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

Dissertation Plan

MASTER IN COMPUTER SCIENCE AND ENGINEERING

NOVA University Lisbon

Draft: *February 6, 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

Dissertation Plan
MASTER IN COMPUTER SCIENCE AND ENGINEERING

NOVA University Lisbon
Draft: *February 6, 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 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

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 [53]. 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 [60]. 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 [1, 17].

Unfortunately, sight is not a universal privilege. According to the International Agency for the Prevention of Blindness Vision Atlas [21], 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 [37], 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 [8] states that everyone if they live long enough, will at least 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, that is not always the case. Exhibitions have historically been visual, as most available art is specifically designed to be consumed that way and due to numerous efforts to preserve it. Such 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 [11, 28, 57, 58, 59].

Though 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 [23, 28, 31] – article 30 of the UN Convention on the Rights of Persons with Disabilities [13].

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 [5, 23, 28, 31, 38]. These efforts tend to be expressed as: tactile replicas and graphics, pre-recorded audio descriptions, specialized tours, workshops, large print and labels in braille, among several others. Though each method serves a relevant and unique purpose, they are not without limitations, which will be enumerated in section 1.2.

Not to undermine the progress achieved over the last few years, as 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, 23, 38, 57]. As a result, low-vision people rarely attend these institutions - in Europe, only 5.5% of them do [58, 59]. The aforementioned ocular centricity is a primary reason for this low affluence, as it significantly limits not only their mobility but also their access to information, which in turn dramatically hinders their independence as well [11, 28, 57].

Such a low percentual for museum adherence is particularly demoralizing, as it is a common observation that BVI people do enjoy and express the desire to visit galleries and experience visual art [5, 10, 23, 27, 28], and much for the same reasons people with sight do [10]. However, they must be provided with the proper access to do so [23] and do not want to rely on others [5, 27] constantly.

While the numerous accessibility methods currently employed at museums merit their welcome addition to any exhibition, these often tend to understate the importance of aestheticism in experiencing the artwork [27, 28, 31]. Instead, the focus leans heavily on describing and educating, sometimes at the cost of a piece's sensory, emotional, and immersive dimensions.

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 [31]. Approaches combining music and

soundscapes have shown promise but can be time and resource-intensive [27].

Technological innovations such as spatial audio, haptics, VR, and AR have enabled the creation of increasingly immersive experiences [12, 29, 46, 62], allowing not only for multisensory engagement with art but also opportunities for remote artistic appreciation. Despite the risk of less engagement than in person [28] and difficulty navigating through a web gallery [52], digital platforms and virtual tours can tear down physical barriers by enabling individuals to experience art from the comfort of home or anywhere else, at their own pace, and unshackled by social tension [25, 28].

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 [46, 62]. 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 visualize virtual worlds [46]. Frequently enhanced with 3D effects to improve spatial perception and experience, 3D audio benefits navigation and immersive engagement in museum and gallery settings [62].

Primarily supporting the above mentioned technologies, the smartphone is the perfect vehicle for accessibility as about 54% of the global population owns at least one [24], and it is rich in accessibility features among several others [2]. 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, 11, 23, 38]. Without delving into too much detail, tours and workshops are infrequent and inconsistent, must be reserved in advance, and are 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 [11].

While a tactile approach seems to be the preferred form of interaction with artwork for BVI individuals [27, 28] since it allows them to feel the artwork up close and personal and sense its various features at a low level, high-level information about the piece is quite limited and the combination of preservation efforts and barriers of intellectual access with the still prevalent visual centricity of exhibitions makes it so that these types of programs are a rare occurrence among museums [28].

In the context of this dissertation, we intend to tap into the potential of spatial audio to create an immersive remote experience that can convey the spatial and emotional dimensions of art through interactive three-dimensional soundscapes. To achieve this, we draw inspiration from the use of spatial audio in other research and accessible games. Although video games may not be the first medium that comes to mind when considering

the visual art experience in museums, they share certain parallels. As an art form and medium of expression, video games focus on immersion and engagement, much like museums aim to captivate visitors. Over time, video games have also evolved to become more accessible and appeal to a broader audience through the innovative use of spatial audio to guide navigation, evoke emotion, and tell a rich, interactive story. This growth aligns closely with our objectives, and while our proposal is not to develop a video game, it adopts a gamified approach to reimagining how art can be experienced inclusively.

Thus, we propose creating a tool that allows museum curators to build a spatial audio environment representing art pieces. It is an intuitive and user-friendly soundscape editor with support for immersive audio and interactivity, not requiring prior experience in more sophisticated tools with a harsh learning curve. The generated environments are then made available via Android mobile devices, where blind and visually impaired clients 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 Expected Contributions

This research is expected to yield the following contributions:

- An intuitive soundscape editor for museum curators to create spatial audio environments representing art pieces, requiring no prior expertise in complex audio design.
- An Android Package Kit (APK) enabling BVI users to explore the generated artwork recreations through immersive 3D audio, providing an experience that is both aesthetic and informational while promoting independence with simple controls.
- A cost-effective approach to accessibility using widely available technologies, such as smartphones and headphones, allowing for both remote and on-site interaction with art.
- Validation through usability testing and feedback from BVI users in order to evaluate the solution's effectiveness in real-world scenarios.

