive audio feedback corresponding to its local content. This procedure employed computer vision and machine learning algorithms to sonify image features from low levels, such as color and texture, to high-level object recognition.

3.2.1 Navmol

Navmol [18] is a molecular browser and editor specifically designed for blind and visually impaired users, aiming to provide BVI accessibility in higher chemistry education. Via a speech synthesizer, it provides an auditory portrayal of the atomic composition of complex molecular structures. Such configurations are bidimensionally depicted using the analog clock metaphor, consisting of mapping directions to the positions of a clock, and are well known among the BVI community. As some users retain some degree of sight, the program has a simple graphical interface (Figure 3.3) with them in mind, visually displaying the selected atoms and some molecular contours.

Rodrigues [40] merged the simple clock analogy with the application of HRTFs on the auditory signal generated by the existing Navmol program at the time (version 2.0) in order to create realistic directional sound cues, perceived to derive from where the atom is positioned. Usability tests demonstrated the efficacy of this integration, with users achieving an average task accuracy of 95.7% in identifying and navigating molecular structures.

Knowing that HRTF performance may greatly vary across users due to inter-individual morphological differences, Rodrigues conducted a study on the performance of 53 distinct HRTF measurements. It confirmed a significant variation in performance across different HRTF datasets for individual users. The five most consistently well-performing measurements – KEMAR, CIAIR, IRC05, IRC25, and IRC44, were then selected for use in one additional study, motivated by the significant variation in performance across different HRTF datasets for individual users. The necessity of allowing users to select their preferred HRTF dataset was apparent, and Navmol was updated accordingly.

Rodrigues' work very much aligns with the goals of this dissertation, as it highlights

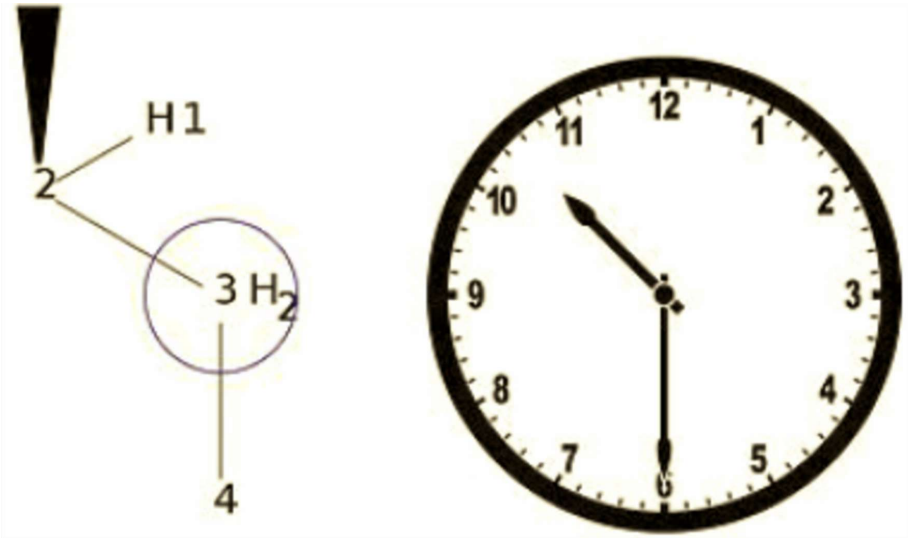

Figure 3.3: Navmol's clock system, where carbon-3 has neighbors at 6 o'clock and 10 o'clock, respectively carbon-4 and carbon-2 (Adapted from [18]).

not only the efficacy of spatial audio (integrated through HRTFs) in the perception of spatial and structural information but also addresses the suitability (or lack thereof) of specific HRTFs to distinct users, allowing them to tailor their own experience to a degree.

3.2.2 MusA

Ahmetovic et al. [1] proposed MusA intending to address the limitations of traditional artwork accessibility methods, such as audio guides. MusA is a mobile application that leverages AR to provide interactive and accessible descriptions of paintings to low-vision visitors. These descriptions are structured into chapters, each representing a specific artwork area. They are linked to an image overlaid on the artwork with a contour highlighting the described section.

The application features artwork recognition via the mobile's camera, interactive navigation across chapters, and touch-based overlays. It was designed for users with some residual sight, presenting a clutter-free and to-the-point interface (Figure 3.4) compatible with system magnifiers, enlarged fonts, and adjustable contrast filters.

After an initial user testing with low-vision participants identifying some challenges, a second and final iteration of the app incorporated audio and haptic feedback, higher contrast contours, and a zoom-supporting virtual mode designed to replace AR if the user cannot frame the painting continuously. User studies revealed MusA to be a significantly more engaging and user-friendly experience than a traditional audio guide, promoting freedom in exploration. Despite some issues in overlay clarity, MusA was effective in supporting the needs of low-vision people, and they were pleased by the ability to use the

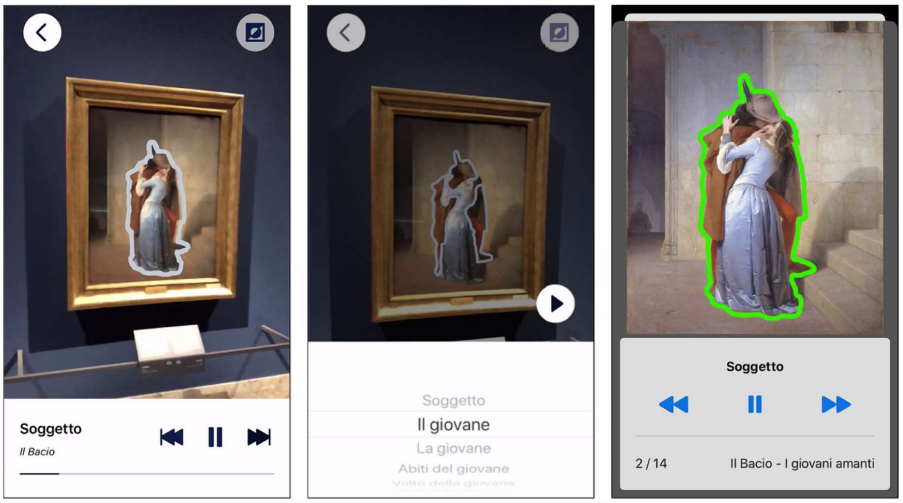

Figure 3.4: Some screens of the MusA app. The first two screens respectively correspond to chapter navigation and selection in its first iteration. The third screen corresponds to the second iteration's virtual mode (Adapted from [1]). © 2021 Copyright held by the owner/author(s).

app at home and on their own devices.

The work of Ahmetovic et al. parallels that of this dissertation in certain aspects, like the chapter navigation available in MusA. The proposed interactive soundscape will offer users control over what details they wish to explore and focus on specific artwork features only with proximity-based interactions. Their findings show that visual feedback through properly contrasting overlays enriches the perception of a painting's structure and details, even for low-vision users. They also show that there exists an appreciation for remote artistic experiences. From them, we take that we must not undervalue visual cues and ensure low visual clutter alongside proper visibility settings, such as contrast and font size.

3.3 Leisurely BVI Inclusive Applications

Leisurely applications such as games designed for BVI individuals aim to merge entertainment with accessibility by integrating multisensory technologies such as spatial audio and haptics. Their goal is to provide experiences as unique as they are inclusive.

Nair et al. [31] introduced a spatial audio-based navigation tool for 3D games, enabling BVI players to explore and create their mental maps of virtual worlds independently. Navstick uses HRTFs for spatialized sound and allows users to scan the contents of their vicinity in specific directions using a thumbstick. Furthermore, Navstick is the directional scanner SAT employed in subsection 3.3.2, and yielded positive feedback. Sánchez and Sáenz [43] designed and evaluated the usability of three distinct interactive 3D virtual environments for visually impaired learners: AudioMUD, AudioVida, and AudioChile. Such environments were navigatable and interacted with through sound, their spatiality

conveyed by spatialized audio and described by voiced narration. Simão [46] proposed Blind Adventure, a mobile audio game aimed at training the orientation and mobility skills of visually impaired children by presenting several challenges requiring physical movement, tracked through the device's sensors. The game's virtual environments were built with the Unity game engine and featured localizable sound cues implemented through HRTFs derived from the SADIE database.

3.3.1 LEAP Tic-Tac-Toe

Drossos et al. [15] adapted an audio-only version of Tic-Tac-Toe to be made appealing and accessible to visually impaired children by empowering it with sonic displays. Simple as it is, Tic-Tac-Toe is a visually reliant game, and spatial perception, game state, positioning, and rules all posed a significant challenge when designing it to be BVI accessible. Additionally, people with residual vision must be accounted for, enabling them to complement their auditory experience with their remaining vision.

To this end, they designed and implemented not only the game itself but also a custom game engine and audio engine specifically intended for the effective development of audio games accessible to the visually impaired. Within the game world represented by the game engine, static objects are the most common type of game object, and sound objects are included.

Carefully ensuring that BVI auditory needs were met, Drossos et al. categorized sound objects into three types that, when collectively employed, enable the development of any stage, be it simple or complex. These are classified as Standard (no interaction with the game; ambient soundtrack is an example), Interruptible (sounds that provide little more than redundant or aesthetic information), and Blocking (sounds carrying important information requiring special attention; interruptible sounds are discarded).

The audio engine functions as the API serving the audio to these objects and handles the sound settings and sonic display implementation. Auditory icons (recognizable real-world sounds) and earcons (synthetic sounds) are used for game state awareness, while binaural processing (utilizing the KEMAR HRTF library) conveys spatial details.

The game's GUI is simple and employs intense contrast alongside significant optical elements (Figure 3.5), accommodating users with residual sight. The auditory interface provides constant feedback on the effects of every action, and players sense direction and three-dimensionality through the localized sound cues. In addition, there are pre-recorded audio instructions for each game component.

In user studies, visually impaired children received the game very positively, though many were experiencing an accessible game for the first time. Drossos et al.'s findings are valuable for this dissertation as they address challenges we will inevitably encounter, such as the need for constant, concurrent, and discernible audio feedback and conveying localization information through the sounds for users to navigate the soundscape. They also address relevant limitations, as not all sounds are equally effective for spatial localization

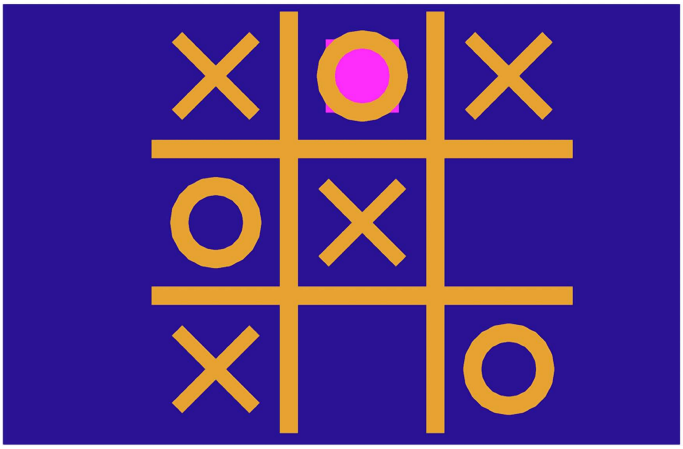

Figure 3.5: Part of the Tic-Tac-Toe game's graphical user interface (Adapted from [15]).

and varying spatial awareness among users.

3.3.2 The Preferred Spatial Awareness Tools for BVI People In Video Games

While the minimap is one of the most employed SATs in video games, crucial even for sighted players in learning the layout of their surroundings, it still has no successful equivalent regarding BVI accessibility. Attempting to bridge this gap, Nair et al. [33] took it upon themselves to tackle the creation of a universal and acoustic BVI-friendly minimap, or more concretely, uncovering the most relevant design factors as well as the merits and limitations of the best acoustic techniques to do so.

