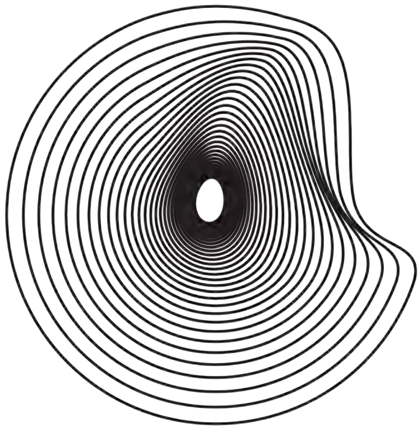
Universität für künstlerische und industrielle Gestaltung Kunstuniversität
Linz
Institut für Medien - Interface Cultures

Masterarbeit zur Erlangung des akademisches Grades Master of Arts



Embracing Sphere,

an Environmental Storytelling with Audio-Tactile Playback System

Mehmet Çolak

Betreut von: Univ. Prof. Dr. Laurent Mignonneau

Datum der Masterprüfung:

Unterschrift des Betreuers/der Betreuerin:

Linz, 2025

Contents

Acknowledgements

I would like to express my sincere gratitude to everyone who supported me throughout my master journey.

First and foremost, I am deeply grateful to my supervisor, Laurent Mignonneau, for his patience and guidance. Especially thankful to Alexander Wöran for his encouragement and insightful feedback.

I would also like to thank everyone at the Interface Cultures department for creating such a supportive and inspiring environment. I learned a lot from everyone, both inside and outside the department.

A special thanks goes to my dear sister, Buse Çolak, for always being there for me, both emotionally and materially. Your support has meant the world to me.

I can not express my gratitude to my mother Zeynep Çolak enough. Her hardworking nature has been my biggest inspiration. Her support and belief in me have been the main source of strength during my studies.

My uncle, Mehmet Derviş, I want to thank him, whose support for my education has been invaluable.

Finally, I owe a debt of love and gratitude to my life companion Ezgi Su Demirci, who has always contributed to me with her interest and valuable knowledge throughout my study, who has never missed her mental and moral support for me for a single day.

I know I could not have done this alone. I am truly thankful for all the help I received along the way.

Thank you all for being part of this journey with me.

Abstract

Conveying a story through the environment is a known technique in game design and theme park train rides[13]. Blood stains under the door, footstep trails on snow and broken chains next to a beware the dog sign; all of these flash scene descriptions are examples of environmental storytelling. A heavy reliance on visual cues, defines the limitations of current methods in environmental storytelling. Utilizing multi-modal stimuli and 3D audio volumes for enhancing environmental storytelling remains niche and open for new perspectives in storytelling and narration through the environment.

This research explores the integration of haptic feedback and acoustic modeling methods. Haptic feedback systems are most commonly used in video games, racing simulations and interactive art installations. Their signals are typically vibrations driven by low-frequency audio, creating pulses on haptic actuators. Complementing tactile experience, Room Impulse Responses (RIR) are utilized as a method of capturing acoustic properties of an enclosed volume/space[54] later to use in reverberation (specifically convolution reverbs) to reconstruct the same acoustic responses while simulating different materials. Combining these distinct modalities, such as tactile pulses and acoustic simulations, offers a novel way to represent environmental snapshots purely through non-visual interfaces.

Building on this potential, a multi-modal stimulation scenario using simultaneous audio signal playback incorporating procedurally generated RIRs and a bass shaker (transducer) will be this article's foundational practical method.

This applied work investigates the current state and potential of such multi-modal experiences in conveying environmental information. Proposed artwork designs, implement and evaluate an interactive audio-tactile system capabilities of procedurally generated environmental snapshots for storytelling purposes.

Chapter 1

Introduction

Using virtual environments as narrative spaces, represents a novel way to tell a story or convey a specific feeling to the audience. Unlike traditional storytelling methods that rely on delivering a story through dialogue, cutscenes or text dumps; environmental storytelling operates more subtly.

To effectively convey environmental storytelling, it's crucial to weave narrative elements seamlessly into the virtual environment that evoke emotions and encourage players to question their surroundings. The design of virtual spaces, like their size, shape, flow, connectivity and materials, can convey narrative meaning. The layout of the virtual environment can guide the player along a specific path, revealing information sequentially or creating feelings of safety, exposure and confusion. The environment becomes a silent narrator, communicating with the audience through carefully crafted, immovable details that enhance immersion and improve the narrative experience[24].

Due to interactive, exploratory and emergent capabilities of video games in storytelling, environmental storytelling is practiced more often, especially after the video game industry's increased popularity[2]. While this practice is increasing, the auditory domain of environmental storytelling remains understudied in both academic literature and design analysis[67]. This thesis focuses specifically on the auditory dimension of environmental storytelling practice.

Movies, video games and interactive art installations often utilize different human sensory fields simultaneously. Understanding and leveraging this multi-modality is crucial when contributing to these mediums as an artistic individual or researcher.

With this multi-sensory understanding, this research will further investigate the interplay/crossplay between auditory and tactile feedback within environmental storytelling. The aim is to explore how audio-tactile experiences can enhance immersion, evoke emotional responses and create richer, more embodied narrative contents for the audience.

1.1 Research Question

The study addresses the following three fundamental questions about auditory environmental storytelling:

1. To what extent does multi-modality play a role in the environmental storytelling?
2. How multi-modal stimulation can be utilized to enhance auditory environmental storytelling?
3. How effectively can an interactive audio-tactile system convey distinct environmental characteristics (space size, material properties etc.) and narrative cues to an audience?

To begin this investigation, the following section reviews foundational concepts of multi-modal interaction, leading into a broader review of existing literature.

1.2 Multi-Modality

Modalities, the plural of modal, is a term primarily related to how media is presented, its form, or its mode. Basically, modalities are modes, a manner, way or method of doing something. Multi-modality refers to the interplay between different modes[53].

The term ”modality” is studied across diverse disciplines such as linguistics, media studies, semiotics and cognitive science. Each examines different aspects of representation, communication, or interaction. In the context of human-computer interaction (HCI), this thesis uses the term modality to specifically refer to sensory channels or the distinct forms of stimuli such as auditory and tactile feedback.

Through this focus, modality can be broadly understood as a distinct sensor, channel, system or manner through which information is represented, perceived, processed or exchanged.

The strategic combination of multiple modalities is often driven by the goal of creating more effective, intuitive and engaging user experiences. Presenting information through various sensory channels can meet different user needs and contexts, creating expressions that can potentially improve narrative and engagement by offering complementary or redundant information.

In the digitized world, different sensory modes have technically become the same at some level of representation and they can be operated by one multi-skilled person, using one interface, one mode of physical manipulation, so that he or she can ask, at every point: ”Shall I express this with sound or music?”, ”Shall I say this visually or verbally?”[53].

Sensory modalities such as auditory cues can effectively convey ambient information, direct attention and shape emotional tone, while the somatosensory system (the system responsible for our conscious awareness of touch, pressure, pain, temperature and body position[71].) offers a direct physical link to digital environments, enhancing feelings of presence and embodiment.

Investigating how these multi-modal experiences can enhance immersion, evoke emotional responses and create richer and more embodied narrative contents, we aim to explore auditory and somatosensory capabilities when an environmental narration is introduced to an audience.

1.3 Environment as a Tool for Narration

"The audience is not a passive observer, but actively investigates the frame both spatially and temporally, responding to the tension created within the narrative space."[63]

The above text, from Interactive Narrative and Transmedia Storytelling, highlights the new role and importance of the audience. Virtual spaces in interactive narratives are more than just digital settings. They are the environments where the story unfolds and where the audience actively participates. Environmental storytelling promotes the virtual environment's role from a basic background to an active narrative agent.

Strategically placed objects in the environment can tell a story on their own. A knocked-over chair, a specific book left open or a series of photographs can hint at past events, reveal aspects of a character's personality or provide clues that help the audience piece together the narrative[24].

In a professionally structured narrative, there are macro and micro storytelling scopes[13]. When it has been done well, this hierarchy creates a depth of narrative design that neither approach could achieve alone, allowing the audience to connect grand macro and micro level events in ways that feel organic and profound[24]. Environmental storytelling has the capability to convey stories that aren't always directly told through dialogue or cutscenes.

The following example images are taken video games, movies and real life They are provided to deepen our understanding of environmental storytelling as a tool for narration;

Figure 1.1: An image from a public road.

On the macro level, 1.1 shows a shortcut made by people walking through the grass instead of using the sidewalk. The red-and-white barrier at the end of the path indicates an attempt to enforce the intended route, but the worn trail on the left shows people kept choosing the faster way.

Figure 1.2: A visual from a video game called Oblivion.

On the micro level, 1.2 tells a humorous story. A man lies exhausted beside a fiery figure on the floor next to a bed with a ”Resist Fire” potion highlighted, hinting maybe he needed magical help to survive an intense, possibly romantic encounter.

Figure 1.3: An image of a building in Istanbul, Turkey.

With another macro level example, 1.3 shows a building from Istanbul, made up of layers from different time periods, each added on top of the previous one. The labels on the right indicates the specific architectural style (starting with the Roman Empire at the bottom and ending with the Republic Era at the top) telling a visual story of Istanbul’s long and complex history. It reflects the city’s rich heritage, shaped by centuries of conflict, change and development, all embodied in a single structure.

Figure 1.4: A visual from a video game called Control.