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 The Fundamentals Of Sound

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.

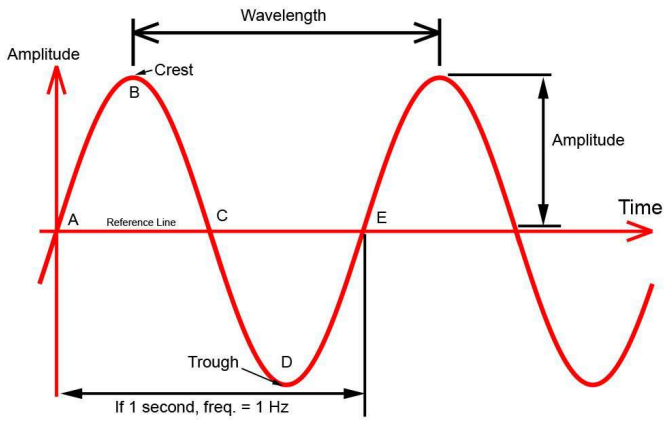

Figure 2.1: Graphical representation of a simple sine wave (Adapted from [54]).

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 [63].

2.2 Sound Perception

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 [63].

2.3 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 [63]. 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 20ºC) [42].

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.

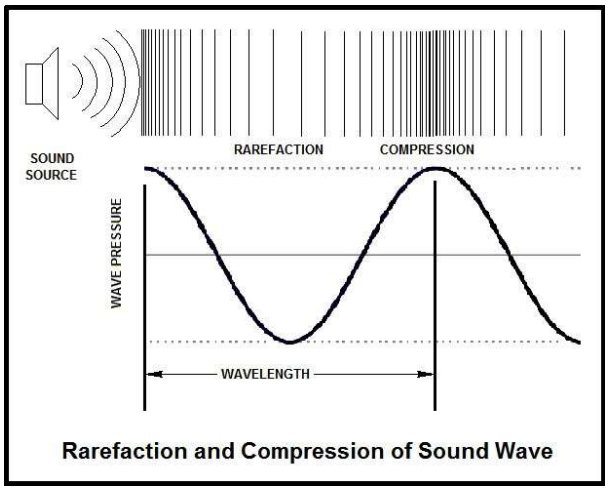

Figure 2.2: Representation of rarefaction and condensation zones (Adapted from [6]).

Sound waves propagate in all directions from a source (circularly in 2D, spherically in 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 [63].

2.4 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 [63].

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) [42, 63]. These are properly illustrated in Figure 2.3.

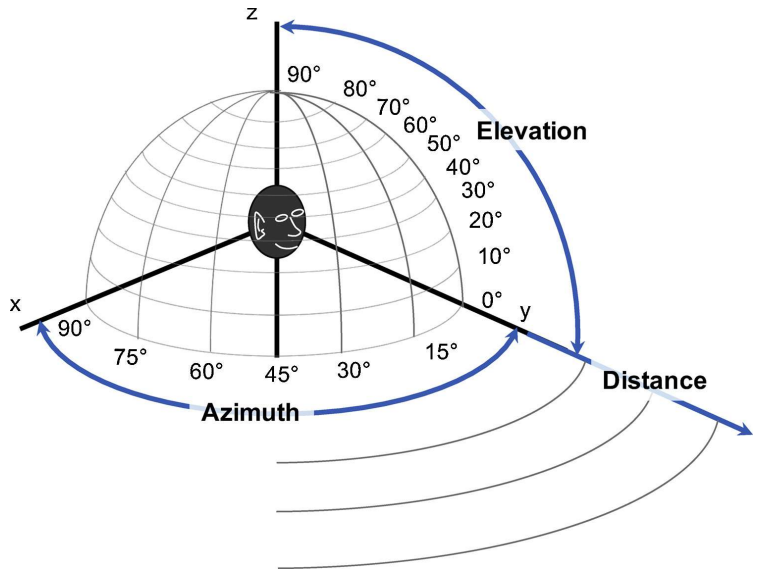

Figure 2.3: Polar coordinates used to locate a sound source in a 3D space centered on the listener (Adapted from [41]).

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 [44].

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 2.4.1.

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 [42]. These alterations are captured by a Head Related Transfer Function (HRTF), which we will delve into in section 2.4.3.

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 [63].