Two main questions were at the center of their study, the first regarding the key aspects of spatial awareness valued by visually impaired players in games and the second focusing on the effectiveness of current SATs in supporting said aspects. Intending to delve into both, Nair et al. [32] investigated the design of the four leading SATs (Figure 3.6) in enhancing spatial perception for the visually impaired in a 3D game world, vastly different approaches developed in previous research. These are the:

- *Smartphone Map* - A touchscreen map working in tandem with the game to provide spatial information through sound effects and text-to-speech

- *Whole-Room Shockwave* – When triggered, emits 3D sounds from certain objects based on their distance, simulating a refined and customizable echolocation (using Steam Audio[3]'s built-in HRTF)

- *Directional Scanner* – Allows players to survey a surrounding direction by tilting the right thumbstick towards it, announcing the first object in line-of-sight with directional sound

[3]https://valvesoftware.github.io/steam-audio/index.html

- *Simple Audio Menu* – Lists the points of interest in a room through their corresponding sounds effects and text-to-speech, in alphabetical order

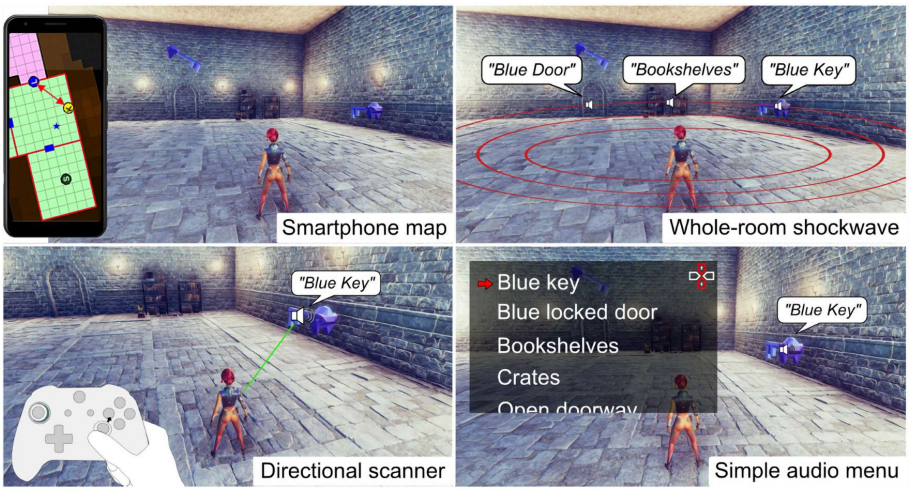

Figure 3.6: The four spatial awareness tools implemented within Dungeon Escape [32]). © 2022 ACM.

To evaluate the effectiveness of each approach in conveying essential aspects of spatial awareness and address the two main questions of the research, Nair et al. implemented them all into Dungeon Escape. Dungeon Escape is an original third-person 3D game in which the players were required to use the given SATs to gain enough spatial awareness to succeed. Implemented in the Unity game engine, it is an adventure game specifically designed for the study. Players navigate dungeons in search of objects that allow them to overcome obstacles and thus escape.

Despite its focus on studying the performance of the SATs within the differing dungeon layouts, it is more than a simple playground for the different SATs, portraying itself as an accessible game in other ways. As in well-known 3D games, the left thumbstick is used for movement and rotation, and in this case, aided by a utility mimicking snap rotation. Collisions have a unique sound effect, and relevant sound is played from any object within a 2-meter radius of the player. Furthermore, players can lock onto objects of interest by placing a looping audio beacon.

Following a user study with nine participants, eight of whom were completely blind, the most critical aspect of spatial awareness was found to be position and orientation. Presence, arrangement, and adjacent areas tied for second place, while shape and scale overwhelmingly came last. One of the key findings was that despite the importance of position and orientation, none of the approaches were entirely satisfactory, though the directional scanner performed the best. Another relevant takeaway was the effectiveness of combining some of the SATs with the most significant spatial awareness provided by the directional scanner and simple audio menu combination. Finally, visually impaired individuals greatly value customizable SATs, as the combination of the directional scanner and whole-room shockwave closely followed the aforementioned SAT combination. This

is despite the whole-room shockwave having been considered overwhelming in several instances (yet customizable) and the simple audio menu a great way to communicate presence (yet disliked by half the participants for being "spoilers").

Nair et al.'s research provides several insights into the current dissertation, primarily by exposing the strengths and weaknesses of the different SATs and the accessible design of Dungeon Escape. Most importantly, it highlights the importance of position and orientation for proper spatial understanding and exploration, which we will prioritize conveying in our work, possibly by incorporating combinations of the most synergetic SATs as was done in their study. Additionally, it touches on how users may be overwhelmed by too much detail, inspiring us to focus on providing cues qualitatively rather than quantitatively.

3.4 Tools for Soundscape Creation

The current market is home to advanced soundscape creation software, enabling creators to design highly detailed and expressive auditory environments tailored to the most diverse applications. Different tools have different purposes, with some focusing on professional sound design, where the spatial accuracy of sound environments is the priority, whilst others aim for dynamism and adaptability.

Regarding sound design and immersive spatial audio, Dolby Atmos[4] stands out as the industry's standard in delivering 3D audio experiences, be it in entertainment, gaming, or Virtual Reality (VR). In Dolby Atmos, environment creation is object-based, meaning that sounds may be placed and moved as discrete objects in a 3D space. Sound Particles[5] is another tool meriting a mention, as it also delivers immersive, high-quality audio for films, gaming, and virtual environments, excelling on large-scale 3D audio effects by simulating spatialized sound with particle systems.

Though not exclusively centered around sound design and production, game engines such as Unity and Unreal Engine[6] have proven effective platforms for soundscape creation, offering robust audio capabilities and integrations. Unity, for instance, provides built-in spatial sound support through its default audio system, with more advanced spatial audio capabilities being made available through integration with third-party tools like FMOD[7] and Google's Resonance Audio. Unreal Engine features a built-in audio engine supporting 3D spatialization, reverberation effects, and audio triggers, more advanced than Unity's default audio system but not accounting for its third-party augmentations.

As expressive and competent as the tools mentioned so far are, they have a significant learning curve, a tradeoff for such feature richness and customization ability. In the context of this dissertation, where soundscape creation should not require prior experience in

[4] https://www.dolby.com/technologies/dolby-atmos/
[5] https://www.soundparticles.com/
[6] https://www.unrealengine.com/en-US
[7] https://www.fmod.com/

sound design, an adequate balance between expressiveness and complexity is a must. Additionally, there are specific design considerations we should adhere to or at least consider regarding the visually impaired, addressed in specific research.

Guerreiro et al. [21] proposed a theoretical framework to support BVI-inclusive auditory representations of object location, behavior, and interaction in virtual environments, utilizing spatialized sound and sonification in Unity. Their proposed design space explored nine distinct categories, such as the audio field, cardinality, concurrency, and spatialization. A user study they conducted showed that sound spatialization was not always preferred, mainly when dealing with moving objects, and that fully concurrent auditory feedback could be mentally overwhelming [21]. Meanwhile, Krol et al. [24] investigated the use of automated musical soundscapes in visual art accessibility for the visually impaired. From their study, several design considerations were highlighted: enhancing narrative comprehension by integrating storytelling elements; soundscape customization options; integration with other accessibility methods such as audio descriptions; contextual and historical accuracy/adequacy; inclusion of ambient sound effects; effective audio reproduction such as spatial sound and reverb to convey location, movement, and atmosphere.

3.4.1 Immerscape

The PASEV[8] project focuses on retaining and promoting the city of Évora's cultural patrimony, namely the rich soundscapes correspondent to the various historical events which took place between 1540 and 1910. As part of this initiative, Ferreira [19] proposed Immerscape.

Intended to provide the PASEV projects's team with the means to reconstruct the city's auditory history from current sound recordings, Immerscape is a soundscape editing tool accessible even to those without prior experience in programming or sound composition software. It allows for the creation of historical soundscapes out of previously collected recordings and the generation of immersive 3D audio files representing such soundscapes, while abstracting away the technical details behind spatial audio generation. The editor itself was implemented within the Unity game engine. The acoustic immersion leveraged the Google Resonance Audio SDK to spatialize sound by applying HRTF filters, requiring headphones to be adequately evaluated.

Besides an accessible interface with minimal complexity (Figure 3.7), Ferreira defined some fundamental requirements for Immerscape, such as the ability to select predefined 3D sound environments (with unique resonance and reverb properties), immersive environmental navigation, creation of editable sound sources (audio properties, movement and triggering events) and real-time playback/recording of the soundscape. Additionally, there are two available camera angles during development, the default being the player view and the alternative an up-view.

[8] https://www.uevora.pt/investigar/projetos?id=3776

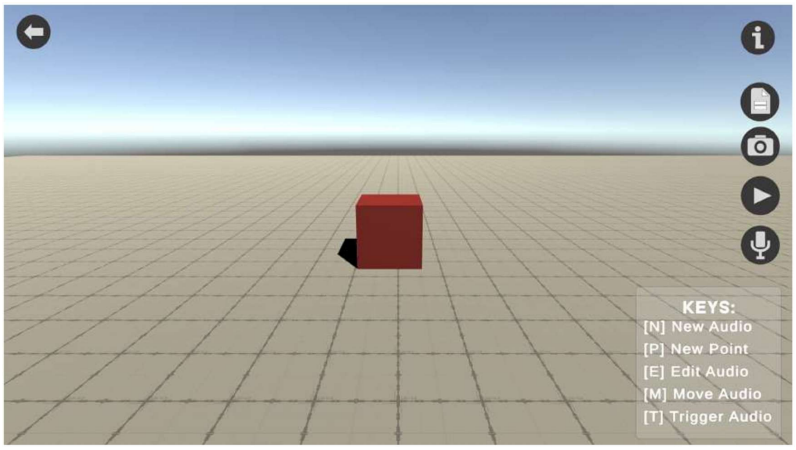

Figure 3.7: Immerscape's environment, through the player view perspective. The red cube represents an audio object (Reproduced from [19]). © 2021 Carolina Ribeiro Dias Ferreira.

An evaluation with 10 developers and 7 non-developer participants highlighted Immerscape's high usability and immersive quality, proven to be an effective way of providing engaging sonic experiences, in this case, by conveying historical soundscapes.

While Ferreira's work targets general audiences rather than the BVI demographic, there is much to take away from it in the context of this dissertation. Immerscape's design and implementation of the HRTF filters inspire what will transpire throughout our work, particularly regarding spatial audio simulation and ease of use for non-experienced users. Furthermore, some of the requirements defined by Ferreira, including predefined 3D sound environments, editable audio sources, and real-time playback of the scene, will likely be imported into our research, aligning with our goal of creating an intuitive yet expressive soundscape editor.

4

Proposed Solution

This chapter provides a detailed overview of the proposed system and briefly exposes the solution's validation methodology. It also addresses the expected technological stack and the envisioned plan for the system's development. The work schedule is split into five distinct and concisely explained tasks mapped in a Gantt chart.

4.1 Proposal Overview

The solution we propose consists of a user-friendly 3D soundscape editor for Windows, specifically designed to support the creation and functionality of immersive virtual auditory environments. These environments will be made available and interactive through a BVI-accessible mobile application, given the widespread use of smartphones. With this approach, we intend to enable curators to generate immersive 3D audio environments representative of specific paintings, and to have BVI users explore these environments interactively. Figure 4.1 attempts to illustrate the intended approach from a purely conceptual standpoint.

Thus, the system we aim to develop has two very different symbiotic parts. The first is the de facto editing of 3D sound environments in a desktop application, catering to curators without previous experience in sound engineering and related software. It abstracts the complexities of spatializing sound sources and optimizing accessibility, automatically generating BVI-accessible environments upon exportation. Furthermore, it features two toggleable modes, one for development and the second for testing. Just as important, the other part consists of a mobile application where the previously generated scenes are made available for interaction. It concerns the interactivity, functionality, and accessibility of said environments for BVI individuals, on a mobile device.

Following the implementation, the proposed system will undergo a structured validation process on the two fronts it caters to through user testing, employing a subjective quality inference mechanism. Through Likert scale surveys and qualitative feedback, the editor tool will be evaluated by individuals with an artistic background and no experience in sound design, to whom testing will focus on ease of use and overall expressiveness.