With another micro level example, 1.4 shows a character stands before a mail room door. The blood spreading under the door serves as an effective environmental narrative device, wordlessly hinting to the player the danger. Signalling that something violent has happened in the space they're about to enter, creating tension and anticipation without explicit exposition.

These examples provided to illustrate how this technique functions across various media and even in our physical world. What makes environmental storytelling special is that it lets each audience member piece together the story in their own way, creating a more personal connection.

With the given level of understanding in environmental storytelling, next section covers my own ideas and proposed installation, a multi-modal playback system directly targeted for environmental storytelling.

1.4 Embracing Sphere and My Own Perception

If I may be a bit more sincere and break the formal language of last couple subchapters, I would like to start by introducing myself.

My name is Mehmet, an audio designer and interactive media artist, born in Bursa, Turkey and currently based in Linz, Austria. My design approach is to create soundscapes and interactive audio experiences with gamification and developing interactions for auditory exploration.

Since I was a child I have been amused by video games and stories that have been told through video games. I studied music technologies in my bachelor's degree, developed my skills around interactive audio and now with the Interface Cultures master's programme, next chapter of my life is going deeper in the field of conjunction of human and technology in the art domain which is highly related to interaction, interaction design and media studies.

Figure 1.5: Video game SSX 3, PlayStation 2 (2003).

I remember the first time I played a game called ”SSX 3” which is a snowboarding game. I can still clearly recall the feeling when jumping off a cliff onto the snow surface and the simultaneous rumble coming from Dualshock 2 (1.6). Feeling of the gritty snow texture when turning a tight corner or imminent hit response when I crashed into a barrier, as a child who experiencing this kind of multi-modal experience for the first time, was really remarkable.

That memory of the Dualshock 2 rumbling in sync with the snowboarding action in SSX 3 really stuck with me. It wasn't just about seeing the snow or hearing the sounds; it was feeling the impact and the texture of that virtual world.

Figure 1.6: Playstation 2 gamepad, Dualshock 2.

Dualshock 2 is the name of main gamepad of the PlayStation 2, which released in 2000 and stayed relevant until nearly 2012-2013. As the physical user interface of this hugely popular and successful console, the Dualshock 2 was also a well engineered interface for gaming.

The vibration motors in the DualShock 2 were quite simple, basically small weights spinning off-center to create a general rumble. It was more of a strong shake or a light buzz[85]. Today, vibrotactile devices/tools are much more advanced, capable of producing a wider range of precise textures and feelings. Modern game controllers, like the Nintendo Switch's ”HD Rumble” or the PlayStation 5's DualSense, take advantage of this. They can create much clearer and more varied tactile effects, like the subtle click of a dial, the tension of a bowstring, or the feeling of raindrops, making virtual interactions feel surprisingly realistic and nuanced.

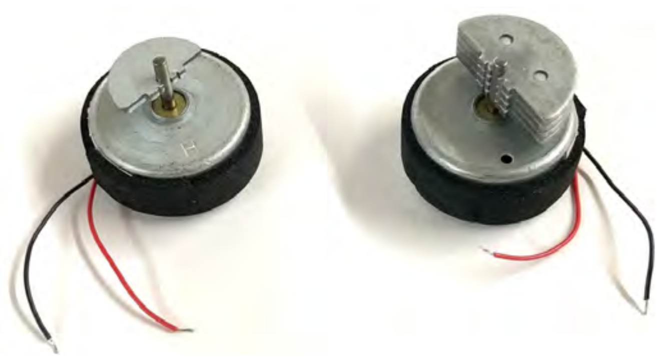


Figure 1.7: Vibration motors of Dualshock 2.

My first hand experience and understanding that combining sound with touch could make a virtual experience feel incredibly real and immersive. With years of other experiences and knowledge about interaction and multi-modal stimuli, made me wonder how much more could be done if these sensory connections were designed intentionally to tell a story or convey specific feelings about an environment.

Embracing Sphere came to life as an idea with the background that I have stated. ”Embracing Sphere”, as a main title that on the one side, poetic yet on the other side a direct analogy of user interaction for a playback system that will be briefly explained later.

The ”playback system” aspect of Embracing Sphere is about creating a personal physical sphere where users can explore these combined audio-tactile sensations. It’s about giving the audience a more embodied way to connect with a narrative, where what they hear is enhanced and expanded by what they can physically sense.

My artwork, therefore, aims to be a practical exploration of the ideas in this thesis. ”Embracing Sphere” is designed to be a tool to investigate how these audio-tactile experiences can shape our perception of a virtual space, its materials and the subtle narrative cues embedded within it. It’s my way of testing how powerful this multi-modality of hearing and touch can be in making environmental storytelling more engaging for the audience.

1.5 Scope and Limitations

Haptic feedback refers to feedback perceived through the sense of touch. The haptic system uses sensory information derived from mechanoreceptors and thermoreceptors embedded in the skin together with mechanoreceptors embedded in muscles, tendons and joints[56].

This research narrows its focus to vibrotactile feedback generated specifically by Voice Coil Actuators (VCAs). Consequently, other forms of haptic interaction, such as force feedback, thermal feedback, or electrovibration, are outside the scope of this study. VCAs are particularly suitable for vibrotactile applications due to their operational principles, which are similar to those of loudspeakers and voice coil actuators[74], allowing for conventional audio recording practices viable on haptic feedback content creation. A further exploration can be found in subchapter 2.3 Haptics and Vibrotactile Perception, where vibrotactile hardwares and human perception capabilities are discussed in detail.

The study is limited by the specific operational range (e.g., frequency, amplitude) of the VCAs employed. Different VCAs possess varying performance characteristics and the chosen model may not represent the full spectrum of possible vibrotactile sensations.

Another limitation is individual differences in vibrotactile sensitivity and perception thresholds among participants. While subchapter 2.3 will explore general human perception, this study does not control for all individual variabilities (age, skin condition, prior haptic experiences etc.).

1.6 Thesis Outline

With the last paragraphs of Chapter 1 - Introduction, this chapter aims to give an overall look for the thesis structure, topics that will covered and brief summary of the rest of chapters.

Chapter 1 - Introduction, established the research context, presented the central research questions, inspirations and personal insights driving to this work. Later defined the scope and limitations of the investigation and provided this outline of the thesis structure.

Chapter 2 - Background and Related Works to the Embracing Sphere, will offer a comprehensive review of existing literature and foundational concepts crucial to this research. This includes defining environmental storytelling and exploring its existing methods across various mediums, before focusing on auditory environmental storytelling. It will then investigate the principles of room acoustics, explaining Room Impulse Responses (RIRs), their measurement, the mathematical concept of convolution and its application in digital audio and sound art. Furthermore, it will provide a detailed review of haptics, different feedback modalities, fundamental aspects of human tactile perception and examples of haptic usage in diverse fields. Finally, the chapter will explore multi-modal interaction, examining related references in literature.

Chapter 3 - System Design and Methodology, will detail the conceptual framework and systemic design of the proposed interactive system. It will cover the methodology for procedural Room Impulse Response (RIR) generation, the design of the audio and haptic playback systems, the interaction design principles and the intended narrative structure built by the system.

Chapter 4 - Implementation, will describe the practical realization of the system. This chapter will cover the specific hardware setup, the software development process and tools utilized, the materialization of the research into an interactive art project and will discuss key challenges encountered during implementation and their solutions.

Chapter 5 - Conclusion, will summarize the entire research endeavor. It will directly address the research questions, acknowledge the limitations of the study.

Chapter 2

Background and Related Works to the Embracing Sphere

This chapter establishes fundamental concepts related with Embracing Sphere by covering in-depth environmental storytelling, acoustics and haptics. The chapter aims to investigate artworks and video games chosen for their relation to the Embracing Sphere on both conceptual and practical levels.

2.1 Environmental Storytelling and Developments in Virtual Environments

Discussions about environmental storytelling as a term in narratology date back only few decades. First defined in 2000 by Don Carson, a former theme park designer for Walt Disney Imagineering, who argues that in themed environments “the story element is infused into the physical space a guest walks or rides through”[13]. During his work in theme park train rides or video games, his objective is to tell a story through the experience of traveling through a real or imagined physical space[16].

These discussions later developed into a game design discourse as the concept of ”story versus play”, ”ludology vs narratology[66]” within transmedia storytelling[45]. Jenkins argues that the story becomes richer and more complex, as the audience is given more opportunities to engage with the narrative.

Although the transmedia storytelling directs something else (a process where integral elements of a fiction get dispersed systematically across multiple delivery channels[46]), a discussion of the narrative potential of video games supported attempts to create narrative spaces in virtual environments[13].

Related Work: Journey

"Like a religious ritual of passage, it is not the spiritual narrative's plot, but rather the poignant symmetry between its metaphorical meaning. The embodied experience of performing the movements it channels, that makes this narrative effective. Journey makes zero use of language and relies entirely on the experience of movement to tell its story[77][60]."

Figure 2.1: A visual from the game called Journey released in 2012 by Thatgamecompany and Santa Monica Studio.

Environmental storytelling in video games is done by staging the game world so that the arrangement of objects, scenery and audio cues naturally conveys the story to the player[83].

Journey is a video game has critic focus on exploration and a great, well awarded (Journey won several "Game of the Year" awards from different organizations) example for environmental storytelling. The game accomplishes this narrative success mostly by not relying on any use of linguistics or semantics.

In Journey, the player controls a figure starting in a vast desert, traveling towards a mountain in the distance in a multiplayer environment, which means you, as a player, can interact with other players on the same journey. The challenge is that the players cannot communicate via speech or text and cannot see each other's names until after the game's credits.