2.4.1 Interaural Differences

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

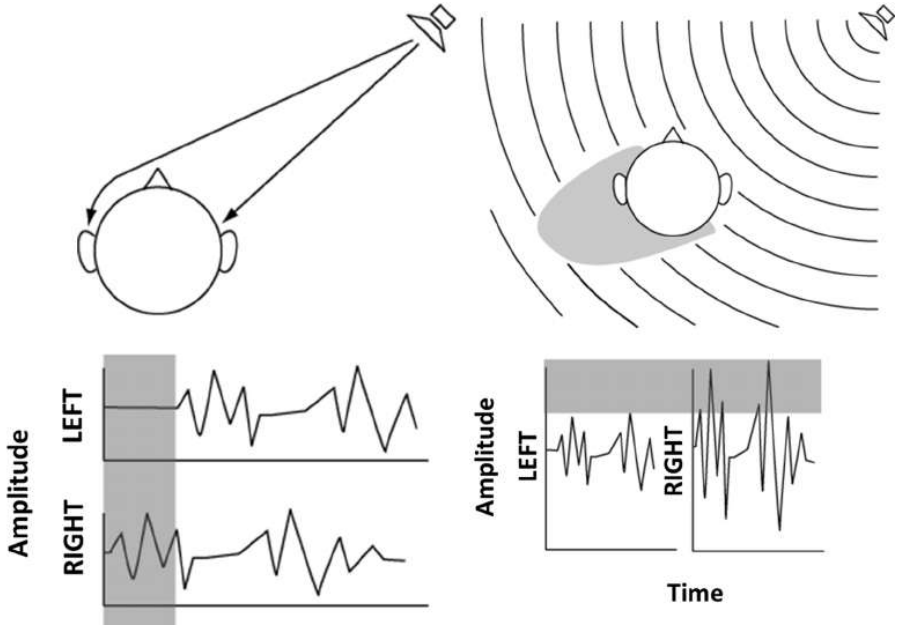

Figure 2.4: Interaural Time Difference and Interaural Level Difference, respectively (Adapted from [64]).

These mechanisms depend on the sound signal's nature and can be influenced by environmental cues that introduce conflicting information [44]. One may classify interaural differences into two primary categories:

- **ITD - Interaural Time Difference:** The 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 IPDs (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 IPD becomes ambiguous [63], and one can no longer tell which ear is leading or lagging [42, 44].

- **2. Interaural Level Difference (ILD):** Also known as the Interaural Intensity Difference, the 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 2.3) [44, 63]. However, the extra distance the wave travels to the farther ear is negligible and the intensity differences minimal [44].

  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 [63]. The same cannot be said for low frequencies, at which the head is not a decent barrier to sound [44].

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 2.4.2.

2.4.2 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 [42, 63].

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 [42, 63]. The plane in which these ambiguous points lie is dubbed the mid-sagittal plane [63], and it forms the so-called cone of confusion (see Figure 2.5), its axis being the interauricular line [42].

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 [42]. Every sound source location has its cone of confusion describing the other possible locations producing the same interaural differences [63].

The most intuitive way to reduce this ambiguity is our natural head movements [42, 44, 63]. 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 Head-Related-Transfer-Functions [42, 63] (we will address them in some detail in section 2.4.3), which are not only critical in determining the vertical direction of sound but also in distinguishing between the ambiguous sounds along the azimuthal plane [42].

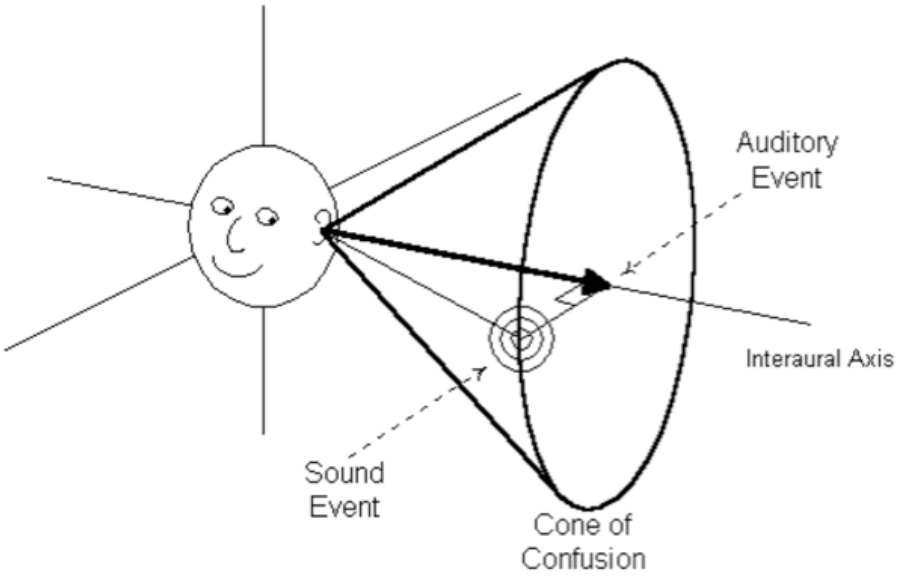

Figure 2.5: Cone of confusion, in which ITDs and ILDs are indistinguishable (Adapted from [32]).

2.4.3 Head Related Transfer Function

As explained in section 2.4.2, 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) [63]. Furthermore, the listed physical features interfere with an incoming sound wave through reflection, diffraction, and absorption, attenuating or delaying specific frequencies composing it [42, 44].

These effects, caused by the filtering action of the pinna and body, are known as spectral cues [63] and are determined monaurally [44]. Depending on both the position of the sound's source as well as its angle of incidence [44, 63], 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) [63].

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) [42]. Figure 2.6 shows how it may be measured.

Vertical localization is generally less accurate than its horizontal counterpart. However,

Figure 2.6: HRTF measurements in an anechoic chamber (Adapted from [30]).

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 [63].

Just as individuals greatly vary in their morphology, so does the shape and size of their pinna. Consequently, each person has a unique set of HRTFs learned throughout their lives [42, 44]. Additionally, research has found that amplifications or attenuations in particular regions of the frequency spectrum seem to correspond to specific sound source positions [44].