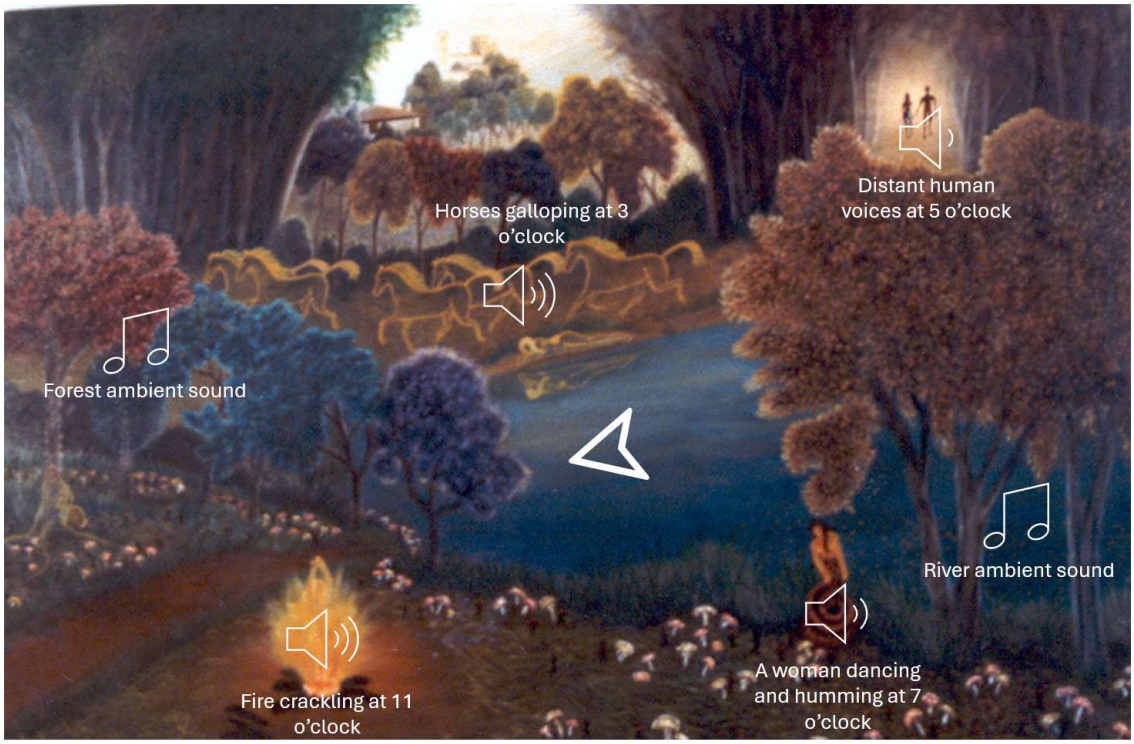

Figure 4.1: Concept representation. Image is *O Bosque Sagrado* by Maria Benamor, 1975. Collection of Museu Nacional Grão Vasco. (Adapted from [36]).

Analogously, BVI individuals will evaluate mobile interaction based on factors such as ease of navigation, spatial awareness enhancement, level of enjoyment, user preferences, and cognitive load assessment.

4.1.1 Mobile Soundscape Player

Loading up a scene exported from the developed soundscape editor and then parsing it, the mobile application will display a top-down 2D representation of the soundscape with large high-contrasting elements representing directional sound sources and collidable zones, making them easier to distinguish within the environment. This aligns with standard BVI accessibility practices for visual interfaces in other related studies [1, 15, 46].

Navigation through the virtual space happens through a virtual joystick control schema, inspired by Nair et al.'s [32] Dungeon Escape game and other mainstream 3D games such as the earliest Resident Evil[1] titles. Spatial position, orientation, and proximity to different parts of the artwork are felt through directional audio (implemented through HRTF filters) and dynamic intensity regulation.

The left joystick will control movement by tilting forward and backward for directional movement and tilting left or right for fixed 15-degree snap rotations complemented by audio cues. The right joystick will work as a directional scanner, akin to Nair et al.'s [31]

[1] https://game.capcom.com/residentevil/en/

NavStick, where the user can scan the environment in a specific direction and receive a spatially emanated verbal description of the first object within their line of sight, through a predefined speech output.

Further, to assist with a more literal interpretation of the environment, double tapping will trigger a complete verbal description of the scene (predefined by the environment's creator) and pause all other sounds apart from ambient sounds. Obstacles encountering the user will trigger real-time audio feedback in the form of specific collision detection sound cues, once again inspired by Nair et al.'s [32] Dungeon Escape.

Especially for BVI people, cognitive overload is a significant concern [21, 32]. As such, users may control the density of sensory inputs to a degree by modifying the number of 3D sound sources being played simultaneously using pinch gestures on the touchscreen. Additionally, a long press gesture toggles between sequential and concurrent playing of sounds. In the former mode, sounds are played in a clockwise direction (starting at the closest sound).

4.1.2 Soundscape Editor for Windows

Inspired by Ferreira's [19] work on Immerscape (addressed in section 3.4.1), the editing tool to be developed will feature predefined environments of varying acoustic properties, such as resonance and reverberation, which may be further customized.

It is convenient for users to easily manipulate and experiment with different sound elements, such as Minecraft[2]'s block-based system, where users can place or erase different types of blocks with distinct functions. Thus, the scene is built with a similar block-based system and, like Minecraft's creative mode, allows for movement in all directions, including free flight. There will be multiple block types, such as 3D sound emitter, ambient audio, collidable block (only simulates physical barriers, does not emit sound), and user-customized ones. These blocks are editable in several ways, some of which are volume, loop settings, sound file association, and object representation.

When developing a scene, insights into the end user's experience are crucial to iteratively improving it. Hence, the editor will have two main modes: Development Mode, where scenes are built and tailored, and Interactive Mode, where users can test the experience from the end user's point of view, though with desktop controls.

Additionally, the user may define default values and boundaries for movement speed, audio cardinality, and zoom level in the exported scene and verbally describe the whole piece. Once a scene is created, it can be saved, loaded, and edited. Furthermore, a built-in help function will always be available, offering a concise explanation of the system's functionalities. Users will supply their sound assets, storing them in a dedicated directory in supported formats (these formats are yet to be decided but are likely to be WAV and/or MP3). On exportation, the editor will package the scene's data (possibly in JSON or XML) and store it in a designated folder.

[2] https://www.minecraft.net/en-us

4.2 Technological Stack

We selected the Unity game engine as the primary development environment for the desktop soundscape editor and mobile application prototype. One factor that solidified this choice was its adoption among some of the most relevant related work we have analyzed [19, 21, 31, 32, 46, 52] in chapter 3.

A real-time 3D development platform, while Unity is mainly known for its role in game development, its versatility has led to widespread use in all sorts of interactive 2D and 3D experiences across various industries [48, 54]. It provides extensive cross-platform build support, including the hardware we are particularly interested in: Windows PC (for the editor application) and Android (for the mobile interaction).

Aside from its powerful rendering capabilities, Unity sports strong community support, a vast amount of learning resources, and a convenient package manager extending its functionality with third-party libraries. Its learning curve is also friendlier than other mainstream 3D game engines like Unreal Engine.

With a built-in 3D audio system, Unity provides a solid foundation for basic sound spatialization with distance attenuation according to positioning, a spatial blend parameter, and adjustable volume roll of, among some other settings. While it may be adequate for rudimentary 3D audio, this built-in system is limited, lacking binaural rendering and other relevant capabilities. Fortunately, specialized plugins for advanced spatial audio address the vanilla system's limitations by adding some of the essential and advanced spatialization functionalities it lacks.

Initially, we had considered the Google Resonance Audio SDK for integration within Unity, as it offered relevant features such as HRTF processing and environmental modeling. Furthermore, some of the analyzed studies [19, 46, 52] had successfully incorporated this library in their Unity-based environment. However, it has been deprecated for quite some time and has no support for recent versions of the Unity engine.

For these reasons, we turned to the Steam Audio framework instead, which has been actively maintained, well-documented, and compatible with Unity engine updates. Moreover, we were motivated by the use of this toolkit in NavStick [31] and Dungeon Escape [32], two of the most influential works for our proposal. Supposedly easy to implement and deploy, Steam Audio encompasses high-fidelity HRTF-based binaural rendering and geometry-based occlusion, reflection, and reverberation effects, among other spatial features contributing to natural sounding immersion. While the specific library to which its default HRTF dataset belongs is not explicitly stated in official documentation, some online discussions suggest it may be based on Phonon[3].

[3] https://www.impulsonic.com/what-is-trueaudio-phonon/

4.3 Implementation

4.3.1 Mobile BVI-Acessible Soundscape Player Prototype

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim

interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

4.3.2 Desktop Beginner-Friendly Soundscape Editor

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus

eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

4.4 Initial Research & Concept Refinement

Prior to the system's technical implementation, conceptual designs of the mobile soundscape player and the desktop editor were produced in Figma, illustrating the key functionalities envisioned for each application. At this stage, the mobile application was identified as the main component to prove the concept, since it would directly serve BVI users.

The first implementation efforts therefore focused on an early version of the mobile prototype, already including most of the features described in section 4.1.1. This build was sufficient to demonstrate the feasibility of the proposed concept for accessible and interactive interpretation of a painting, as was ultimately intended. However, compared to the final application it lacked ambient sound regulation, the proximity sensor, and haptic feedback, among several other adjustments - as these hadn't yet been planned at the time. Along with the Figma mockups of the desktop editor, some of the mobile functionalities were presented in a remote meeting with researchers from the Department of Conservation and Restoration of NOVA School of Science and Technology. The researchers saw potential in the idea and expressed interest in its evolution.

The most important stage of this early phase came in late May, when the mobile prototype was tested on-site at the Raquel e Martin Sain Foundation - a social solidarity institution dedicated to the professional education and social integration of BVI individuals. Two blind participants volunteered their time: Cláudia Pires, a student at the foundation, and Ana Inês Colares, one of its instructors. Both of these interviews followed a simplified protocol with similarities to that of the final study but with less specific tasks and more open-ended questions, for the goal was refinement rather than evaluation. The painting "O Bosque Sagrado" by Maria Benamor was used as the painting for the test scenario, also being used in the final study of the application (section 5.2).

The first interviewee, Cláudia Pires, pointed to several areas for improvement in the experience. She felt that ambient sound should be lower and that the sound effects used for movement and rotation were too similar, sometimes hard to distinguish. Though she could discern sound direction clearly, perceiving proximity to the sound sources was more difficult. While she found the directional scanner useful, she noted that it required alignment that was much too precise, with narrations cutting off if not perfectly targeted. Cláudia also reported little difference between sequential and simultaneous reproduction modes and sometimes had difficulty tracking how many sounds were active. The clock hour orientation feature was very helpful to her but she stated that it should have been introduced earlier in the protocol. Finally, she strongly preferred the console scheme over

the touch gestures, suggesting haptic feedback and larger virtual joysticks as possible improvements to touch usability.

Many of the points brought up by Cláudia were echoed by the second interviewee, Ana Inês Colares, who also provided further insights. Accentuating the importance of the clock hour orientation system, she too felt that this feature should be introduced in an earlier task. Ana Inês suggested that sound sources should increase in volume more noticeably when approached and that ambient sound should be toggleable, as it could distract from the other elements at times. Like Cláudia, she noticed little difference between the two reproduction modes and found sound source direction easier to perceive than proximity, occasionally struggling to locate specific sounds in the environment. She also preferred the console controls and noted the difficulty of touch gestures and especially virtual joysticks without haptic feedback. Lastly, she proposed the addition of a menu to allow switching between different soundscapes, if the application were to support several painting representations. Most importantly, from the conversation with Ana Inês and her emphasis on the amplification of sound according to proximity, derived the idea for the currently implemented proximity sensor.

The purpose of this early study was to garner real input from the individuals to whom the application was intended, and in this way assess whether the proposed features were indeed adequate for BVI needs, rather than only in theory. These preliminary interviews were fundamental in shaping the application into what it eventually became, as both participants provided detailed feedback which revealed not only the prototype's shortcomings but also some of its strengths. As much as possible, their suggestions were incorporated into the application's final design and in the later evaluation (section 5.2.3) these proved both practical and meaningful.

5

Evaluation and Results

This chapter presents the evaluation of the two developed prototypes as their own distinct study and starts by outlining the methodology common to both. It then details the assessment of the mobile soundscape player with BVI participants, followed by the desktop soundscape editor with users inexperienced in sound design. Each study is generally described in terms of participant characterization, the tasks undertaken, and the main findings drawn from survey responses. A summary of the results of both studies concludes the chapter, discussing the extent to which the prototypes achieved their intended objectives.

5.1 User Studies

To evaluate the proposed system in its entirety, two final user studies were conducted, each corresponding to one of the two developed prototypes: the mobile soundscape player designed for BVI users, and the Windows-based soundscape editor intended for museum curators without prior experience in sound design. The objective of these studies was to assess not only the technical functionality of each application but also their usability, accessibility, and the overall adequacy of the implemented features for their respective target audiences.

Despite addressing distinct user groups, both studies shared a common methodological structure. Each followed a predefined protocol specifying the study's general objectives, the equipment to be used (all provided by the researcher) and the usage scenario. In both cases, the researcher's assistance was deliberately minimal, limited to clarifying questions about controls or task instructions.

Participants first completed a sequence of predefined tasks designed to cover the core functionalities of the respective application. Some tasks were highly specific, while others allowed for more freedom in execution. The focus of the tasks wasn't evaluation in and of itself, but rather a pratical introduction to the system's features and controls. This structured segment was followed by free-use optional sessions, during which participants

could continue to interact with the system without constraints. These sessions aimed to familiarize the user further with the system, independently and at their own pace.