Players have basic navigational controls like walking, jumping, sliding on the dunes and the ability to emit a wordless shout to another. The length and volume of the shout depend on how long the button is pressed. [1]

[1] Journey shout example: `https://youtube.com/clip/UgkxMXBXc4aZmHuOL2f3PUoEVQ57Og5Suyks/`

Through the player's path in Journey, one distinct element always catches the eye. The big shining mountain peak on the horizon, pictures an unspoken, ultimate goal or a direction for our journey. Each player is trying to reach the peak either by helping each other or going this path individually.

Figure 2.2: The mountain in the video game, Journey.

The subtractive design of Journey's game environment forces the player to focus on the environment all the time. Directing to consume the story through cryptic glyphs, symbols and figures carved into the walls, artistically placed in the game environment.

By collecting these symbols, players gain more movement ability to explore deeper in the game. This mechanic is also shown in a diegetic way with a scarf wrapped around the player character's head. The scarf is giving information about your energy left to jump and fly around, like the fuel in your tank or stamina left. The scarf grows as you progress through the game by picking up collectible symbols.

Figure 2.3: The player's scarf in the Journey.

In Journey, the narrative aspect of the non-linguistic communication and the movement through the designed space, generate the story. The player reconstructs the story by interpreting different objects in ruins, interacting with the others and initiating events embedded in the game environment.

Video game in that sense has one characteristic feature, which is their spatiotemporality. In comparison to other multimedia mediums like cinema, which is purely temporal sequences, we can view video game narrative as a blend of the temporal and the spatial[13].

2.1.1 Existing Environmental Storytelling Methods in Different Mediums

Environmental storytelling is not entirely specific to video games. First in theaters, then later adopted in screenplay, there is a French term called ”mise-en-scene”, literally translated as ”putting into the scene”. Mise-en-scene describes the arrangement of scenery, props, lighting and other visual elements to support the story. Filmmakers and production designers use settings to provide backstory and mood. For example, a character’s cluttered apartment, lit by a flickering lamp, can imply their personality and situation before they even speak. In cinematic terms, every object in the frame is there by choice to serve the narrative.[76]

Details like wall posters, broken objects or photographs can signal past events. A close-up of a train ticket on a nightstand might hint where a character planned to go. In film theory, such embedded clues function like visual foreshadowing. Viewers can pick up these details either consciously or subconsciously. Production designers and directors use every element of mise-en-scene to make the environment feel lived-in and story-rich.

For the comparison between cinema and video games, we can broadly categorize video game narrative broadly into 2 categories: ”embedded narrative”, which follows pre-authored contents as temporal narrative sequences that still exist without audience action and ”emergent narrative” which is directly linked with audiences meaningful action, exploration and interaction with the virtual environment[13]. Cinema, doesn’t have emergent narrative due to its fixed media characteristics. According to Natalia A. Bracikowska, environmental storytelling exists in a liminal space between embedded and emergent narrative[13] and this aspect of environmental storytelling opens a pretty viable path to convey multi-branched story bits for the audience.

Across interactive media, game design and film theory, environmental storytelling is a shared strategy to embed narrative in the space itself. In all domains, meaning is conveyed indirectly. Games and interactive media emphasize the user’s role in uncovering that meaning players must explore and investigate to “do the work” of interpretation. In contrast, films present a fully authored scene but still rely on designed space to impart story context. There is a common goal to see, which is to make every detail of the world serve the story.

2.1.2 Auditory Environmental Storytelling

"Sound is an integral part of every performative and aesthetic experience with an artifact. Yet, in design disciplines, sound has been a neglected medium, with designers rarely aware of the extent to which sound can change the overall user experience."[29]

According to Stefania Serafin, humans are sensitive to sounds arriving from anywhere within the environment, whereas the visual field is limited to the frontal hemisphere, with good resolution limited specifically to the foveal region. Therefore, while the spatial resolution of the auditory modality is cruder, it can serve as a cue to events occurring outside the visual field-of-view[33]. Therefore, effective auditory environmental storytelling relies on a sophisticated understanding of how different sound elements function individually and collectively to shape the user experience.

The concept of "presence" and "immersion" highly relates to this subject from the perspective of auditory domain's characteristics on fully spherical perception and cognition.

Immersion is a studied concept in multiple fields such as film, video games and music. It is a subject that can have different meanings depending on the context of the field of study[33]. Immersion is a metaphorical term that derives from the physical experience of being surrounded. In Janet H. Murray's words, "being submerged in water"[66].

Presence, on the other hand, is a term that is used in a much broader sense. According to Lombard and Ditton, presence is described as the feeling of "being there" in a mediated environment, even though the experience is happening through a screen or device. Presence can occur in different forms, such as feeling physically in a place shown on screen, socially connected to others through media, or involved in an environment that reacts to one's actions[59].

By understanding how we naturally interact with the world, how we interpret information provided by sensory stimulations, we can apply this understanding to enhance and elevate the immersion and presence in our intended mediums.

Therefore, auditory stimulation, by its nature, envelops a lot of human sensory aspects. It is explicitly useful for its spatiotemporal narrative capabilities. With a related work investigation that explores a video game called Return of the Obra Dinn we can have a deeper understanding of auditory environmental storytelling.

Related Work: Return of the Obra Dinn

Return of the Obra Dinn, is an adventure-puzzle video game developed by Lucas Pope in 2018. In the backstory of the game, The Obra Dinn, a merchant ship missing for five years, has reappeared off the coast of England with no known surviving crew or passengers. The player's task is to determine every crew member's fate, including their names, how they met their fate, who or what killed them and if anybody is alive, where are they?

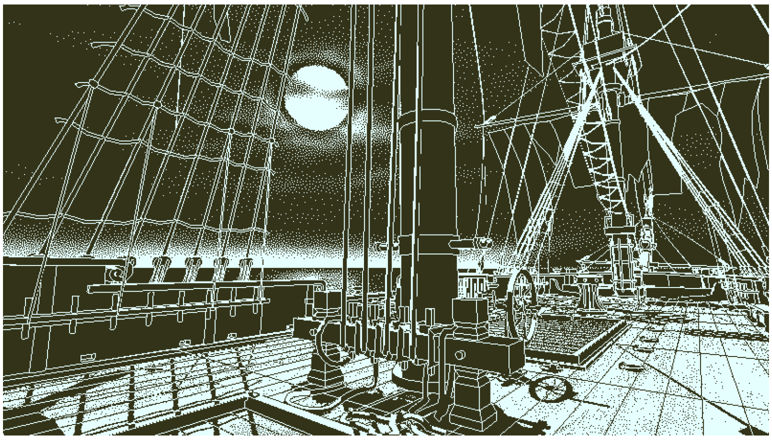


Figure 2.4: A visual of the ship deck from the game called Return of the Obra Dinn released in 2018 by Lucas Pope.

Players can hop onto the Obra Dinn and navigate on the main deck. The main mechanic is portrayed with a Latin phrase, "Memento Mortem", a mystical pocket watch that allows the investigator to witness the final moments of any corpse discovered aboard the ship.

Figure 2.5: A visual of the pocketwatch in Return of the Obra Dinn.

When a player approaches a remaining of a crew member the pocket watch can be activated and it triggers a sequence that includes 2 parts: first, an audio snippet right before the event happened, secondly a static, frozen explorable visual tableau of the event that results the death scene.

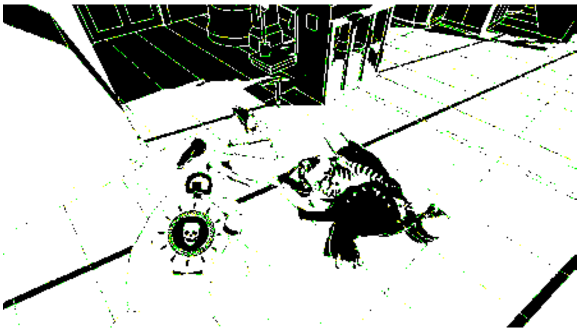


Figure 2.6: A visual of the of a remaining of crewmember in Return of the Obra Dinn.

The crew member shown in the 2.6 is our first case to solve in this adventure-puzzle game. As we interact with the body remain, a musical cue transports us to the past and we hear an audio snippet [2] which transcription is like that:

- *"Captain! Open the door..."*
- *"Kick it in!"*
- *"...lest we break it down... ...and take more than those shells."*
- *"You bastards may take... ...exactly what I give you.[DOOR OPENING AND GUN SHOT]"*

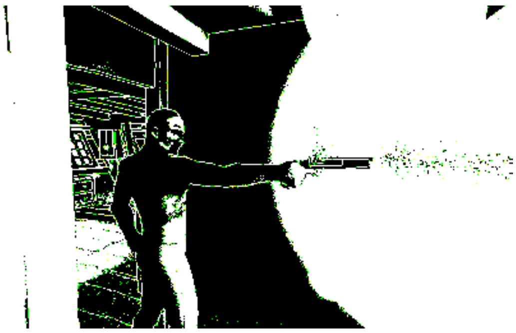


Figure 2.7: A visual the static scene of Return of the Obra Dinn.

At the point 2.7, we have enough sensory cues to solve at least 1 crew member's identity, which is the Captain himself and we have enough hints to define how this crew member died. Because according to the auditory cues, one crew member calls "Captain!" and knocks the door in anger. The other crew member is suggesting to

[2]Return of the Obra Dinn, first audio snippet: `https://youtu.be/UXC6Sjsedcg`

kicking the door and forcefully getting into the room. The only person we hear behind the door opens it and shoots one of the rebellious crew members immediately.