Generalizing the spectral characteristics that compose HRTFs across a vast demographic is convenient as it simplifies their implementation process while ensuring reasonable accuracy for many listeners. However, to no surprise, an individual's best HRTFs are their own [44].

Nonetheless, there have been studies in which subjects are fed audio signals through HRTFs other than their own [42, 44] – to find their ability to localize sound significantly hampered [44]. After some time, the participants appear to begin learning the HRTFs they have been given and eventually regain normal vertical localization [42]. Some people are known to localize audio better than others, so their HRTFs tend to be considered more applicable for general use [44].

Delivering audio stimuli over headphones is convenient and establishes a controlled environment[63]. However, certain headphone types induce a feeling of the sound emanating from inside the head, rather than the natural tri-dimensional externalization (azimuth, elevation, and range) we are accustomed to [44, 63].

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.

Luckily, 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 [63].

2.5 Soundscape

By R.M. Schafer's [47] 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 [4, 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 [4, 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 [4].

In the context of this dissertation, we are particularly interested in the role of the soundscape as an accessibility tool for 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.

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 [45].

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. [26] 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. [38] designed Eyes-Free Art, 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.

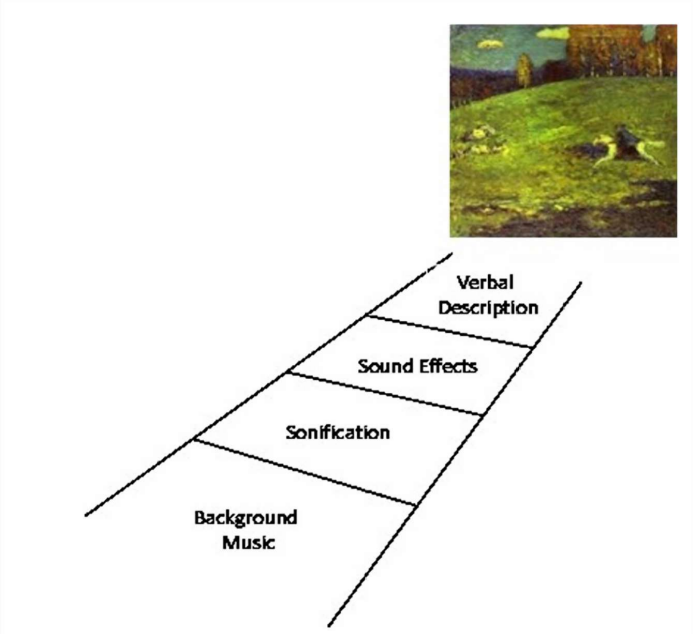

Figure 3.1: The four zones of the Eyes-Free Art proxemical interface. Image is *The Blue Rider* from Wassily Kandinsky (Reproduced from [38]). © 2017 Copyright held by the owner/author(s).

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. [62] 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. At the same time, the Unity3D game engine was employed to model the room and map the various sound sources.

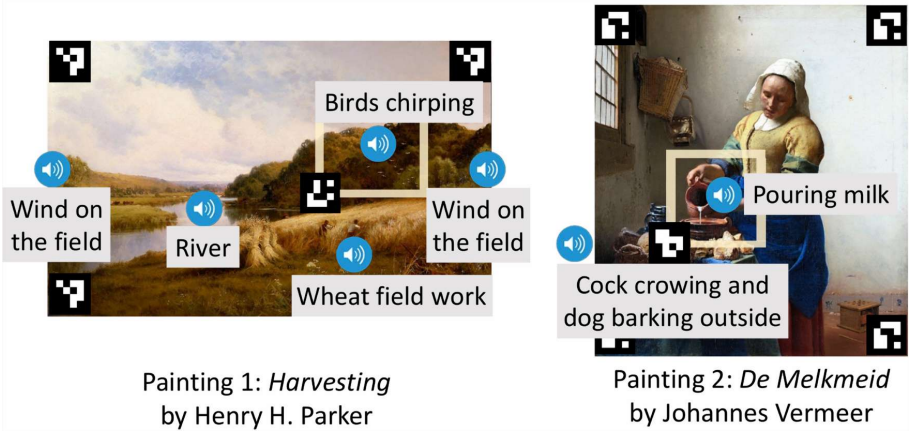

Figure 3.2: Two of the paintings used in the application. Blue audio icons represent virtual sounds accurately spatialized relative to the user (b). (Adapted from [62]). © 2019 Copyright held by the owner/author(s).

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.