At the end of each session, participants filled out a survey consisting mainly of Likert-scale items supplemented by open-ended qualitative questions. The closed questions gauged ease of use and functionality intuitiveness and adequacy. The open-ended questions allowed participants to share their general impressions, report difficulties, and suggest improvements. Comments and notable observations made during the testing sessions were also considered as part of the feedback.

The focus of the two protocols diverged according to the nature of the application. For the mobile soundscape player, emphasis was placed on accessibility, immersion, and the ability to meaningfully interpret a painting through sound. For the desktop soundscape editor, the main focus was on comprehensibility and intuitiveness, specifically whether non-expert users could construct an expressive auditory representation of a painting without prior technical training.

To guarantee anonymity, each participant was assigned a numerical identifier, which is how they will be referenced in the subsequent analysis. Participation in both studies was voluntary and uncompensated.

5.2 Mobile Sonic Painting Exploration

The first study focused on the evaluation of the mobile soundscape player, which enables BVI users to autonomously explore a painting through auditory cues and several spatial awareness features. As outlined in section 5.1, this study assessed the application for accessibility, immersion, and interpretability.

All user testing sessions were conducted on-site at the Raquel e Martin Sain Foundation, as in the preliminary interview discussed in section 4.4. Ana Inês Colares, an instructor at the foundation and also a former interviewee, provided the participants for this study and also the same fairly quiet room for all sessions. All tests took place in comparable conditions, with the same equipment and research protocol.

The setup comprised a smartphone running the application, headphones, and a PlayStation 4 controller. This configuration supported the two distinct modes of interaction: a physical control scheme through the gamepad and touch-based controls on the device itself. The headphones were required for the perception of spatial sound. After a quick overview of the application, participants were first introduced to the system through a series of structured tasks using the PS4 controller. As they were not yet acclimated to the application or its controls, the tactile nature of the physical joystick and buttons was intended to give them an initial sense of stability, as they are yet learning the functionalities.

The utilization scenario consisted of navigating a sonic representation of "O Bosque Sagrado", a painting by Maria Benamor, guiding participants to locate specific sound sources, regulate sounds and reproduction settings, and make use of several orientation aids. Following this exposure to the prototype's main features, the participants were

encouraged to explore the soundscape more freely, initially continuing with the controller, and then transitioning to touch controls, so they could compare both input types.

The presented soundscape's content was developed by the research team (in this case, the author), rather than by art professionals, using sound clips available online for the scene elements and Google Cloud's text-to-speech for narrations. Participants were made aware of this ahead of the tasks, and were also informed that the soundscape itself was not being evaluated — only the application was.

5.2.1 User Characterization

A total of eight participants took part in the main evaluation of the mobile soundscape player. None of them had been previously exposed to the application or its functionalities, ensuring that their feedback reflected first-time user impressions instead of familiarity. Table 5.1 characterizes the participants.

Participants' ages were relatively balanced across middle adulthood, ranging from 28 to 56 years old, for an average of 42.5 years.

Table 5.1: User characterization by identifier.

ID	Age	Gender	Degree of Vision
1	54	Male	Low Vision
2	41	Male	Blind
3	56	Female	Blind
4	37	Female	Low Vision
5	37	Male	Blind
6	36	Female	Blind
7	51	Male	Blind
8	28	Male	Blind

The group consisted of five male participants (62.5%) and three female participants (37.5%). In terms of visual condition, six participants (75%) were blind and two participants (25%) were classified as having low vision. Among the blind participants, some reported having lost their sight later in life rather than being blind from birth, which in some cases may have influenced the way they described their perception and interaction with auditory cues.

When asked about their prior familiarity with joystick interfaces and gamepad controllers, most participants indicated that they were not accustomed to such devices. However, several noted that they had used them occasionally, most in the context of other user studies and a minority for gaming. This background might have influenced their degree of comfort with the PlayStation 4 controller used in the evaluation, as all participants were able to complete the required tasks with minimal guidance.

5.2.2 Task Overview

Participants were guided through a set of six tasks designed to gradually introduce the main features of the mobile soundscape player. As previously discussed in section 5.1, these tasks were not conceived as a strict evaluation in themselves, but rather as a structured way of teaching participants how to use the system. These ensured that each participant could get comfortable with the core functionalities before moving on to the optional free exploration sessions.

The set of tasks was intentionally varied: while some were highly specific (such as finding a particular sound source), others left some room for interpretation and exploration, allowing participants to interact with the environment at their own pace. During subsequent tasks, participants were encouraged to use previously learned functionalities, but were free to ignore them as well.

Standing out from all six other tasks, Task 0 was exclusively intended for low-vision subjects, as it required a certain degree of residual vision to adjust the zoom level and toggle between different visual display modes. It was not presented to blind participants and actually carried little relevance even for those with low vision. Due to its optional character, it did not interfere with the main evaluation flow.

Task 1 focused on basic movement and spatial perception, introducing participants to the joystick-based control of direction and rotation, while Task 2 expanded on this by training orientation with respect to the original alignment of the painting, using the clock-based narration system. Task 3 introduced the directional scanner and the global narration feature, allowing users to hear localized descriptions of individual elements or a full scene overview. Task 4 presented the proximity sensor, which mimicked a car parking sensor by giving intermittent auditory feedback as the user approached a sound source. Task 5 allowed participants to regulate the number of active sound sources and toggle ambient sound, introducing the idea of controlling auditory complexity. Finally, Task 6 exposed participants to the two available playback modes — simultaneous and sequential — focusing on how these modes altered the listening experience.

These tasks served as a walkthrough of the application's feature set, that prepared them for the optional free exploration sessions that followed.

5.2.3 Results Overview

The evaluation of the mobile soundscape player yielded encouraging results overall, with participants rating most of the features positively. It is important to note that the Likert items in the survey were phrased as declarative sentences, and scores reflect participants' level of agreement with them (1 = strongly disagree, 5 = strongly agree). A higher score indicates only strong agreement with the statement, which may or may not correspond to a direct quality judgment depending on the item in question. The structured tasks served their purpose as a guided introduction to the system, while the subsequent exploration phases revealed both strengths and points for refinement.

Task 0, which involved zoom and visual mode adjustments, was only presented to low-vision participants and, as anticipated, had little overall impact on the evaluation.

In terms of navigation, participants generally felt quite able using the left joystick to move through the scene and orientating themselves via the clock system, as can be seen in figure 5.1. Only participant 8 found himself neutral to the orientation feature's usefulness, with 38% of participants scoring it highly and 50% very highly. As for the movement and rotation via the left joystick, half of the participants were comfortable with these controls while the other half stated to be very comfortable. These results could in part be due to their experience with gamepads in other user studies, but even participants who had less previous exposure to such interfaces learned the movement controls very quickly.

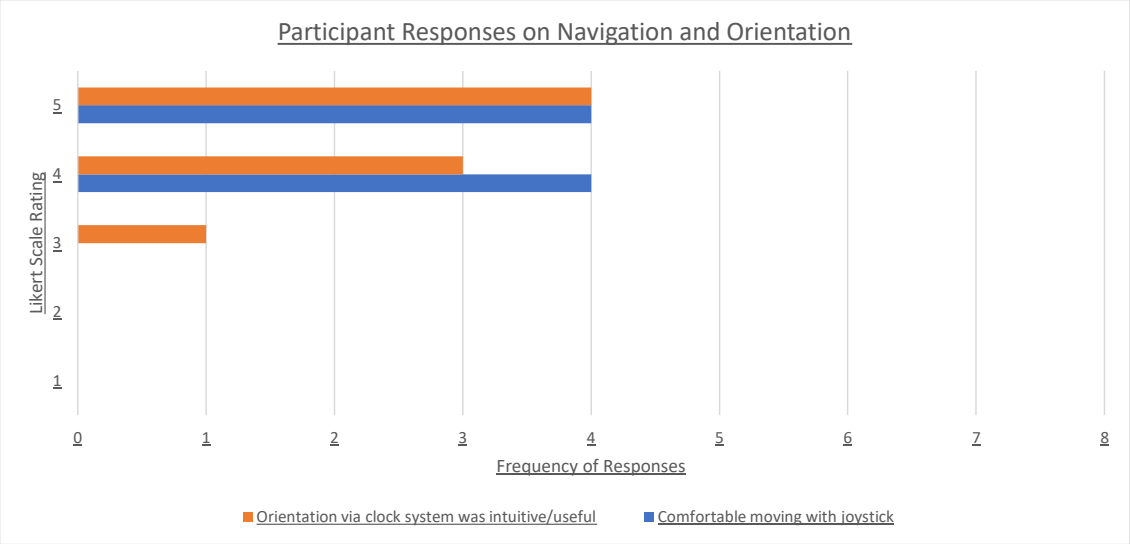

Figure 5.1: Histogram of participant responses regarding navigation comfort and orientation.

In what pertains to spatial awareness, the vast majority of users reported that they could clearly identify both the direction and proximity of sound sources. Figure 5.2 displays the perceived efficacy of the implemented spatial awareness tools, as well as the overall perception of sound source direction and proximity. On average, clearness in direction and proximity were rated the same value - 4.1 out of 5. Aside from positive, this result is particularly interesting as in the early prototype's refinement study (section Y), sound source direction had been considered significantly clearer than proximity. In the current application, that gap has apparently been bridged, likely due to the inclusion of the proximity sensor and adjustments to the sound elements in the scene. The said proximity sensor, which emitted increasingly frequent signals as users approached a sound source, was frequently classified as useful and intuitive. Similarly, the directional scanner allowed participants to identify specific objects in the painting with relative ease, by leaning their finger in the direction they want to listen in. Both functionalities were well received and close in ratings.

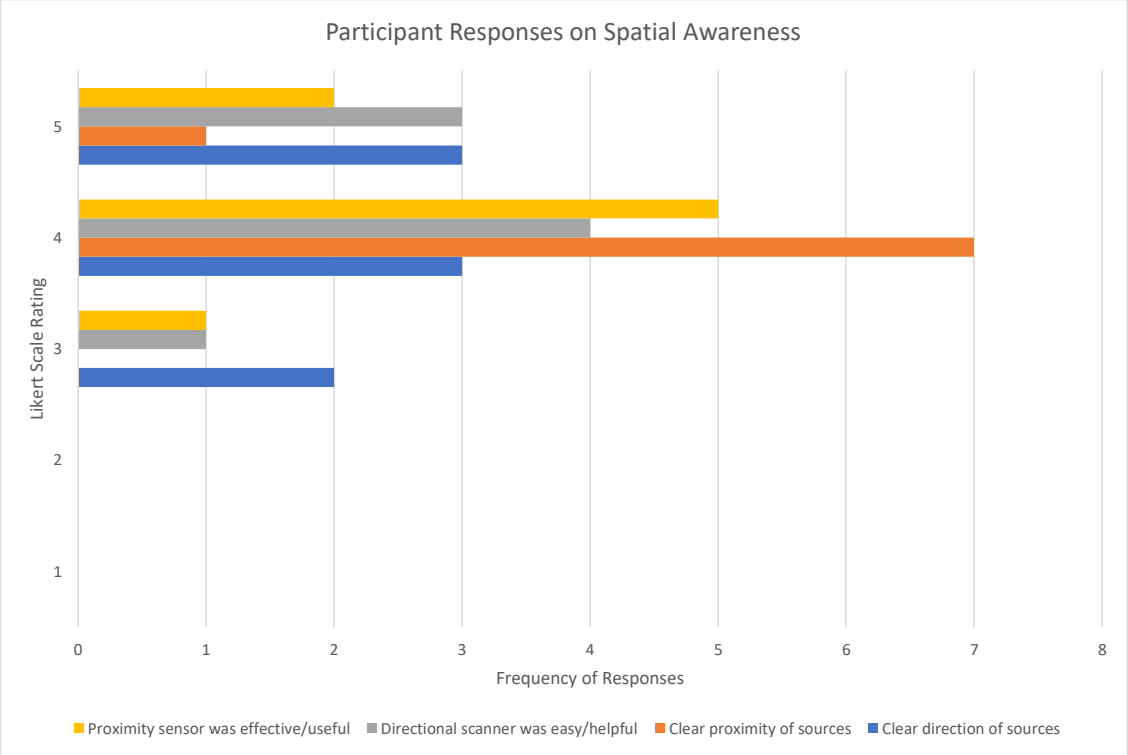

Figure 5.2: Histogram of participant responses regarding spatial awareness and related features.