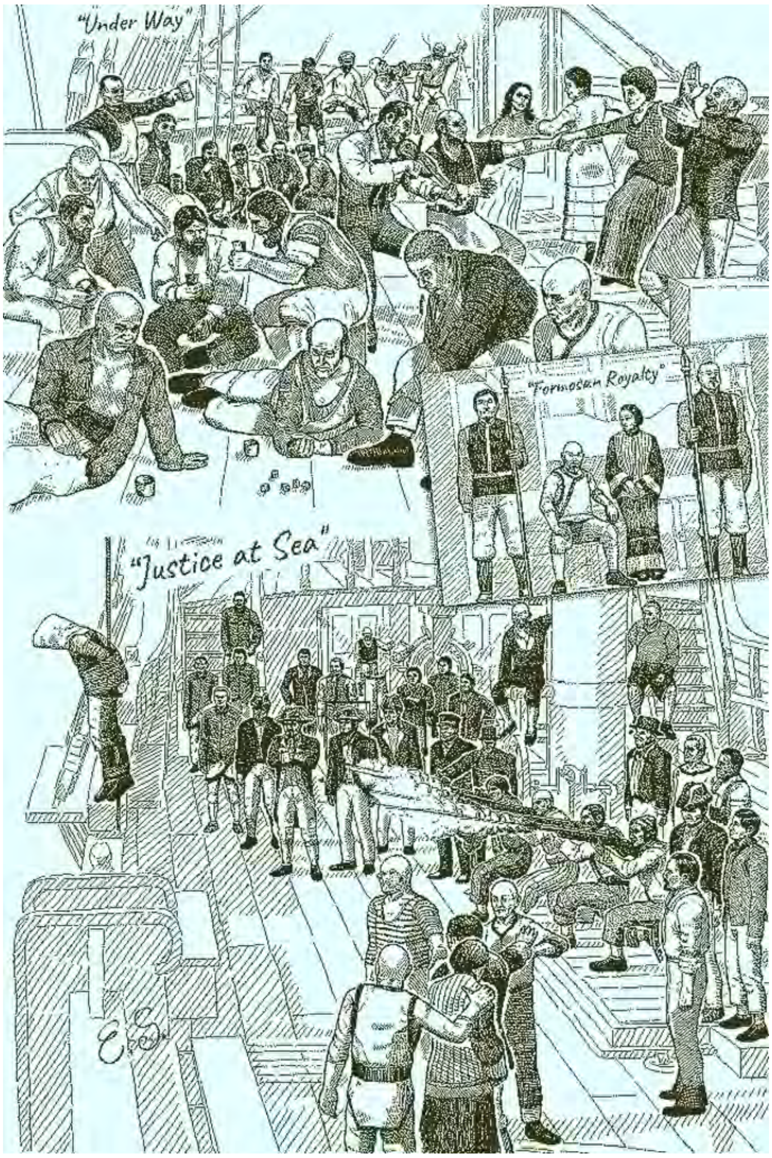


Figure 2.8: An illustration of all crew members of the Obra Dinn.

While the visual scene is frozen, the audio snippet captures the ambient soundscape of that specific moment and location. This ambient layer logically contributes to differentiating scenes and grounding the abstracted visual tableaus in a more tangible sense of place. It is in my perspective, a well done auditory environmental storytelling.

My experience with this video game was a great inspiration for Embracing Sphere after all. It showed me that the possibility of environmental storytelling is indeed strong and it has benefits for spatiotemporal story cases.

After covering the necessary perspective from narratology and video game studies, sound and perception are going to be investigated in more detail, specifically acoustics and psychoacoustics subjects within the next subchapter.

2.2 Space and Acoustics in Artistic/Scientific Perspective

"Sound is something most people take for granted. Our environment is full of noises, which we have been exposed to from before birth. What is sound, how does it propagate and how can it be quantified?"[41]

The purpose of this subchapter is to introduce fundamental knowledge ground for room acoustics and describe the usage of room impulse responses in the Embracing Sphere context.

2.2.1 Definitions for Room Acoustics

Sound is simply a mechanical disturbance of the medium. Medium in that context may be air, solid, liquid or other gaseous matters. Depending on the medium's state, the sound can be propagated and while this propagation occurs it interacts with physical objects and other sound waves. In room acoustics, the interactions of the sound with the medium basically can be listed as refraction, absorption, reflection and interference. Psychoacoustics is the study of how humans perceive sound after all these interactions with the medium[41].

The sound heard, is the result of all these complex physical interactions in the place where our ears are located.

When sound moves through a room, its behavior is shaped by how it interacts with surfaces, mainly through absorption and reflection. Sound reflection occurs when sound waves hit a hard boundary and bounce back into space. In contrast, sound absorption is when a material takes in the sound's energy, which is converted into small amounts of heat through internal friction, reducing the amount of sound that reflects[41]. For example, a wooden surface absorbs more sound than a rough concrete one. These interactions, along with others like refraction and interference, collectively determine the acoustic character of a room.

These interactions do not occur in isolation. They collectively define the complex behavior of sound waves and their propagation within the room. Each time a sound interacts with a surface in a room, it loses some of its energy due to absorption and reflection. The time that it takes for a sound to gradually fade out in a room is called the reverberation time.

Reverberation time is an important aspect of sound behavior in a room. Mentioning different absorption coefficient values in different materials and frequencies shapes the perception of the room. If the sound dies away very quickly, we perceive the room as being "dead" or if the sound dies away very slowly, we perceive the room as being "live". To calculate reverberation times, there is a simple formula known as the "Sabine formula", named after its developer, Wallace Clement Sabine[41].

$$RT_{60} = \frac{0.161 \cdot V}{A}$$

- RT60: This is the reverberation time in seconds. It's defined as the time it takes for the sound pressure level in a room to decrease by 60 decibels (dB) after the sound source has stopped[54].
- 0.161: This is a constant. Its units are seconds per meter (s/m). This constant is derived empirically and is based on the speed of sound in air at a typical room temperature.
- V: This represents the volume of the room in cubic meters. Calculated by multiplying the length, width and height of the room.
- A: This is the total sound absorption of the room in Sabins. It's calculated by summing the absorption of all surfaces in the room. The absorption of each surface is found by multiplying its surface area in square meters by its sound absorption coefficient at a specific frequency.

According to the Sabine formula, reverberation time depends on the volume, surface area and the average absorption coefficient in the room. However, the absorption coefficients of real materials are not constant with frequency. This difference in absorption strength at different frequencies changes the timbre of the room as the sound in the room decays away. Apart from being useful, the Sabine formula has assumptions for the speed of sound and static response of the reflective materials in the room. To more accurately measure reverberation time, another method called room impulse response capturing was introduced in 1964 by M. R. Schroeder [75].

This method uses tone bursts (or filtered pistol shots) to excite the enclosure (room). The captured smooth decay curves resulting from the new method improve the accuracy of reverberation time measurements and facilitate the detection of non-exponential decays[75].

These impulse response recordings can be used later to reconstruct a virtual environment with the same reverberation curves as the captured room with a mathematical operation called convolution. An anechoic (no reflections) sound is convolved with an RIR and this mathematical process applies the room's acoustic snapshot to the sound. The convolution effectively embeds the reflections and reverberation captured in the RIR into the original sound.

2.2.2 Room Impulse Response Measurement Methods

"Between stimulus and response there is a space. In that space is our power to choose our response. In our response lies our growth and our freedom."[33]

The quote above is highly controversial because it has been attributed to an Austrian neurologist and psychologist, Viktor Frankl, by Stephen R. Covey in his book called The Seven Habits of Highly Effective People, but the quote that was attributed to Frankl's book, Man's Search for Meaning, does not include the quote. Although I wanted to share this quote as a poetic mood setter for this chapter.

From an artistic perspective, Embracing Sphere refers to this metaphorical concept of the relation between stimulus and response of space. Either directly or metaphorically it is a powerful representation in spatial perception. Embracing Sphere is about creating environments/spaces that are capable of telling stories. These spaces should contain both stimulus and response. Although in Embracing Sphere, the audience doesn't have the power of choice, space has that power because I view the space as an actor that has a lot to tell. The response of the space, grows and establishes its freedom.

Acoustics, mathematics and methodologies around it can be covered so deeply in detail, but in the end, all of these descriptions and formulas exist in my research to support my artwork to understand the background and design choices of the Embracing Sphere. Therefore, RIRs are important in my artwork to give an acoustic and spatial context to the audience. It is useful to render the depth and the aural character of the environment into audio content.

In previous sections, we covered room impulse response description at the surface level. It is a captured audio file that contains acoustic space characteristics such as frequency changes, reverberation decay and length.

There are several practical ways to capture the RIR of a room. The easiest one is popping a balloon in the desired room and recording the balloon pop. This method is practical but not accurate enough and it is not easy to recreate because every balloon pop sound parameters may differ in frequency-wise[26]. The more accurate method is the sine sweep technique [3] , which includes 1 reference sine wave sweep audio (5-10 seconds long) that starts from 20Hz and sweeps every frequency until 20kHz[87]. In this method, the room is acoustically excited by this sine wave and with a microphone, the room response is recorded, then processed with a mathematical process called deconvolution that extracts RIR by subtracting the reference sine sweep from the reverberant recording.