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 [29] developed an audio-only inclusive prototype for navigating 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. [11] 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 [7] 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 [43] 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

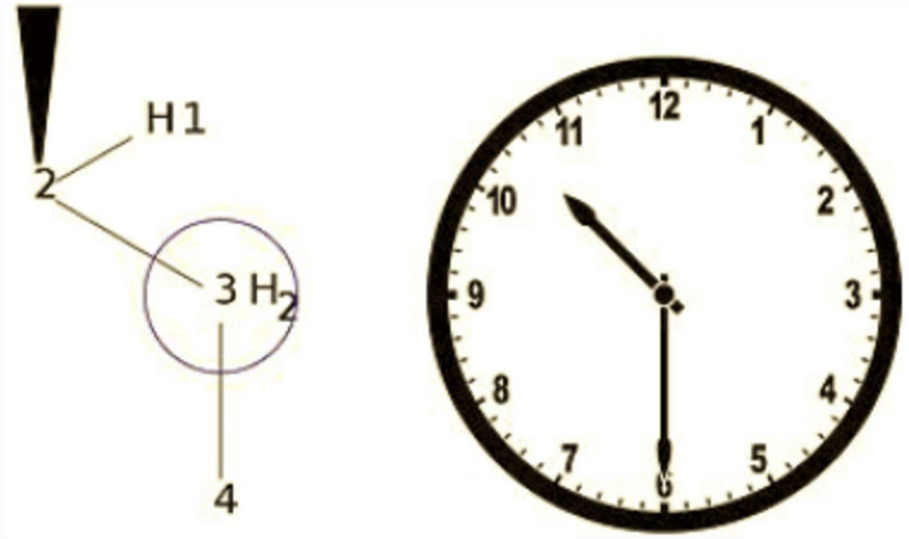

Figure 3.3: Navmol's clock system, where carbon-3 has neighbors at 6 o'clock and 10 o'clock, respectively carbon-4 and carbon-2 (Adapted from [18]).

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 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. [3] 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 LV 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.

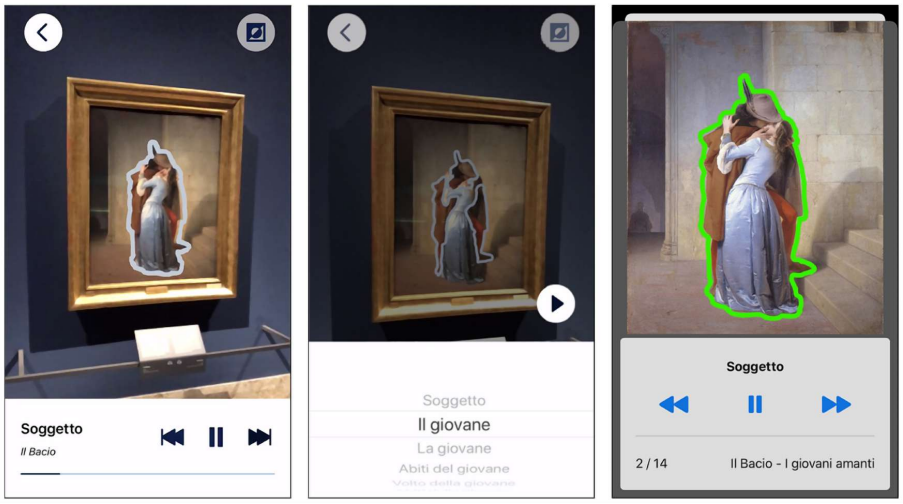

Figure 3.4: Some screens of the MusA app. The first two screens respectively correspond to chapter navigation and selection in its first iteration. The third screen corresponds to the second iteration's virtual mode (Adapted from [3]). © 2021 Copyright held by the owner/author(s).

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 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. [34] 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 [46] 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 [48] 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.

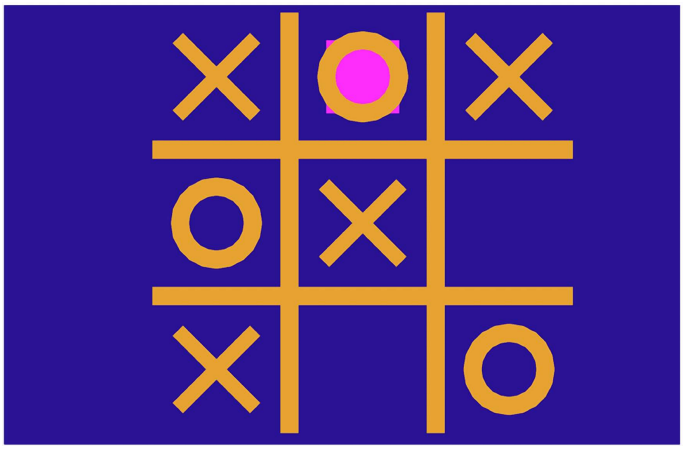

Figure 3.5: Part of the Tic-Tac-Toe game's graphical user interface (Adapted from [15]). © 2015 ACM.

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 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 spatial awareness tools (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 in accessibility, Nair et al. [35] 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. [36] 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
- *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
- *Simple Audio Menu* – Lists the points of interest in a room through their corresponding sounds effects and text-to-speech, in alphabetical order