Opinions were more divided regarding sound environment regulation, namely in whether ambient sound distracted from the localized elements of the painting. As can be seen in figure 5.3, half of the participants thought that it did, while the other half thought the opposite. Independent of their stance on ambient sound, all participants considered the ability to toggle it on or off indispensable, unanimously rating its usefulness a 5 out of 5. In addition, participants appreciated being able to adjust the number of active sound sources, indicating that the ability to manage the auditory scene's complexity was highly valued.

Regarding the two available playback modes - simultaneous and sequential, participants felt both modes were distinct and impactful in their own way, but favored simultaneous reproduction by a fair margin, as illustrated in figure 5.4. While sequential playback allowed some participants to focus more closely on the painting's individual elements, most felt that the simultaneous mode provided a greater sense of immersion. Participant 7, in particular, saw no purpose in the sequential reproduction mode, stating that a similar experience could be accomplished in the simultaneous mode by decreasing the number of active spatial sounds to one, and also felt that each sound source would play for too long sequentially. In contrast, participant 3 had no preference in mode, having enjoyed both equally. As preferences appear to depend on personal strategies for exploring the painting, offering both modes is valuable, even if one is considered the "better" one.

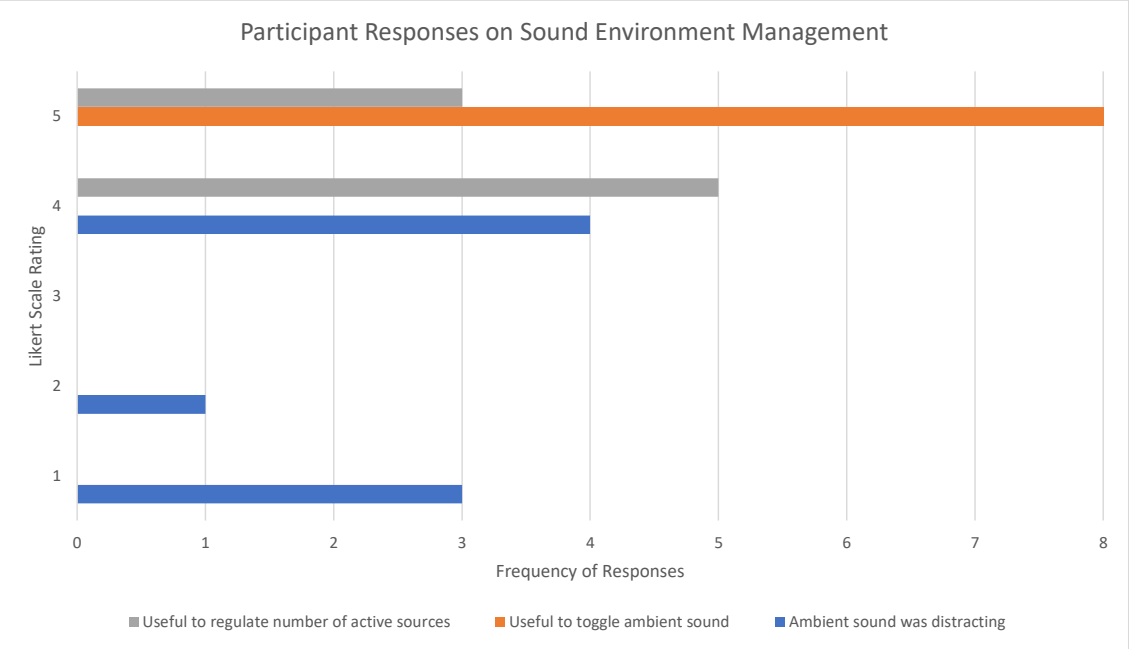

Figure 5.3: Histogram of participant responses regarding sound environment regulation.

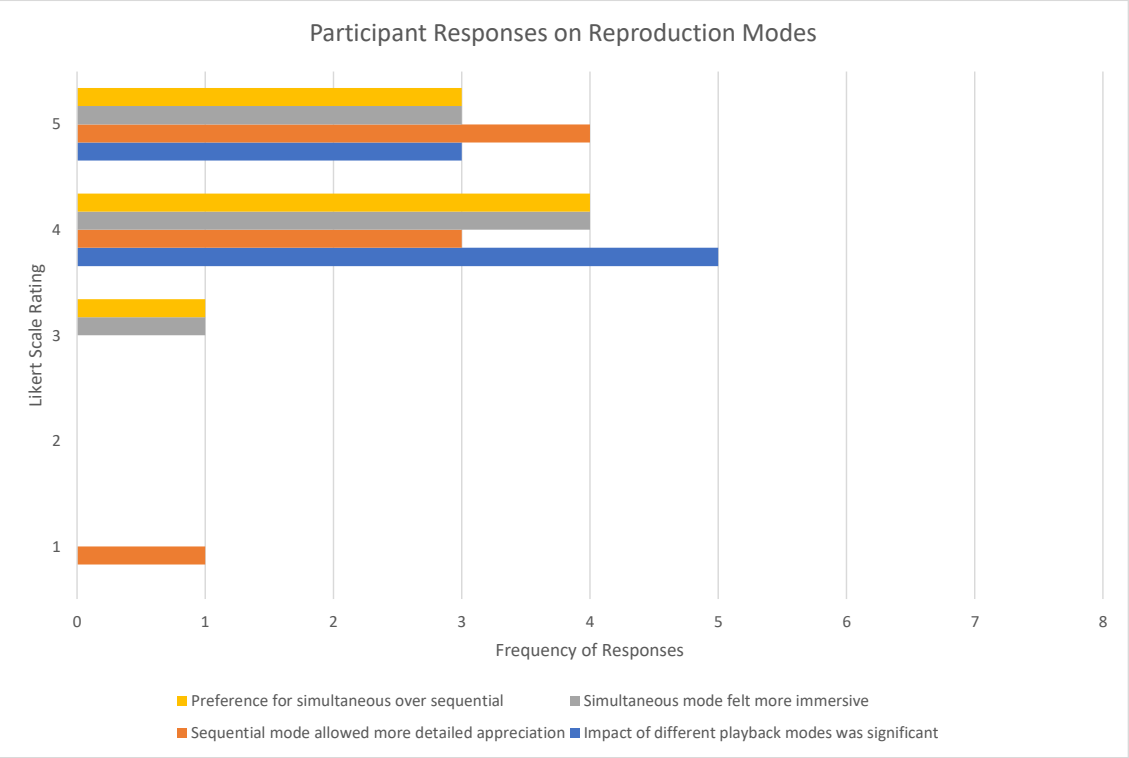

Figure 5.4: Histogram of participant responses regarding the two existing reproduction modes - simultaneous and sequential.

As for the application's control schema (results illustrated in figure 5.5), console controls were rated very positively, with an average agreement of 4.6 out of 5 on their intuitiveness and ease of use. While the touch controls were also described as intuitive by the majority, the responses were less consistent, with an average score of 4.1 out of 5. Participant 2 rated the touch controls' intuitiveness a 2 out of 5, due to some difficulty with the swipe gestures and especially the position and size of the virtual joysticks. While the rest of the participants did not experience difficulties to this extent, they generally reported greater comfort and confidence with the physical controller, thus preferring this interface when asked to compare both input modes (3.75/5). However, the console scheme was not one-sidedly favored in this case, as 25% of subjects actually preferred the touch controls and participant 1 appreciated both control schemes equally. The moderate success of the touch controls can in part be explained with the positive reception of the haptic feedback present in this schema, which may have compensated for the lack of tangible button-based physical interaction. The smartphone vibrations, triggered by collisions with the painting's borders or virtual joystick interactions, received an average score of 4 out of 5 in effectiveness. Though some users did point out that they would prefer these vibrations to be more intense.

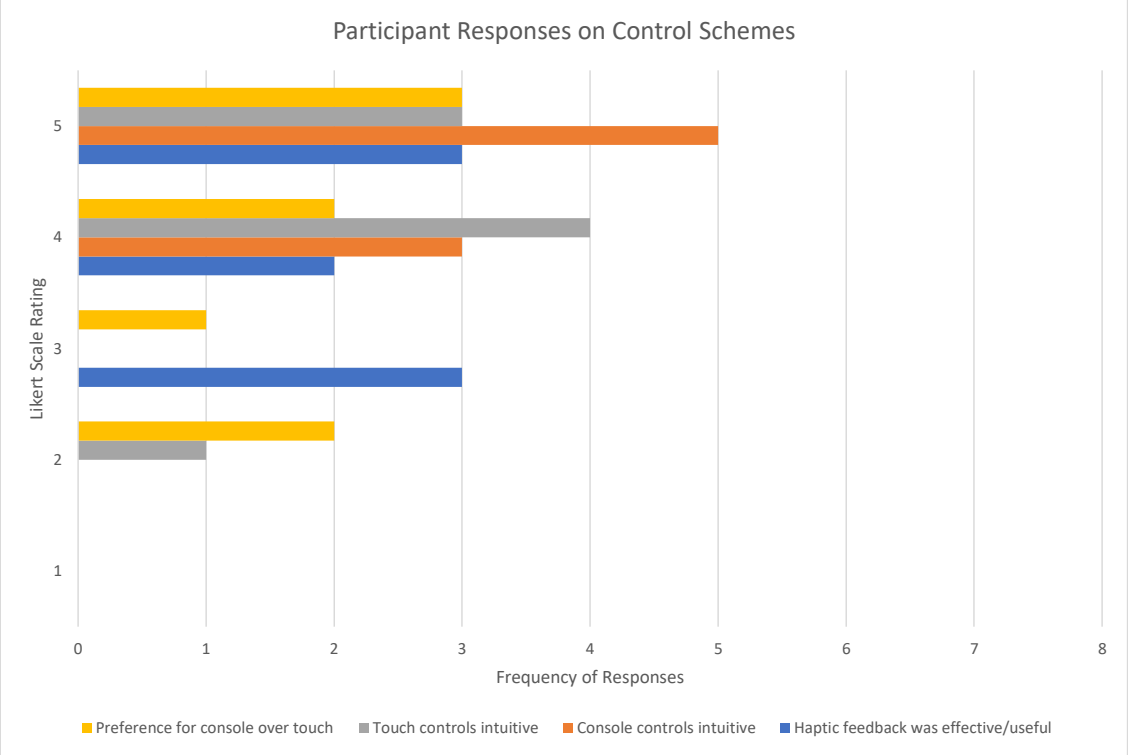

Figure 5.5: Histogram of participant responses regarding the available control schemes - physical gamepad and touch.

The general reception of the application was highly positive, with participants strongly agreeing that the system provided an immersive and engaging with the painting (average

score 4.4/5). Even more notably, they agreed that the set of available features allowed them to form a somewhat complete idea of the painting's content and essence (4.6/5).

Responses to open-ended questions reinforced the quantitative findings of the study, as several participants highlighted the immersion and stimulation of the experience, some also suggesting improvements. Out of the suggestions, that of participant 5 was a particularly interesting one - incorporating clock hour narration alongside the directional scanner's narrations, for further awareness assistance. As expected, most users commented on initial difficulties in orienting themselves within the scene but reported that these lessened with practice and with new spatial awareness tools being introduced. When asked about usage context, most felt the application would be most meaningful in a museum setting alongside the artwork, others couldn't choose between using it at a museum or remotely, and one participant preferred a fully remote interaction.

The results of this study suggest that the application succeeded in delivering both an enjoyable and meaningful interpretation of the artwork, beyond the usability of individual functionalities - its potential acknowledged as both an independent exploration tool and a complement to in-person museum visits.

5.2.4 The Perspective of a Returning User

One of the two blind individuals who had previously taken part in the early prototype's concept refinement sessions (addressed in section 4.4), Cláudia Pires, returned to test the application in it's final state. With her participation came a unique perspective, as she was the only person in this evaluation with any prior exposure to the application, though this occurred a month and a half earlier and under a different protocol. For this reason, her results are considered separately from those of the other participants in section 5.2.3, who all experienced the system for the first time.

The early prototype that Cláudia had tested already included most of the core functionalities, but lacked features such as the proximity sensor, toggleable ambient sound, and haptic feedback. Additionally, between versions the sensitivity of the directional scanner was adjusted, and the sound attributes of the test soundscape were slightly refined. The protocol of the earlier study was similar in structure to the final one, using the same painting and a comparable sequence of tasks, though these were less specific and accompanied only by qualitative questions. Consequently, some of the tasks in the final evaluation were familiar to her, while others introduced entirely new elements.

Cláudia performed these final tasks with relative ease, quickly adapting to both the revised button bindings and touch gestures. She highlighted the improvements in the directional scanner's usability and especially the mobile touch controls, which she found more reactive and supported by proper haptic feedback. Cláudia's responses to the survey aligned with the positive trends observed in section 5.2.3, and emphasized the importance of iterative improvements, especially when coming from someone able to compare versions of the system.