[3] A sine sweep IR capture example: `https://youtu.be/1egKAtC16e8?feature=shared`

Figure 2.9: A visual from a recording session for capturing room impulse response in an old house in Balat, Istanbul.

The visual in 2.9 is a recording session in 2021 when I was working as an audio designer in a company called Vadi Sound. In that frame we were capturing RIR with balloon popping in an old house in Balat, Istanbul. The house specifically interesting for capturing RIR because it was one of the few historical wooden houses remain in the Istanbul from early Turkish Republic times. In that situation we hadn't got any high quality full range loudspeaker to excite the room so we choose to pop a balloon at several places in the room to capture as many different positions RIR.

Figure 2.10: A visual of a RIR capturing session in a studio using sine sweep method. Retrieved from Audioease website `https://www.audioease.com/altiverb/browse.php`

An RIR capturing session that using the sine sweep method can be seen in the figure 2.10. The specific setup uses 2 omnidirectional microphone positioned in the A/B stereo miking technique (a well known stereo miking technique in ambience recordings.)[20][42] and 1 binaural microphone to capture the studio room impulse response.

Figure 2.11: An illustration of raytracing of a sound in a close environment. Drawn in AMROC website `https://amcoustics.com/tools/amray`

Either method works in basic principle shown in the figure 2.11. A source that exciting the room and a capturing point that captures the room's acoustic character.

RIRs are artistically useful for virtual aural architecture. In real life architecture uses volumes to evoke intended emotion or paradigms. Some spaces can evoke feeling about privacy or loneliness, others can invite for social cohesion or some space can establish an hierarchy between different roles in the society[11]. Artistically I see enclosed spaces and its response as an actor in my story. Same dialog or sonic events can convey many different meanings and feelings according to environments relation with the stimulation.

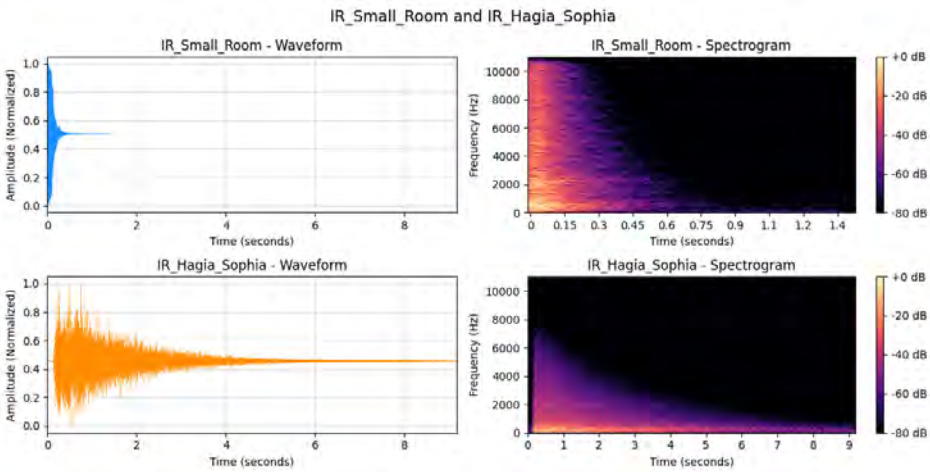


Figure 2.12: Plotting of RIR files of 2 distinct spaces, a small room and The Hagia Sophia. Plotted with matplotlib and librosa in python.

To explain the frequency difference and reverberation time affect in any sonic event, the 2 completely different character RIR waveforms and spectrogram graphics are shown in the figure 2.12. As we can see in the left column in the graph, these 2 RIR files have quite different lengths and densities.

The first row with the blue waveform is from a small room RIR shown above, with a picture of me holding a balloon. The length of this small room RIR is around 500ms, which indicates a small volume and highly absorbent environment with carpets on the floor, many cushions and curtains.

The second row with the orange waveform is from architecturally and historically famous building Hagia Sofia in Istanbul, built in the 6th century. The Byzantine church Hagia Sophia transformed into a mosque after conquest of Constantinople (modern day Istanbul) by Ottoman Empire in 1453 and functioned as a mosque until it transformed into a secular museum with the political and social reforms made in 1935 under Mustafa Kemal Atatürk then reverted back into a mosque in 2020 by Turkish government to support their religious ideologies and symbolic political endeavors[6]. This iconic wonder has stood against time and many diverse cultural influences.

To study the acoustics of old Byzantine churches and mosques built by the famous architect Sinan (known as Mimar Sinan), the CAHRISMA project (Conservation of the Acoustical Heritage by the Revival and Identification of the Sinan's Mosques Acoustics) started in 2003[69]. In this project, the sine sweep technique is used, an omnidirectional loudspeaker emits a sinusoidal sweep signal in the Hagia Sophia and the response of the environment is simultaneously recorded with a microphone. The sweep response is then deconvolved with the reference sweep signal[79].

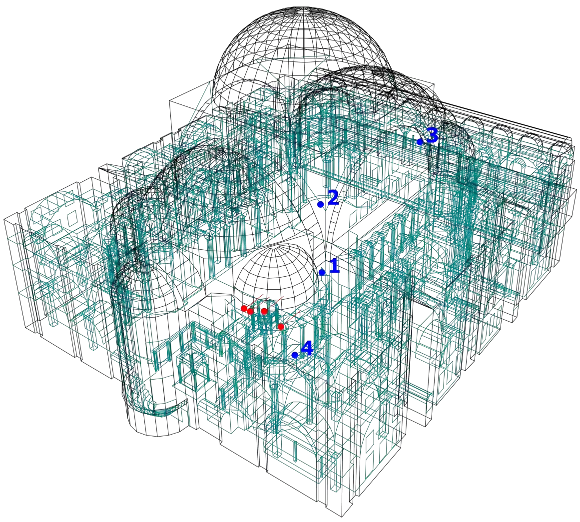


Figure 2.13: Wireframe visualization of Hagia Sophia from ODEON software. Retrieved from ODEON web-blog `https://odeon.dk/the-church-hagia-sophia`

The RIR of Hagia Sophia then the analyzed and full acoustic model built in an acoustic software called ODEON[79]. This model can simulate auralization of any anechoic sound inside Hagia Sophia in different positions and the aural difference between Hagia Sophia as a mosque (with carpet floor and wooden panels with Arabic inscription) and Hagia Sophia as a church (mainly marble floor, stone walls)[7].

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz
Painted Wood	0.04	0.04	0.07	0.06	0.06
Carpet	0.01	0.02	0.06	0.15	0.25
Marble	0.01	0.01	0.01	0.01	0.02

Table 2.1: Absorption coefficients for painted wood and marble at different frequencies[82].

As shown in the figure 2.12 Hagia Sophia's reverberation time is much longer than our small room, around 6-8 seconds. The frequency response of both of the examples also shown in the figure 2.12. The frequency response also depends on volume, absorption coefficient and affects the heard sound dramatically. High frequencies are easier to absorbed by air and surfaces[54] in the room and this can be seen in the figure 2.12 as a colorful slope in the spectrogram panels in the right column.

In artistic view, ability to snapshot an environment and using for creating virtual environments[1] highly interesting and useful for my perspective. Using RIRs to develop such virtual environments require a specific signal processing called "convolution". The next section will cover the practical mathematical description of convolution and its capabilities in digital audio.

2.2.3 Convolution in Math and Digital Audio

Convolution is a mathematical operation that combines two functions to produce a third function. This new function expresses how the shape of one function is modified by the other. In simple terms, convolution tells us how one signal changes when it passes through a system described by another signal. The convolution algorithm is often interpreted as a filter, where the kernel filters the feature map for certain information[21].

We can say that convolution is fancy multiplication[10]. It is important in physics and mathematics as it defines a bridge between the spatial and time domains and the frequency domain through the convolution theorem. Convolution is essentially used in computer graphics, digital signal processing and lately in machine learning algorithms.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

The above formula is the formal definition of the convolution operation. Instead of starting a calculus lecture, we can present a metaphoric example to deepen our understanding:

Let's say you are running a restaurant. In this restaurant, you have a fixed menu and your kitchen uses 2 eggs for one meal. In Monday rush hour, 10 meals are ordered in the first hour, 11 in the second, and 12 in the third. How many eggs did you use? It's simple multiplication.

$$(2 \cdot 10) + (2 \cdot 11) + (2 \cdot 12) = 66$$

The answer is 66, but on Tuesday, your chef added a dessert that requires an egg to make. Desserts are getting served after 1 hour of each meal. How many eggs do you use for each hour? This is now a complex problem because the amount overlaps after the first hour.

First hour is simple, you prepare initially 10 meals with 2 eggs each.

$$2 \cdot 10 = 20$$

Then the next hour you have to prepare 11 meals, but also 10 desserts for the visitors from the first hour.

$$(2 \cdot 11) + (1 \cdot 10) = 32$$

The third hour after your last visitors came, you have to prepare 12 meals and an additional 11 desserts for the second hour visitors.

$$(2 \cdot 12) + (1 \cdot 11) = 35$$

And you work overtime for the last visitors and serving 12 desserts.

$$1 \cdot 12 = 12$$

To summarise all the details,

- The Input (Orders): [10, 11, 12]
- The Plan (First meal, next hour dessert): [2, 1]
- The Result (Total eggs used per hour): [20, 32, 35, 12]

This calculation is a high-level example of a convolution operation. In digital audio, with the lists that have at least 44100 samples[4] per second, many more convolution operations are needed to convolve an input audio (input) and an impulse response (kernel).