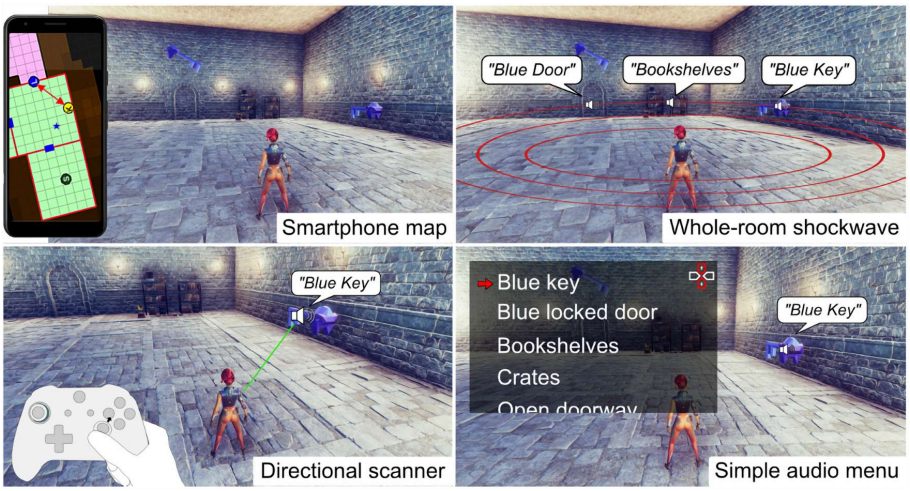

Figure 3.6: The four spatial awareness tools implemented within Dungeon Escape [36]). © 2022 ACM.

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, 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, Dungeon Escape 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 [14] stands out as the industry's standard in delivering 3D audio experiences, be it in entertainment, gaming, or VR. In Dolby Atmos [14], environment creation is object-based, meaning that sounds may be placed and moved as discrete objects in a 3D space. Sound Particles [49] 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 [55] and Unreal Engine [56] have proven effective platforms for soundscape creation, offering robust audio capabilities and integrations. Unity [55], 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 [20] and Google's Resonance Audio [40]. Unreal Engine [56] features a built-in audio engine supporting 3D spatialization, reverberation effects, and audio triggers, more advanced than Unity 's [55] 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 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. [22] 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. Meanwhile, Krol et al. [27] 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

Contributing to the PASEV project, focused 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, 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, allowing 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.

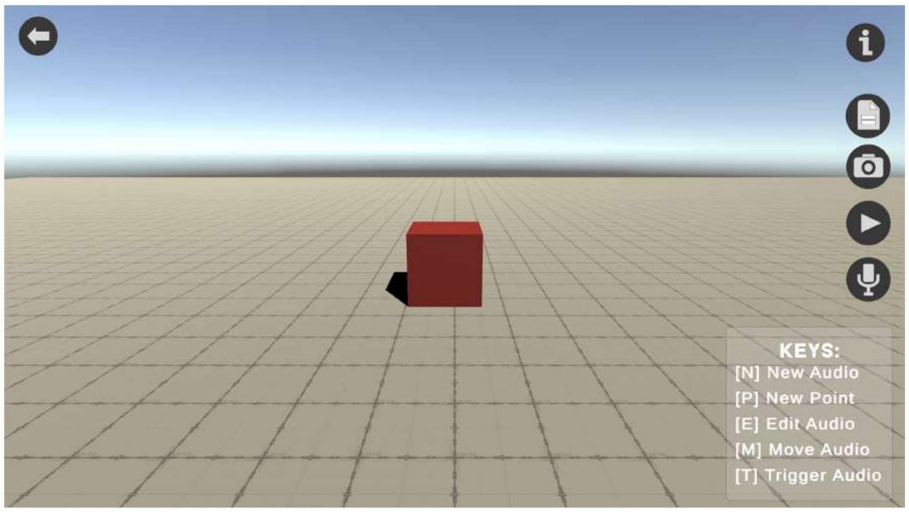

Figure 3.7: Immerscape's environment, through the player view perspective. The red cube represents an audio object (Reproduced from [19]). © 2021 Carolina Ribeiro Dias Ferreira.

An evaluation with 10 developers and 7 non-developer participants highlighted Immerscape's high usability and immersive quality. It is one of many examples where spatialized audio with the HRTF technique was proven to be an effective way of providing engaging sonic experiences, in this case, by immersively 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 undoubtedly 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 current dissertation focuses on accessibility for the blind and visually impaired in experiencing visual art, specifically through the use of audio generation techniques. It aims to provide an immersive auditory experience through interactive 3D sound environments representative of specific paintings.

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.

Thus, this system we aim to develop has two very different symbiotic parts. One is the de facto editing of 3D sound environments in a desktop application, catering to users without previous experience in sound engineering and related software. Just as important, the other part concerns the interactivity, functionality, and accessibility of said environments in mobile devices operated by BVI individuals and is to be experienced with headphones. The system abstracts the complexities of spatializing sound sources and optimizing accessibility, automatically generating BVI-accessible environments upon exportation.

Following the implementation, the proposed system will undergo a structured validation process on the two fronts it caters to through interviews and questionnaires after user testing. The tool will be evaluated by individuals with an artistic background and little to no experience in sound design, and testing will focus on its ease of use, usability, and overall expressiveness. 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 3D Soundscape Editor for Windows

Inspired by Ferreira's [19] work on Immerscape (addressed in section 3.4.1), the elaborated tool 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 's [33] 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 has 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 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 packages the scene's data (possibly in JSON or XML) and stores it in a designated folder.

4.1.2 Mobile Soundscape Player (APK-based proof of concept)

Loading up a scene exported from the developed soundscape editor and then parsing it, the mobile application displays 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 [3, 15, 48].