It is also important to acknowledge Ana Inês Colares, the other blind participant who had taken part in the early prototype study and also volunteered for the final study. Ultimately, she was unable to attend the final evaluation due to the researcher's time constraints. Nonetheless, her role was indispensable in supporting this stage of the research by greatly facilitating the recruitment of the BVI participants, essentially managing this process in its entirety.

5.3 Desktop Soundscape Creation

The second user study focused on the Windows-based soundscape editor, intended for museum curators without prior experience in sound design to build auditory representations of paintings in an intuitive manner. As stated in section 5.1, the objective of this evaluation was to assess the tool's usability and comprehensibility, but also to determine whether participants could successfully construct an immersive soundscape capable of conveying the essence of an artwork to BVI visitors.

The evaluation followed a similar structured methodology to the one applied in the mobile prototype study (section 5.2), with a sequence of tasks introducing participants to the editor's main features, followed by a short free-use period and a survey. The utilization scenario paralleled that of the mobile application's study. Participants were tasked with creating a auditory environment of "O Bosque Sagrado" by Maria Benamor, essentially reconstructing the very same soundscape that BVI participants had explored within the mobile application.

Most tests were carried out in person under similar controlled conditions, using the researcher's equipment. Participant 13 was one single exception, who took part remotely while closely adhering to the established methodology. For this case, the application executable, media folder, and protocol were provided in advance, and the evaluation was conducted on the participant's own computer via a Discord screen-sharing session.

All participants used a computer running the application and headphones for reliable spatial sound reproduction when testing the created soundscape. For the convenience of the participants, all media files were prepared in advance and placed in a dedicated folder within the application's directory. Though not required for the tool's use, this allowed participants to concentrate on evaluating the editor itself rather than creating or searching for external media resources.

5.3.1 User Characterization

Twenty participants took part in the evaluation of the desktop soundscape editor, none of whom had any previous exposure to the tool. Ideally, this study would have involved museum curators or professionals with backgrounds in art interpretation, the editor's target audience. However, due to time constraints and limited connections to such professionals, the participant pool consisted instead of volunteers without formal ties to

the museum sector. Despite this limitation, the study provided valuable insights into the tool's usability and comprehensibility from the perspective of novice users.

The group, caracterized in table 5.2, was composed of fifteen male participants (75%) and five female participants (25%). The age distribution was quite narrow: 35% participants were 22 years old and the remaining 65% were 23 years old, resulting in an average age of 22.7 years.

Table 5.2: User characterization by identifier.

ID	Age	Gender	Experience in sound design
1	22	Male	None
2	22	Male	Intermediate
3	23	Male	None
4	23	Male	None
5	23	Female	None
6	23	Male	None
7	23	Female	None
8	22	Male	Beginner
9	23	Male	Beginner
10	22	Female	None
11	23	Female	None
12	23	Male	None
13	23	Male	None
14	23	Male	None
15	22	Male	None
16	22	Male	None
17	23	Male	Beginner
18	23	Male	None
19	23	Male	None
20	22	Female	None

Regarding experience in sound design, 80% of participants reported no prior background at all. 15% of individuals identified themselves as beginners, and only participant 2 described their experience as intermediate - the most experienced user. Thus, the participant group was predominantly representative of non-specialist users, aligning with the audience for whom the editor was conceived.

5.3.2 Task Overview

Similar to the mobile study in section 5.2.2, participants were guided through six tasks that introduced the core functionalities of the desktop soundscape editor. Reinforcing what was stated in section 5.1, these tasks acted as a structured tutorial to familiarize participants

with the system's interface and feature set, rather than strict evaluation measures in themselves. While tasks were highly specific in general and had clearly defined steps to follow, some allowed for a small degree of freedom, such as experimenting with sound attributes or engaging with the interactive mode.

Task 1 consisted in importing the background painting into the scene, followed by adjusting the grid system used to place, remove and edit scene elements (Task 2). In Task 3, participants used the grid to place the main components of the soundscape: ambient sounds, spatial sounds, and a player starting point. In the most laborious task, Task 4, users edited all existing elements by assigning audio files, adjusting several audio attributes, defining volume attenuation ranges, and setting player properties such as initial orientation and movement speed. Task 5 consisted only in defining a global narration track for the scene, while Task 6 transitioned participants into the interactive mode, allowing them to simulate and test the soundscape they had just created.

This set of tasks exposed participants to a complete workflow of the editor: from constructing and refining a soundscape to finally experiencing it in an interactive preview. With this, participants developed a reasonable and functional understanding of the system in preparation for the optional free-use session that followed.

5.3.3 Results Overview

The evaluation of the soundscape editor produced encouraging results, with participants generally finding the tool intuitive and effective for creating detailed auditory scenes. As with the mobile study's survey (addressed in section 5.2.3), the survey's Likert-scale items were phrased as statements, scored by level of endorsement (1 = strongly disagree, 5 = strongly agree).

As shown in figure 5.6, in terms of basic operations, such as importing the background painting and setting a global narration, participants strongly agreed that these were straightforward, with average agreeability scores nearing 5 for both of these functionalities. Comfortability adjusting the grid's cell size and line thickness was similarly rated, a 4.8. In addition, with an average score of 4.6 out of 5, the available element types (ambient sounds, spatial sounds, and player entity) were generally considered sufficient to adequately represent the painting.

As can be seen in figure 5.7, participants expressed confidence in the clarity and intuitiveness of the implemented placement system, with an average score of 4.1 and editing operations generally perceived as simple and comprehensible. However, participant 12 rated the statement a 2, possibly due to his reported difficulties in choosing audio and narration files when editing the elements. The grid system was considered extremely easy to use by 85% of participants, with an average score of 4.8. Well received as it was, 40% of participants were undecided on whether the system in place was more intuitive and suitable for soundscape creation than alternatives such as drag-and-drop and point-and-click. 10% of users had a preference for these alternative interaction models, while the

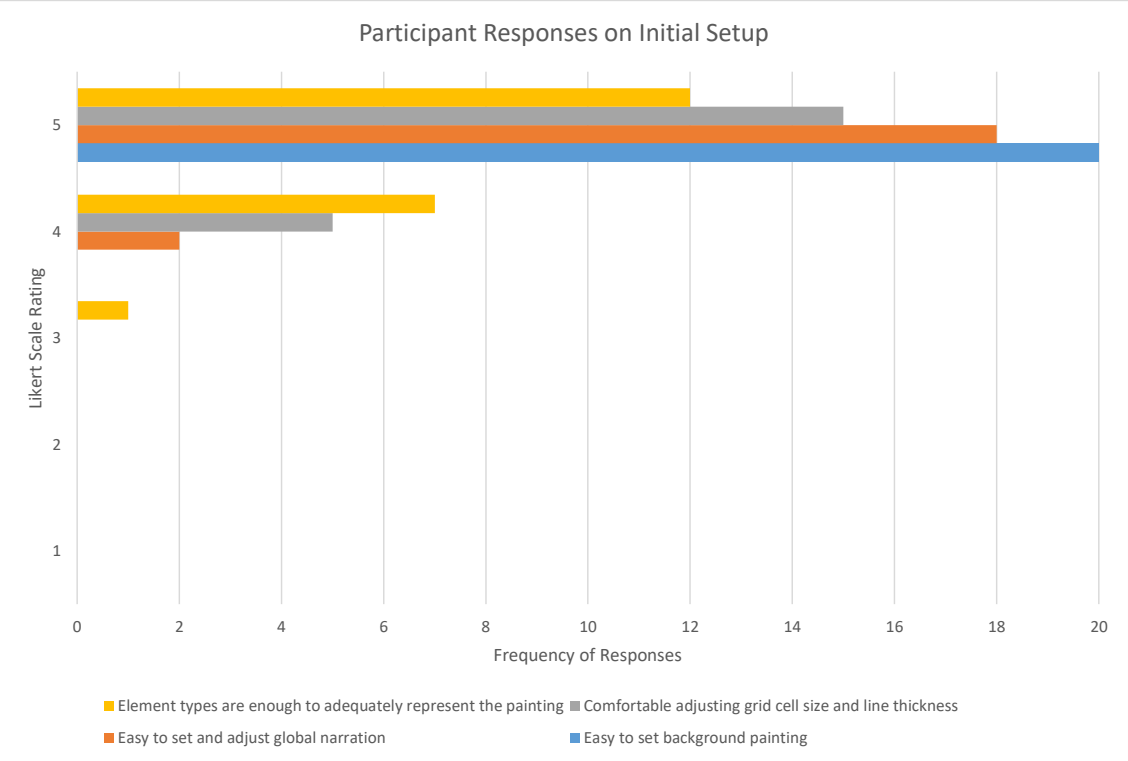

Figure 5.6: Histogram of participant responses regarding setting up the background painting, global narration and initial setup.

remaining 50% preferred the current grid system.

When asked about the adequacy of editing options (responses shown in figure 5.8), participants agreed that the configurable properties of ambient sounds, spatial sounds, and player attributes were sufficient to represent the desired soundscape, all with an average agreeability score equal to or over 4.5/5. Additionally, the available sound editing options were considered very comprehensible, also with a rating of 4.5, suggesting an appropriate balance between simplicity and expressiveness.

Participants responded very positively to the editor's two modes - development and interactive - as one can observe in figure 5.9. With an average score of 5, alternating between modes was classified as extremely quick and intuitive by every participant. Furthermore, their distinct purposes were clearly understood and participants felt capable of effectively emulating and testing the soundscape in interactive mode, having created it within the development mode. Both previous statements were similarly rated and nearing an agreement score of 5, confirming the application's suitability for both soundscape construction and its validation.

According to figure 5.10, only 35% of participants did not feel the absence of a help button. While 20% of users were indifferent, the remaining 45% stated a need for a help button, claiming that it would have been useful to clarify doubts on demand, despite generally completing the tasks successfully without it.

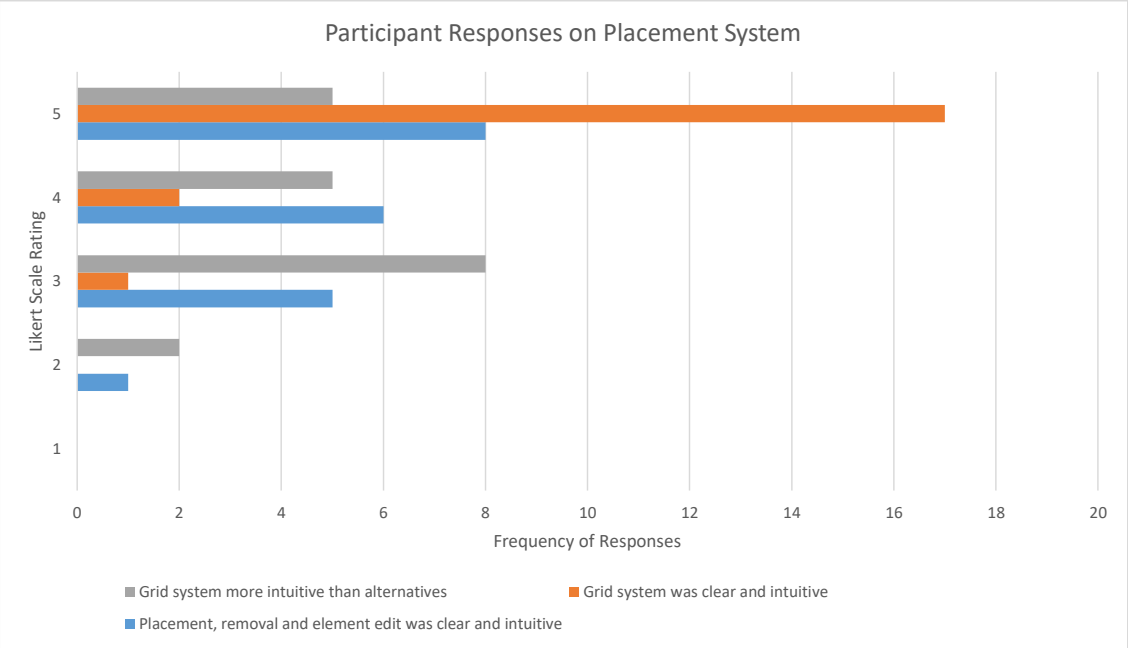

Figure 5.7: Histogram of participant responses regarding the current grid-like placement system.