[4]The Nyquist-Shannon sampling theorem states that a continuous signal can be accurately reconstructed from its discrete samples when the sampling frequency exceeds twice the maximum frequency present in the original signal[9].

Convolution is a key tool for processing sounds. It is used to apply the characteristics of one sound (such as the acoustics of a room or the response of a loudspeaker) to another sound (like a musical recording or a voice).

Convolution is also used in digital filters, such as equalizers and effects and in spatial audio to simulate how sound arrives at the ears from different directions. For example, by convolving audio with a head-related transfer function (HRTF), we can make sounds appear to come from specific locations in three-dimensional space[70].

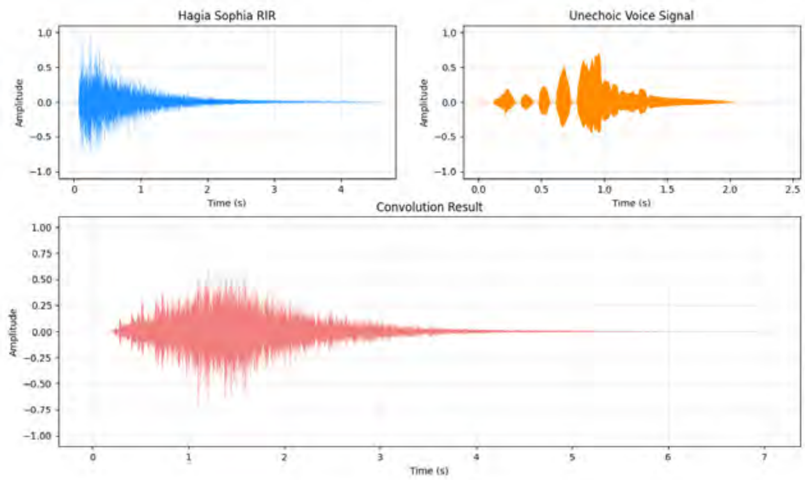


Figure 2.14: Waveform plotting of convolved signal of The Hagia Sophia RIR and anechoic singing voice. Plotted with matplotlib and librosa in python.

The visual explanation of convolution is shown in the figure 2.14, where an anechoic voice signal and RIR from Hagia Sophia signal are processed with convolution and the result is indicated in the second row. The aural effect of convolution makes us perceive that the singing voice is happening in a big reverberant church instead of an anechoic chamber.

Embracing Sphere utilizes this process to procedurally generate virtual acoustic environments and embed existing narrative into various virtual spaces. Procedurally generating new snapshots of imaginary spaces that serve as an environment that tells a story. Procedural generation and the system behind it will be covered in detail in Chapter 3 and Chapter 4.

Deriving from the technical perspective of acoustics and artistic approach with these acoustic technologies, in the next section, the usage of the room's acoustics as both medium and subject will be covered with a related work investigation process-based art called "I am sitting in a room" from renown sound artist Alvin Lucier.

2.2.4 Room Acoustics in Sound Art

"I am sitting in a room different from the one you are in now.
I am recording the sound of my speaking voice and I am going to play it back into the room again and again until the resonant frequencies of the room reinforce themselves so that any semblance of my speech, with perhaps the exception of rhythm, is destroyed.
What you will hear, then, are the natural resonant frequencies of the room articulated by speech. I regard this activity not so much as a demonstration of a physical fact, but more as a way to smooth out any irregularities my speech might have."
- Alvin Lucier[61]

Related Work: I am sitting in a room

The emphasized text is from a sound art piece "I am Sitting in a Room" [5] . This work is widely recognized in the experimental music and sound art scene as one of the most important pieces in the history of minimalist sound art[28].

Figure 2.15: Cover visual of the artwork I am sitting in a room from Alvin Lucier.

[5]Recording of I am sitting in a room: `https://alvinlucierlovely.bandcamp.com/album/i-am-sitting-in-a-room`

Figure 2.16: Alvin Lucier recording I Am Sitting in a Room at The Museum of Modern Art, New York, on Saturday, December 20, 2014. Retrieved from `https://www.moma.org/explore/inside_out/2015/01/20/collecting-alvin-luciers-i-am-sitting-in-a-room/`

The artwork follows a simple iterative process. Lucier records himself reading the text at the start of this chapter, explaining exactly what is going to happen in the performance. In the iterative process, Lucier re-records this recording by playing back in a room with a loudspeaker and a microphone. This process repeats over and over again, with each new iteration containing the previous recording and the acoustic interaction of the room[61].

The room acts like a filter, emphasizing certain frequencies that match its natural resonances (modal frequencies of the room[54]) while reducing others. With each iteration, Lucier's words become more blurred and the room's acoustic characteristics become the leading aspect of the performance. Eventually, Lucier's speech transforms into pure tones that reveal the acoustic signature of the space itself. The transformation happens gradually in 45 minutes of tape[48].

This approach turns the space into an active participant in the composition, rather than an empty container for sound. It highlights our environment's physical role in our artistic output. This concept has inspired many artist to consider the acoustic properties of spaces as a factor in their works[28].

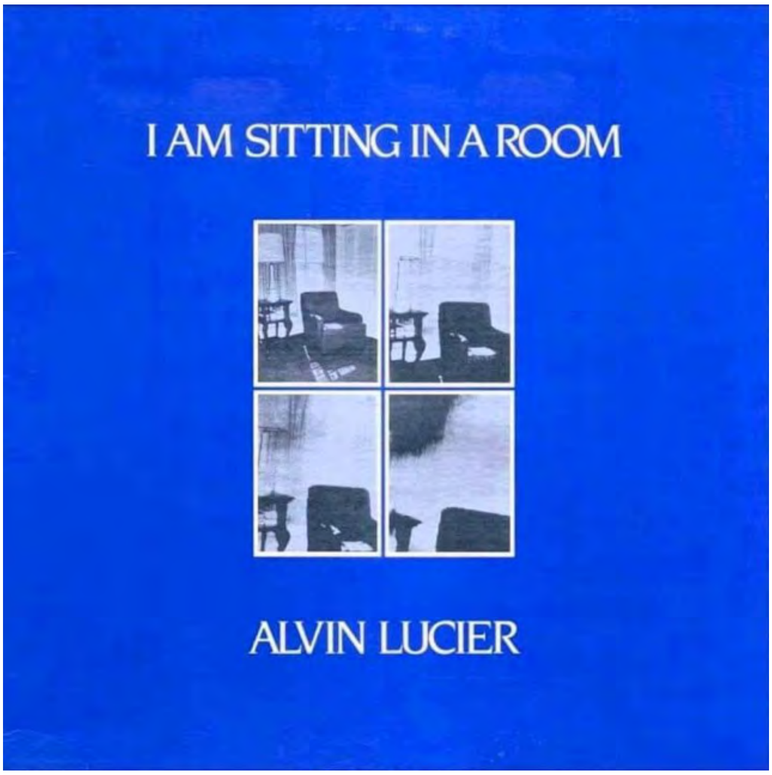


Figure 2.17: Minimalist cover visual of the artwork I am sitting in a room from Alvin Lucier.

My inspiration, which I derived from I am sitting in a room, was one of the driving forces in my research about auditory environmental storytelling. After I found out about this artwork and took my time to digest the composition, I figured out that our surrounding environments have their own character and affect on the information we constantly perceive. This concept of environment as a narrator/actor has been utilized in cinema and video games but developing an environmental storytelling without any visuals or semantic structures intrigued me most. Thus, Embracing Sphere promotes the space into the main narrator role, through auditory or tactile sensory modes.

In summary, the acoustic perception is important in the Embracing Sphere. The next sections are going to cover haptics and human vibrotactile perception, with a focus on Embracing Sphere and why haptics are an efficient sensory mode for perception and cognition.

2.3 Haptics and Vibrotactile Perception

Spatio-temporal narration can be done with many different sensory modalities, but auditory and somatic senses fundamentally need a temporal ground to be perceived and have capabilities for conveying spatial cues[56]. The temporal ground is basically the time required for our perceptual and cognitive systems to fully evaluate a stimulus[56]. In the auditory medium, we distinguish relative pitch, volume, amplitude and distance mostly through this temporal information and with this information, we interpret the environment we are stimulated by; thus, a spatial meaning is created.[57]

In haptics (somatic system), the human perception behavior is similar in many ways and the benefits of somatic stimulus on interpreting an environment is quite the same as the benefits of the auditory stimuli[56].

Both of them can be studied and examined with the same scientific subject ”waves” because fundamentally, both of them are transferred via vibrations[47]. Some of the auditory concepts can be perceived with the somatic system simultaneously, such as rhythm perception[17][8].

For Embracing Sphere, I intentionally chose the auditory and haptic modalities to use the ability to create contrasts and harmonies within the same scenery. From my perspective, haptic stimulus is mostly interpreted as intimate and close in distance. In the meantime, auditory perception is the best way to convey the distance and localization, even better than vision because it's not constrained by our peripheral vision.

If you imagine my craft as a painter, I can draw depth into my artwork much more efficiently with audio-tactile stimulus.

To use this wide palette, this section, Haptics and Vibrotactile Perception, is going to investigate human haptic capabilities and benefits. Later, we are going to use this knowledge to create more detailed virtual environments and the ultimate objective is to convey better environmental storytelling.