Navigation through the virtual space happens through a virtual joystick control schema, inspired by Nair et al.'s [36] Dungeon Escape game and other mainstream 3D games such as the earliest Resident Evil [39] 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 controls 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 works as a directional scanner, akin to Nair et al.'s [34] 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 triggers a complete verbal description of the scene (predefined by the environment's creator) and pauses all other sounds apart from ambient sounds. Obstacles the user encounters trigger real-time audio feedback in the form of specific collision detection sound cues, once again inspired by Nair et al.'s [36] Dungeon Escape.

Especially for BVI people, cognitive overload is a significant concern [22, 36]. 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).

4.2 Technological Stack

We selected the Unity [55] 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, 22, 34, 36, 48, 62] in chapter 3.

A real-time 3D development platform, while Unity [55] 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 [51, 55, 61]. 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 [55] 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 [56].

With a built-in 3D audio system, Unity [55] 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 [40] for integration within Unity [55], as it offered relevant features such as HRTF processing and environmental modeling. Furthermore, some of the analyzed studies [19, 48, 62] 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 [55] engine.

For these reasons, we turned to the Steam Audio [50] framework instead, which has been actively maintained, well-documented, and compatible with Unity [55] engine updates. Moreover, we were motivated by the use of this toolkit in NavStick [34] and Dungeon Escape [36], two of the most influential works for our proposal. Supposedly easy to implement and deploy, Steam Audio [50] encompasses high-fidelity HRTF-based binaural rendering and geometry-based occlusion, reflection, and reverberation effects, among other spatial features contributing to natural sounding immersion.

4.3 Work Plan

To adequately manage and organize the work to be developed throughout this dissertation, the proposed solution's implementation will constitute five distinct stages of development, each representing a relevant milestone:

- **Initial Research & Concept Refinement:** Initially, some time will be reserved to develop simple conceptual mockups of both the desktop editor and the mobile prototype, showcasing their key functionalities and accessibility considerations. These will then be presented to supposed end-users, such as non-experts in sound design and BVI individuals. The feedback gathered will assess whether the currently proposed feature set is adequate for the audience's needs and ultimately refine it.

- **Soundscape Editor (PC) – Development:** Upon finalizing the architecture details for both applications based on user feedback and confirming what technologies will be utilized, the focus will shift onto the editor's core development. Such will mainly involve integrating advanced spatial audio with the Steam Audio [50] framework and implementing movement alongside an expressive block-based system. Other subtasks include developing a simple UI for adjustable settings, scene creation, editing, and erasure. Since the proposed interactive mode emulates mobile interaction, it will likely be developed alongside it.

- **Mobile Interactive Prototype – Development:** The design of the mobile interaction's accessible map view parsed out of a previously editor-generated scene will be the first subtask under focus. Immediately after comes the joystick-based navigation, assisted with snap rotation and directional scanning. Finally, spatialized audio feedback

and collision detection will be integrated alongside the ability to toggle between sequential and concurrent stimuli, adjusting audio cardinality and triggering the scene's verbal description.

- **Results and System Evaluation:** Before proceeding with evaluation, we will take the time to thoroughly test both parts of the developed system and debug it where needed. We will then define an adequate methodology for evaluating each part of the solution, catering to its specific end-user demographic. Having established the evaluation methodology to apply, user tests with the targeted audiences will be conducted. The gained insights will be used to assess the system's quality both in usability and accessibility.

- **Dissertation Writing & Revisions:** The dissertation document will be written in parallel with some of the aforementioned tasks and will mainly incorporate the implementation details and evaluation results into the discussion. Additionally, previously written chapters may be revisited and refactored where needed.

The following chart (Figure 4.1) exposes the expected duration of each of the tasks defined in the proposed work plan.