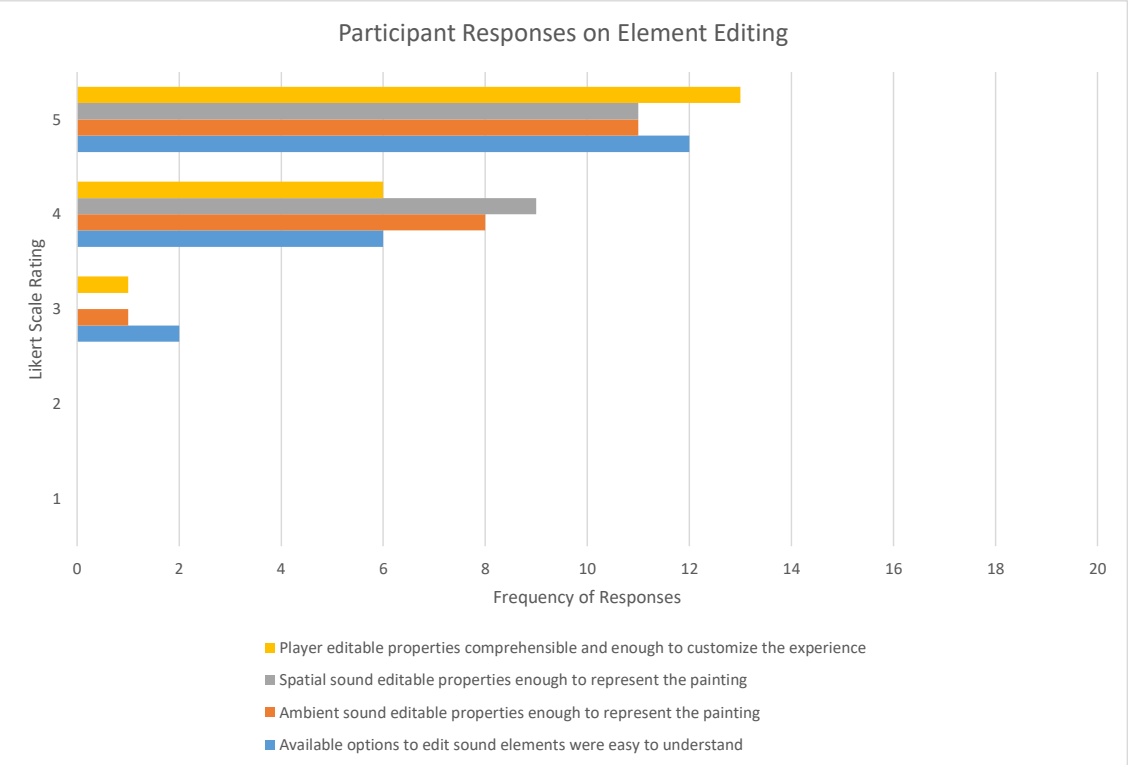

Figure 5.8: Histogram of participant responses regarding the editing options for the existing element types.

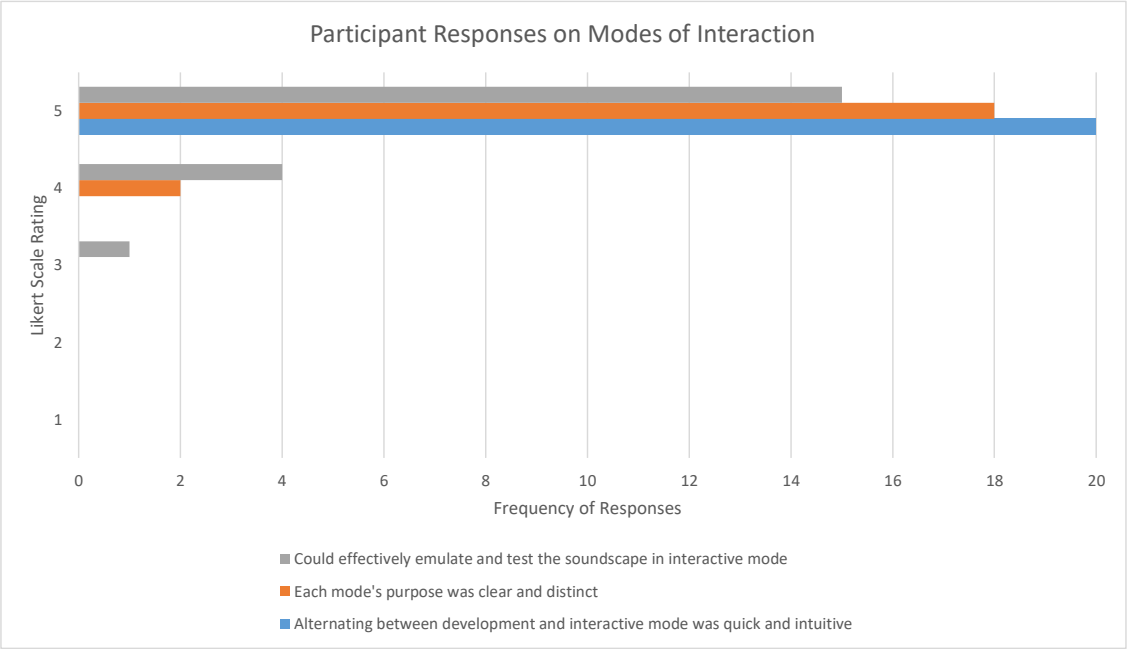

Figure 5.9: Histogram of participant responses regarding the two existing modes of interaction.

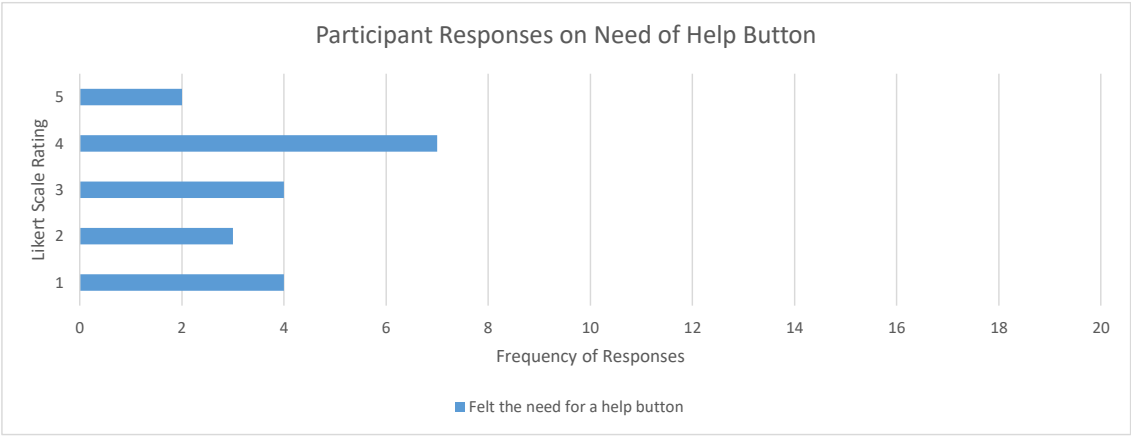

Figure 5.10: Histogram of participant responses regarding the need for a help button.

Lastly, as displayed in figure 5.11, participants strongly considered the editor to be expressive, allowing for the creation of a detailed and immersive representation of "O Bosque Sagrado" (average score 4.6/5). They also described the experience of creation itself as both intuitive and pleasant, rating it an average of 4.8 out of 5.

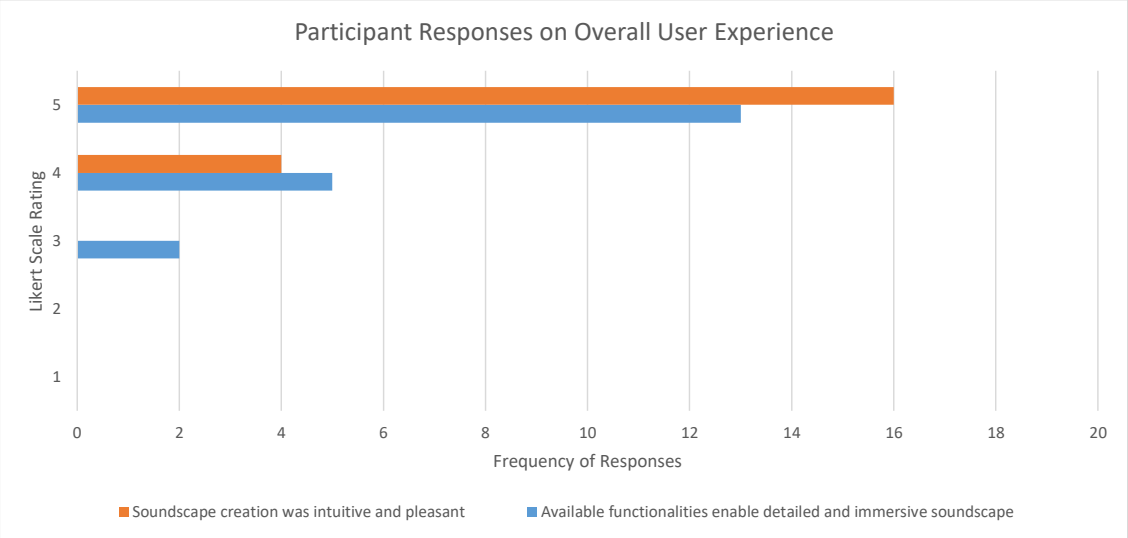

Figure 5.11: Histogram of participant responses regarding the overall user experience.

From the responses to the open-ended questions, derived the most common points for improvement, which included:

- clearer button tooltips, since not all functionalities were implicitely self-explanatory;
- a visible submit/OK button in editing popups, as some participants were uncertain whether changes had been applied;
- an audio preview option within editing popups, to immediately hear the effect of adjustments;
- the ability for most buttons to function as toggles (activating and deactivating features with the same control);
- support for moving already placed elements rather than having to delete and recreate them;
- quality-of-life enhancements, such as a visual indication of edited elements, preventing spatial sound popups from closing after adjusting volume rollof and support for editing multiple elements simultaneously.

The findings of this study confirm that the desktop editor meets its central goals of simplicity and effectiveness for users without former sound design experience, producing

an interactive and immersive soundscape equivalent to the one presented to BVI individuals in section 5.2. Most importantly, these results shed light on a clear set of refinements that could further improve the experience for all of its users.

5.4 Summary

Taking together the results of both sections 5.2.3 and 5.3.3, the two prototypes achieved their intended goals, with clear potential for further improvement. The mobile soundscape player effectively enabled BVI participants to autonomously explore and interpret a painting through audio stimuli - be it sound effects or narrations, movement within the painting and SATs. Besides valuing the accessibility features and sound regulation options, the participants reported a strong sense of immersion, also feeling a solid grasp over the painting's content and essence.Although the console controls were generally preferred over touch gestures, the mobile interface was also considered broadly accessible, likely due to the inclusion of haptic feedback and some gestures somewhat resembling the physical gamepad's bindings.

In turn, the desktop soundscape editor was found intuitive and comprehensible for users with little or no prior experience in sound design. Users were able to construct a complete and immersive soundscape of "O Bosque Sagrado" with ease, mostly appreciating the editor's grid system, editing options, and dual development and interactive modes. The absence of a help feature was the most common limitation identified, along with several smaller usability and quality of life improvements.

These studies confirmed the viability of both components of the proposed system. While the editor empowered curators to create expressive auditory representations of artworks, the mobile application allowed BVI individuals to engage with these representations in a manner that was both immersive and accessible.

6

Conclusion

This chapter provides a detailed overview of the proposed system and briefly exposes the solution's validation methodology. It also addresses the expected technological stack and the envisioned plan for the system's development. The work schedule is split into five distinct and concisely explained tasks mapped in a Gantt chart.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus.

Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

6.1 Future Work

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend

consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Bibliography

[1] D. Ahmetovic, C. Bernareggi, K. Keller, and S. Mascetti. "MusA: artwork accessibility through augmented reality for people with low vision". In: *Proceedings of the 18th International Web for All Conference*. 2021, pp. 1–9 (cit. on pp. 19, 20, 28).

[2] F. Aletta, J. Kang, and Ö. Axelsson. "Soundscape descriptors and a conceptual framework for developing predictive soundscape models". In: *Landscape and Urban Planning* 149 (2016), pp. 65–74 (cit. on p. 13).

[3] B. Ansel. *Two-way Radio Basic Theory - By Bill "BillaVista" Ansell*. URL: http://www.billavista.com/atv/Articles/Offroad_Radios_and_Comms/index.html (visited on 2025-01-04) (cit. on p. 6).

[4] S. Asakawa, J. Guerreiro, D. Sato, H. Takagi, D. Ahmetovic, D. Gonzalez, K. M. Kitani, and C. Asakawa. "An independent and interactive museum experience for blind people". In: *Proceedings of the 16th International Web for All Conference*. 2019, pp. 1–9 (cit. on p. 2).