2.3.1 Overview of Haptics

Within the human body, the somatic system can be subdivided into three elements: kinesthetic, visceral and cutaneous. Kinesthetic sensation uses signals from proprioceptors in the joints, muscles and tendons to provide feedback to the brain on the position and forces within segments. Similarly, visceral sensation uses receptors in the abdomen. Cutaneous sensation consists of the combined response of four types of nerve endings in the skin[47]. The haptic system uses sensory information derived from mechanoreceptors embedded in the skin, muscles, tendons and joints[56].

Haptic interaction can be stimulated by different devices, such as thermal feedback and electro-vibration which are outside the scope of this study. Therefore, a vibrotactile device was chosen for this research. These vibrotactile devices are similar to those of loudspeakers and voice coil actuators[74], allowing for conventional audio recording practices to be viable on haptic feedback content creation.

Figure 2.18: Voice coil actuator, vibrotactile device, Dayton Audio BST-1.

The sensory system is a network that enables your body to receive information from the environment and its own internal state, converting stimuli into signals for the brain to process. The human sensory system doesn't just process this sensory information as a single stream; it organizes it to answer fundamental questions about the environment[73]. Research in sensory neuroscience suggests a fundamental distinction in how the brain processes sensory information framed as "what" an object (disturbance) is versus "where" it is located.

How do you feel the difference between rough stone, resonant wood, or soft earth? This relates directly to identifying the "what" of the surrounding environment. The "where" pathway provides spatial information, helping us to understand the location of a stimulus in relation to our body and within the environment. The location of a distant explosion felt through the floor is related to identifying "where" information.

As we are exploring environmental storytelling through audio-tactile stimulation, we can extend this framework with a third component: "how". This component can include "cause and effect" relations within our sensory perception and cognition. Where "what" and "where" components answer material and spatial questions, "how" components can answer temporal questions derived from the first two components. These questions will focus on a more interactive concept of "event" rather than static stimulation characteristics.

This "what, where and how" taxonomy provides a conceptual tool for environmental storytelling, thus embedding temporal relations into the environment, moving beyond simple rumbles to convey specific information about an environment's materials, spatial layouts and past/ongoing events.

Figure 2.19: A visual shows road rumble strips.

As shown in the 2.19, rumble strips are designed to alert drivers by creating vibrations and noise when a vehicle leaves its intended lane or crosses the edge of the road. Stimulation from a rumble strip on the side of the road, not just indicating physical position or road surface information, is an immediate warning.

The sensation isn't a random patch of bad road; it's a deliberately engineered, rhythmic pattern. This pattern connects the what (the ribbed texture) and the where (the edge of the lane) to create a temporal meaning: ”You are currently in the process of making a mistake.” The ”how” pathway interprets this sequence as a cause-and-effect event, because you are drifting, you are feeling this vibration. In a structured narrative context, this type of haptic stimulation can be utilized for environmental storytelling.

2.3.2 Human Tactile Perception

This section will cover human response to haptic stimulation, detection thresholds of vibrotactile stimulation in comparison with auditory perception.

- What are the detection thresholds of vibrotactile stimulation?
- How do vibrations affect the material texture feeling of an object?
- How do whole body vibrations affect our feelings?
- How vibration amplitude, frequency and frequency pitch intervals affect haptic cognition?
- Which body parts or positions are most effective in haptic stimulation?
- How is audio-tactile stimulation affecting our localization capabilities?

These questions were introduced in the context of Embracing Sphere to basically have a certain knowledge ground while crafting/designing content for Embracing Sphere. As the ultimate goal is to convey environmental storytelling through an audio-tactile interface, it is crucial to explore and study haptics in detail.

The first answer defines tangible human limits and scales the most effective playground for vibrotactile experience. The detection threshold of vibrotactile stimuli varies with frequency, transmission position and measurement conditions. Human response to vibration is a multidisciplinary topic that involves many sciences such as biology, psychology, physics and biomechanics[47].

People are primarily exposed to vibration, either localized or vibration that affects the whole body[47]. We experience our surroundings through these 2 different vibration ways. The question, "What is the required magnitude of a vibration for it to be perceived?" has to be restructured as "What is the required magnitude of a vibration in each frequency domain for it to be perceived by hand and whole body?". This restructuring embeds more accurate investigations in human response to vibration, assembling more useful data for Embracing Sphere. In many researches[47][17][44][68], the vibrotactile sensitivity of the hand has been measured to be highest (lowest threshold) at mid-range frequencies (40-150Hz) and decreases (threshold increases) at very low and very high frequencies (below 20 and above 400).

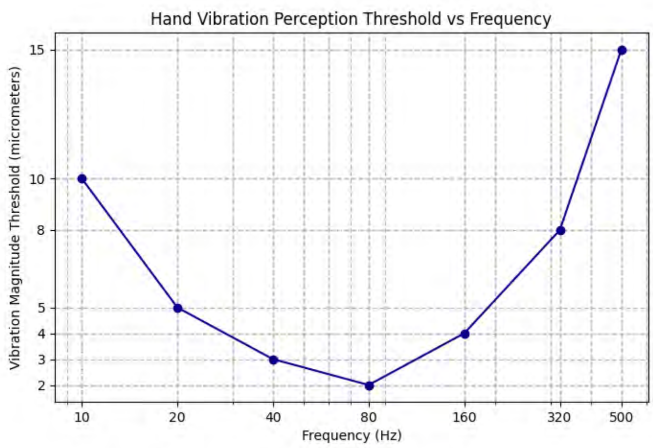


Figure 2.20: A plotting of minimum vibration magnitude required for the hand to perceive vibration. Data collected from sources[30][56][47], plotted in matplotlib using python.

Whole body vibrations are important because humans are used to perceive whole vibrations in situations like being a passenger or driver in vehicles such as cars, trucks and helicopters; experiencing massive motions such as earthquakes and waves in the sea[3]. Whole body vibration is a vibration that affects the whole body, the perceptive result of a stimulus in the brain is a sum of every somatic sensor of our body[68] and it may make us feel diverse feelings from nausea, sickness, to refinement and rigidity[56].

According to a research[49], the perceived quality of an instrument is slightly affected by the vibration level of an instrument and according to other studies[17][74], vibrotactile perception of musical concepts (rhythm, pitch, melody cognition) is quite similar to auditory perception.

In the Audio-tactile Rendering experiment[74], tactile renderings and perceptive effects on different types of auditory concepts such as rhythm, pitch, melody, timbre and loudness were explored and in the Consonance of Vibrotactile Chords experiment[17], musical theory was involved to express musical feelings of consonance and dissonance via vibrotactile interfaces. Both experiments are highly related to my research and Embracing Sphere.

According to study[74], tactile rhythm perception is explained in a basic way of filtering the music signal and using the filtered signal as an exciter for an actuator. Rhythm is described as a pattern of pulses in discrete time and rhythm as a musical feature can be perceived by multiple sensory channels, such as visual, auditory, and touch.

Pitch perception with vibrotactile stimuli is a complex task, as touch has frequency perception limitations. The simplest way is to translate pitch and loudness to vibrotactile stimuli, using loudspeakers or VCAs, which directly convert pitch to frequency and loudness to intensity of vibrations. However, the frequency response of these actuators exceeds skin perception thresholds we described, because that information embedded in high-frequency bands (i.e., over 1000 Hz) might be lost.

Melody builds up as a suitable combination of pitch changes over time. Therefore, most of the limitations for pitch conversion also apply to melody.

Timbre allows the listener to differentiate between tones played from one musical instrument or another. Timbre relies on the frequency content called ”overtones” (i.e., spectral content) of audio signals and overtones are frequencies of sound that are higher than the fundamental frequency of a vibrating object. In a higher registry fundamental tone, an overtone series may start from higher than 600Hz. Therefore, tactile translation of timbre represents a challenge.

The concepts derived from music theory and composition can be utilized for creating abstract but effective audio-tactile scenes where meaning is not so direct but more intuitive.

Another research[5] caught my interest, which experimented with an illusion that occurs during the evaluation of the multi-modal events. According to the study, visual, tactile and auditory information interact and influence each other. Multi-modal illusions like the ventriloquism effect an auditory illusion in which sound is misperceived from a source when it has a different position than the visible source. The effect is most powerful for speech sounds and it happens because of visual dominance over auditory information. It is exploited by stage ventriloquists who practice the art of speaking without moving their lips while manipulating the movements of a puppet.[5].

Figure 2.21: A picture of a ventriloquist Peter Kerscher. Retrieved from `https://commons.wikimedia.org/wiki/Category:Ventriloquism/`

Although the multi-modal events determined by physical rules and usually have the same position in the physical world, it is possible to break these rules in virtual environments to have control and optimization over object interaction.

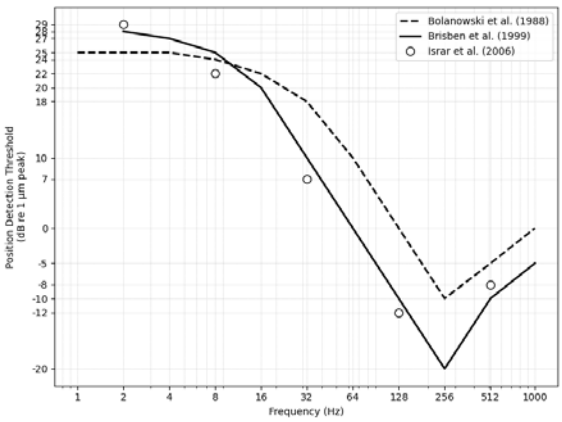


Figure 2.22: A plotting of minimum vibration magnitude required for the localization of a vibration. Data collected from sources[56], plotted in matplotlib using python.