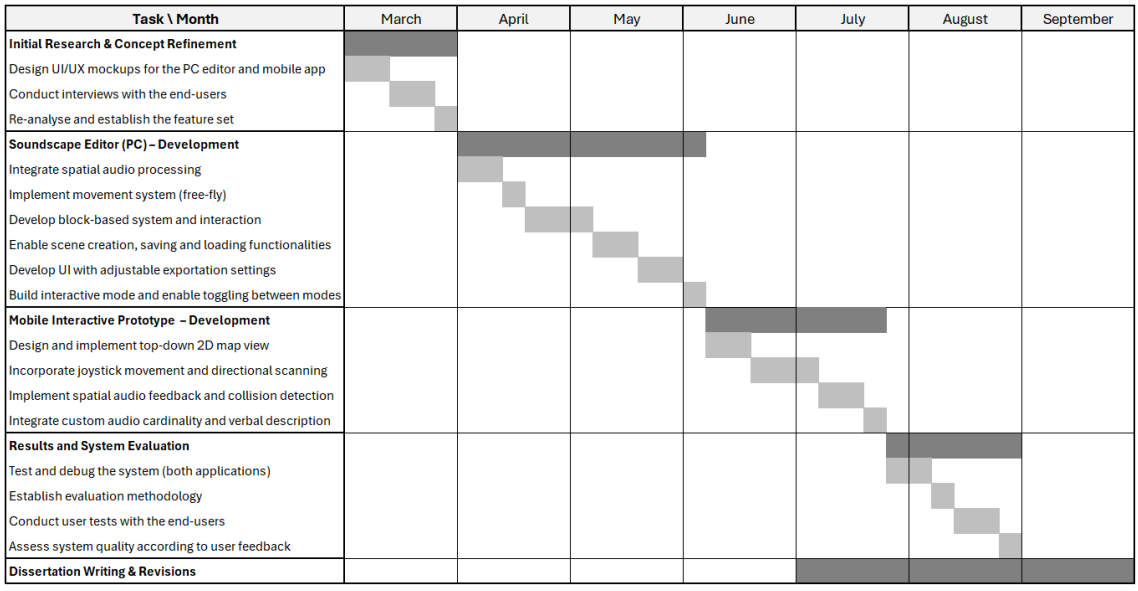

Figure 4.1: Gantt chart displaying the proposed work distribution over the remaining months.

Bibliography

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

[2] *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).

[3] D. Ahmetovic et al. "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. 18, 19, 27).

[4] 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).

[5] S. Asakawa et al. "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).

[6] *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).

[7] 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. 17).

[8] *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).

[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] 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, 17).

[12] R.-C. Chang et al. "Sound Unblending: Exploring Sound Manipulations for Accessible Mixed-Reality Awareness". In: *arXiv preprint arXiv:2401.11095* (2024) (cit. on p. 3).

[13] *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).

[14] *Dolby Atmos*. URL: https://www.dolby.com/technologies/dolby-atmos/ (visited on 2025-01-21) (cit. on p. 23).

[15] K. Drossos et al. "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. 20, 21, 27).

[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 et al. "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 et al. "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. 17, 18).

[19] C. R. D. Ferreira. "Creating immersive audio in a historical soundscape context". MA thesis. Universidade NOVA de Lisboa (Portugal), 2021 (cit. on pp. 14, 24, 25, 27–29).

[20] *FMOD*. URL: https://www.fmod.com/ (visited on 2025-01-21) (cit. on p. 24).

[21] *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).

[22] J. Guerreiro et al. "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. 24, 28).

[23] L. Holloway et al. "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).

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

[25] *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. 3).

[26] 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. 14).

[27] S. J. Krol et al. "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, 24).

[28] F. M. Li et al. "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).

[29] 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. 3, 17).

[30] 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).

[31] 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).

[32] K. McMullen. "Interface Design Implications for Recalling the Spatial Configuration of Virtual Auditory Environments." In: (2012-01) (cit. on p. 11).

[33] *Minecraft*. URL: https://www.minecraft.net/en-us (visited on 2025-02-01) (cit. on p. 27).

[34] V. Nair et al. "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. 19, 28, 29).

[35] V. Nair et al. "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. 21).

[36] V. Nair et al. "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. 21, 22, 27–29).

[37] *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).

[38] K. Rector et al. "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, 14, 15).

[39] *Resident Evil*. URL: https://game.capcom.com/residentevil/en/ (visited on 2025-02-01) (cit. on p. 27).

[40] *Resonance Audio*. URL: https://resonance-audio.github.io/resonance-audio/ (visited on 2025-01-21) (cit. on pp. 24, 29).

[41] M. Risoud et al. "Sound source localization". In: *European Annals of Otorhinolaryngology, Head and Neck Diseases* 135.4 (2018), pp. 259–264. ISSN: 1879-7296. DOI: https://doi.org/10.1016/j.anorl.2018.04.009. URL: https://www.sciencedirect.com/science/article/pii/S187972961830067X (cit. on p. 8).

[42] M. Risoud et al. "Sound source localization". In: *European annals of otorhinolaryngology, head and neck diseases* 135.4 (2018), pp. 259–264 (cit. on pp. 7–12).

[43] I. N. Rodrigues et al. *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. 17).

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

[45] 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).

[46] 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. 3, 20).

[47] R. M. Schafer. *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster, 1993 (cit. on p. 13).

[48] 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. 20, 27–29).

[49] *Sound Particles*. URL: https://www.soundparticles.com/ (visited on 2025-01-21) (cit. on p. 23).

[50] *Steam Audio*. URL: https://valvesoftware.github.io/steam-audio/index.html (visited on 2025-02-01) (cit. on p. 29).

[51] *Technical Overview of Unity Game Engine*. URL: https://www.pubnub.com/guides/unity/ (visited on 2025-02-01) (cit. on p. 28).

[52] *The future of art spaces: How accessible is the virtual space?* URL: https://www.1854.photography/2021/06/the-future-of-art-spaces-how-accessible-is-the-virtual-space/ (visited on 2024-12-23) (cit. on p. 3).

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

[54] *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).

[55] *Unity Real-Time Development Platform*. URL: https://unity.com/ (visited on 2025-01-21) (cit. on pp. 23, 24, 28, 29).

[56] *Unreal Engine*. URL: https://www.unrealengine.com/en-US (visited on 2025-01-21) (cit. on pp. 23, 24, 28).

[57] P. Vasilakou et al. "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).

[58] 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).

[59] 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).

[60] *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).

[61] *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. 28).

[62] 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. 3, 16, 28, 29).

[63] 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. 7–13).

[64] 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).