[5] E. C. Baig. *Accessibility Settings, Tools on Your Smartphone Can Make Life Easier*. URL: https://www.aarp.org/home-family/personal-technology/info-2020/smartphone-accessibility.html (visited on 2024-12-23) (cit. on p. 3).

[6] M. Banf and V. Blanz. "Sonification of images for the visually impaired using a multi-level approach". In: *Proceedings of the 4th Augmented Human International Conference*. 2013, pp. 162–169 (cit. on p. 18).

[7] I. A. for the Prevention of Blindness Vision Atlas. *Global Estimates of Vision Loss*. URL: https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/global/ (visited on 2024-12-23) (cit. on p. 1).

[8] I. A. for the Prevention of Blindness Vision Atlas. *Projected Change in Vision Loss 2020 to 2050*. URL: https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/projected-change/ (visited on 2024-12-23) (cit. on p. 1).

[9] A. L. Brown, J. Kang, and T. Gjestland. "Towards standardization in soundscape preference assessment". In: *Applied acoustics* 72.6 (2011), pp. 387–392 (cit. on p. 13).

[10] F. Candlin. "Blindness, art and exclusion in museums and galleries". In: *International Journal of Art & Design Education* 22.1 (2003), pp. 100–110 (cit. on pp. 2, 3).

[11] F. E. Care. *Vision: Our dominant Sense*. URL: https://www.newjerseyeyesite.com/vision-therapy-optometrist/the-17-visual-skills-assessed-during-your-childs-functional-eye-exam/vision-our-dominant-sense/#:~:text=Vision%20is%20our%20dominant%20sense&text=It%20is%20a%20complex%2C%20learned,activities%20are%20mediated%20through%20vision (visited on 2024-12-23) (cit. on p. 1).

[12] L. Cavazos Quero, J. Iranzo Bartolomé, and J. Cho. "Accessible visual artworks for blind and visually impaired people: comparing a multimodal approach with tactile graphics". In: *Electronics* 10.3 (2021), p. 297 (cit. on pp. 2, 3, 18).

[13] R.-C. Chang, C.-S. Hung, B.-Y. Chen, D. Jain, and A. Guo. "Sound Unblending: Exploring Sound Manipulations for Accessible Mixed-Reality Awareness". In: *arXiv preprint arXiv:2401.11095* (2024) (cit. on p. 2).

[14] T. Deakin. *How museums can remove barriers to access for blind and partially sighted people*. URL: https://www.museumnext.com/article/how-museums-can-remove-barriers-to-access-for-blind-and-partially-sighted-people/ (visited on 2024-12-23) (cit. on p. 2).

[15] K. Drossos, N. Zormpas, G. Giannakopoulos, and A. Floros. "Accessible games for blind children, empowered by binaural sound". In: *Proceedings of the 8th ACM international conference on pervasive technologies related to assistive environments*. 2015, pp. 1–8 (cit. on pp. 21, 22, 28).

[16] S. L. Dumyahn and B. C. Pijanowski. "Soundscape conservation". In: *Landscape ecology* 26 (2011), pp. 1327–1344 (cit. on p. 13).

[17] J. Enoch, L. McDonald, L. Jones, P. R. Jones, and D. P. Crabb. "Evaluating Whether Sight Is the Most Valued Sense". In: *JAMA Ophthalmology* 137.11 (2019-11), pp. 1317–1320. ISSN: 2168-6165. DOI: 10.1001/jamaophthalmol.2019.3537. eprint: https://jamanetwork.com/journals/jamaophthalmology/articlepdf/2752217/jamaophthalmology_enoch_2019_br_190017.pdf. URL: https://doi.org/10.1001/jamaophthalmol.2019.3537 (cit. on p. 1).

[18] R. P. Fartaria, F. Pereira, V. D. Bonifácio, P. Mata, J. Aires-de Sousa, and A. M. Lobo. "NavMol 2.0–a molecular structure navigator/editor for blind and visually impaired users". In: *European journal of organic chemistry* 2013.8 (2013), pp. 1415–1419 (cit. on pp. 18, 19).

[19] C. R. D. Ferreira. "Creating immersive audio in a historical soundscape context". MA thesis. Universidade NOVA de Lisboa (Portugal), 2021 (cit. on pp. 3, 15, 25, 26, 29, 30).

[20] D. Foundation. *The Senses: Vision*. URL: https://dana.org/resources/the-senses-vision/ (visited on 2024-12-23) (cit. on p. 1).

[21] J. Guerreiro, Y. Kim, R. Nogueira, S. Chung, A. Rodrigues, and U. Oh. "The design space of the auditory representation of objects and their behaviours in virtual reality for blind people". In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2763–2773 (cit. on pp. 25, 29, 30).

[22] L. Holloway, K. Marriott, M. Butler, and A. Borning. "Making sense of art: Access for gallery visitors with vision impairments". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–12 (cit. on pp. 2, 3).

[23] E. Kabisch, F. Kuester, and S. Penny. "Sonic panoramas: experiments with interactive landscape image sonification". In: *Proceedings of the 2005 international conference on Augmented tele-existence*. 2005, pp. 156–163 (cit. on p. 15).

[24] S. J. Krol, M. T. Llano, M. Butler, and C. Goncu. "Design Considerations for Automatic Musical Soundscapes of Visual Art for People with Blindness or Low Vision". In: *arXiv preprint arXiv:2405.14188* (2024) (cit. on pp. 2, 3, 25).

[25] N. Kumar. *How Many People Own Smartphones (2024): Worldwide Data*. URL: https://www.demandsage.com/smartphone-users/ (visited on 2024-12-23) (cit. on p. 3).

[26] F. M. Li, L. Zhang, M. Bandukda, A. Stangl, K. Shinohara, L. Findlater, and P. Carrington. "Understanding visual arts experiences of blind people". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–21 (cit. on pp. 2, 3).

[27] J. Li. "Beyond Sight: Enhancing Augmented Reality Interactivity with Audio-Based and Non-Visual Interfaces". In: *Applied Sciences* 14.11 (2024), p. 4881 (cit. on pp. 2, 18).

[28] J. J. Lopez, P. Gutierrez-Parera, and M. Cobos. "Compensating first reflections in non-anechoic head-related transfer function measurements". In: *Applied Acoustics* 188 (2022), p. 108523. ISSN: 0003-682X. DOI: https://doi.org/10.1016/j.apacoust.2021.108523. URL: https://www.sciencedirect.com/science/article/pii/S0003682X21006174 (cit. on p. 12).

[29] P. R. Martins. "Blindness in art museums: A Portuguese case study". In: *Journal of museum education* 45.3 (2020), pp. 340–349 (cit. on p. 2).

[30] K. McMullen. "Interface Design Implications for Recalling the Spatial Configuration of Virtual Auditory Environments." In: (2012-01) (cit. on p. 10).

[31] V. Nair, J. L. Karp, S. Silverman, M. Kalra, H. Lehv, F. Jamil, and B. A. Smith. "Navstick: Making video games blind-accessible via the ability to look around". In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2021, pp. 538–551 (cit. on pp. 3, 20, 28, 30).

[32] V. Nair, S.-e. Ma, R. E. Gonzalez Penuela, Y. He, K. Lin, M. Hayes, H. Huddleston, M. Donnelly, and B. A. Smith. "Uncovering visually impaired gamers' preferences for spatial awareness tools within video games". In: *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 2022, pp. 1–16 (cit. on pp. 3, 22, 23, 28–30).

[33] V. Nair, S.-e. Ma, H. Huddleston, K. Lin, M. Hayes, M. Donnelly, R. E. Gonzalez, Y. He, and B. A. Smith. "Towards a Generalized Acoustic Minimap for Visually Impaired Gamers". In: *Adjunct Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*. 2021, pp. 89–91 (cit. on p. 22).

[34] U. Nations. *Convention on the Rights of Persons with Disabilities (CRPD)*. URL: https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-crpd (visited on 2024-12-23) (cit. on p. 2).

[35] W. H. Organization. *Blindness and vision impairment*. URL: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (visited on 2024-12-23) (cit. on p. 1).

[36] RAIZ. *O Bosque Sagrado*. URL: https://www.pubnub.com/guides/unity/ (visited on 2025-02-11) (cit. on p. 28).

[37] K. Rector, K. Salmon, D. Thornton, N. Joshi, and M. R. Morris. "Eyes-free art: Exploring proxemic audio interfaces for blind and low vision art engagement". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3 (2017), pp. 1–21 (cit. on pp. 2, 3, 15, 16).

[38] M. Risoud, J.-N. Hanson, F. Gauvrit, C. Renard, P.-E. Lemesre, N.-X. Bonne, and C. Vincent. "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. 8).

[39] M. Risoud, J.-N. Hanson, F. Gauvrit, C. Renard, P.-E. Lemesre, N.-X. Bonne, and C. Vincent. "Sound source localization". In: *European annals of otorhinolaryngology, head and neck diseases* 135.4 (2018), pp. 259–264 (cit. on pp. 7–12).

[40] I. N. Rodrigues. *Espacialização de Som no Navegador e Editor Molecular Navmol Dissertação para obtenção do Grau de Mestre em Engenharia Informática*. URL: https://github.com/joaomlourenco/unlthesis (cit. on p. 18).

[41] F. Rumsey. *Spatial Audio*. 2001-01. ISBN: 9780080498195. DOI: 10.4324/9780080498195 (cit. on pp. 8–12).

[42] M. Rychtáriková, J. Herssens, and A. Heylighen. "Towards more inclusive approaches in soundscape research: The soundscape of blind people". In: *Inter-noise and noise-con congress and conference proceedings*. Institute of Noise Control Engineering New York. 2012 (cit. on p. 13).

[43] J. Sánchez and M. Sáenz. "Usability of Audio-Based Virtual Environments for Users with Visual Disabilities". In: *Virtual Reality and Human Behavior Symposium, LAVAL Virtual*. 2007, pp. 18–22 (cit. on pp. 2, 3, 20).

[44] R. Saville. *Audio, Image and Video Processing*. URL: https://beccasaville.blogspot.com/2012/09/compression-and-rarefaction.html (visited on 2025-01-04) (cit. on p. 7).

[45] R. M. Schafer. *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster, 1993 (cit. on p. 12).

[46] A. R. L. Simão. "Jogo sério para treino de orientação e mobilidade de crianças cegas e amblíopes". PhD thesis. Universidade Nova de Lisboa, 2018 (cit. on pp. 21, 28, 30).

[47] A. E. Team. *77% of People Say Sight is Their Most Important Sense*. URL: https://assileye.com/blog/is-sight-the-most-important-sense/ (visited on 2024-12-23) (cit. on p. 1).

[48] D. R. Team. *Technical Overview of Unity Game Engine*. URL: https://www.pubnub.com/guides/unity/ (visited on 2025-02-01) (cit. on p. 30).

[49] P. Vasilakou, S. Mineiko, T. M. Hasioti, Z. Gavriilidou, and A. Drigas. "The accessibility of visually impaired people to museums and art through ICTs". In: *Technium Soc. Sci. J.* 35 (2022), p. 263 (cit. on p. 2).

[50] R. Vaz, D. Freitas, and A. Coelho. "Blind and visually impaired visitors' experiences in museums: increasing accessibility through assistive technologies". In: *The International Journal of the Inclusive Museum* 13.2 (2020), p. 57 (cit. on p. 2).

[51] R. Vaz, D. Freitas, and A. Coelho. "Perspectives of visually impaired visitors on museums: towards an integrative and multisensory framework to enhance the museum experience". In: *Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*. 2020, pp. 17–21 (cit. on p. 2).

[52] J. Yang and C. Y. Chan. "Audio-augmented museum experiences with gaze tracking". In: *Proceedings of the 18th international conference on mobile and ubiquitous multimedia*. 2019, pp. 1–5 (cit. on pp. 2, 3, 17, 30).

[53] W. Yost and R. Schlauch. "Fundamentals of Hearing: An Introduction (4th edition)". In: *Journal of The Acoustical Society of America - J ACOUST SOC AMER* 110 (2001-10). DOI: 10.1121/1.1398047 (cit. on pp. 7–12).

[54] Zenva. *What is Unity? – A Top Game Engine for Video Games*. URL: https://gamedevacademy.org/what-is-unity/ (visited on 2025-02-01) (cit. on p. 30).

[55] X. Zhong, W. Yost, and L. Sun. "Dynamic binaural sound source localization with ITD cues: Human listeners". In: *Journal of the Acoustical Society of America* 137 (2015-04), pp. 2376–2376. DOI: 10.1121/1.4920636 (cit. on p. 9).