The localization capabilities of humans with vibrotactile stimulation can be seen in the figure 2.22.

The sum experience of simultaneous audio and tactile stimulation is going to be covered in the next section but while composing an audio-tactile experience, these perceptive concepts and illusions have to be taken into account. Being able to author feelings and embed meaning utilizing these concepts is a novel way of conveying a story, in my perspective.

2.3.3 Haptics in Media Arts and Video Games

Both in the media art scene and the video game industry, immersion is occasionally an important subject. This section will cover several case studies in such domains that utilize haptic feedback to translate information, feeling or meaning.

Related Work: Emoti-Chair

The Emoti-Chair is an audio-tactile display designed to translate music, speech, and environmental noises into physical vibrations that can be felt on the body. This system is aimed at improving music accessibility for deaf and hard of hearing individuals, but it also offers a novel sensory experience for all users[51].

Figure 2.23: A visual of audio-tactile display Emoti-Chair.

This work is physically and technically one of the closest works to Embracing Sphere. Both of them are core concepts translating audio signals into vibrotactile stimuli via an array of voice coil actuators embedded in a chair form factor.

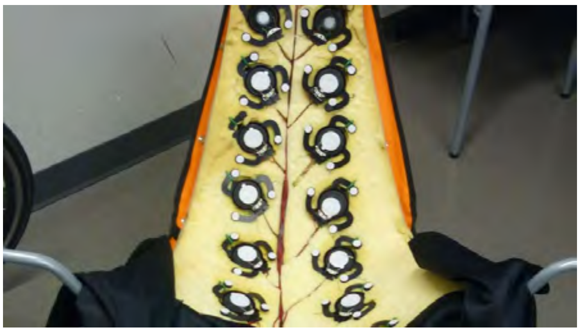


Figure 2.24: Inside parts visual of audio-tactile display Emoti-Chair.

Specifically, Emoti-Chair uses voice coil actuators that are arranged in a two-column by eight-row array, as shown in the figure 2.24, each corresponding to a specific frequency band. This spatial mapping allows users to feel different frequency components of sound at different locations on their body.

In one of the interviews, they have made [6] , researchers explain the system and describe their work as confirming the idea that music is essentially multi-modal and maybe even a-modal.

Figure 2.25: A picture of audio-tactile display Emoti-Chair.

Multimodal interfaces such as Emoti-Chair incorporate multiple forms of input and output to provide a variety of devices to support human-computer interactions. In further experiments, researchers introduced Emoti-Chair to professional film-makers and singers in creating and experiencing tactile music on the Emoti-Chair[8]. They reported on responses to pre and post questionnaires that collected participant views about the workshop and about vibrotactile stimulation in general.

[6] SmartLab TMU News, Emoti-Chair: `https://youtu.be/gA--cOs87p4?feature=shared`

During workshops with professional film-makers, singers and artists, participants either composed vibrotactile music for the first time or experienced their voices as tactile vibrations through the Emoti-Chair. Across these diverse participants, the technology has received positive feedback, with strong interest in using it for future projects. Future directions include enabling artists to further explore tactile composition and developing new instruments specifically for vibrotactile music.

Related Work: Movement and Impact by Yvonne Weber and Sabine Haerri

Up to six million vehicles a year pass through the Gotthard Road Tunnel, Switzerland's most important north-south traffic artery. "Movement and Impact" gives you a completely new feeling for the Gotthard Tunnel and the cars and trucks incessantly pouring through it.[38]

The quote above is taken from an artwork description named "Movement and Impact" co-created by Yvonne Weber, Sabine Haerri and the Ars Electronica Futurelab. I have discovered this artwork through a suggestion from my professor, Manuela Naveau and immediately I felt compelled to research further. It was exhibited in the Ars Electronica Festival in 2009 and unfortunately, I have no chance to experience it again. I will do my best to explain the piece using the resources and visuals I have found online.

Figure 2.26: Movement and Impact. Ars Electronica Festival in Hauptplatz Linz 2009. Foto: ARCHIPICTURE Mag. Dietmar Tollerian.

In this artwork, the artists translated the heavy flow of vehicles passing through Switzerland's Gotthard Road Tunnel into a tactile experience. Sensors laid on the ground captured real-time data on traffic volume, vehicle size, weight, and direction. That data converted into gentle, rhythmic vibrations on a reclining platform[38].

Figure 2.27: Gotthard Road Tunnel, Switzerland.

In Movement and Impact, artists explored the translation of digital traffic data into a physical experience. Through transforming overwhelming, invisible data, they created an experience that was somehow intimate, tangible and even therapeutic.

Figure 2.28: Openstreetview visual of Gotthard Road Tunnel, Switzerland.

Coincidentally, I have chosen a similar practice in haptic content creation for Embracing Sphere. I recorded a bridge rumble which is emotionally close to me, Neue Eisenbahnbrücke in Linz.

Figure 2.29: A picture of Neue Eisenbahnbrücke, Linz.

Since starting my master's studies at the University of Art and Design Linz, Linke Brückenstraße was my first place of accommodation. For a year and a half, I crossed this bridge nearly every morning and night. Sometimes I sat and listened to the waves, sometimes I put my headphones on and only felt the heavy rumble of the big metal frames of this beautiful bridge.

Eventually, I decided to document my experience in an auditory way. However, the experience of this bridge is far from purely auditory, so I grabbed my geophone (an electronic seismic recording device) and recorded the bridge's rumble and the vibrations caused by passing vehicles.

Figure 2.30: A picture of my recordings on Neue Eisenbahnbrücke, Linz.

Figure 2.31: A picture of my recordings on Neue Eisenbahnbrücke, Linz.

In conclusion, Movement and Impact was not something I knew about before making these recordings or starting to develop Embracing Sphere. However, researching this artwork has convinced me to continue along this path. I may be on the right track to express my own experience. The original files of this recording session will be used in Embracing Sphere content.

Related Work: Astro's Playroom

Astro's Playroom is an exclusive PS5 tech demo for the DualSense [7] . In Astro's Playroom, players can feel the texture of sand, the grittiness of ice and the tension of a bow string. Rainfall is simulated as tiny, distinct taps and pulling a bow uses adaptive triggers for realistic tension.

PS5 DualSense utilizes voice coil actuators and adaptive triggers to produce complex vibration signals[25]. Each side of the controller has a combination of a vibrotactile

[7]Short documentary of Astro's Playroom tech demo from Noclip: `https://youtu.be/_WpD8PvH0QA?feature=shared`

actuator, designed by Foster Electric [8] .

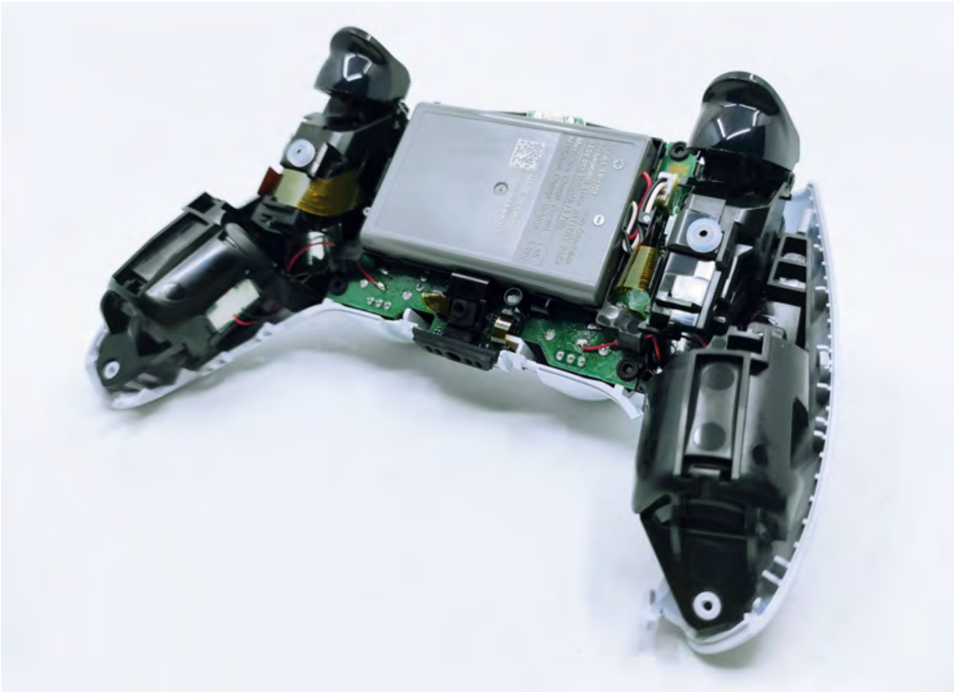


Figure 2.32: Playstation 5 gamepad DualSense.

The haptic capabilities of DualSense are quite intuitive. In ”Chapter 1 - Introduction, Embracing Sphere and My Own Perception” section we mentioned an old technology of Playstation 2 Dualshock 2 in figure 1.7, the haptic technology at that time was using weighted motors that spinning off-center to create a general rumble and it was quite low resolution and not expressive enough in comparison today’s haptic technologies that uses voice coil actuators[74]. The DualSense haptic actuators are basically small speakers that can playback low frequencies clearly for a haptic feeling.

To test DualSense haptic capabilities, I prepared a recording setup [9] . The setup includes 1 low frequency contact microphone (Lom Geofon) and a sound device (Audient EVO 4) for recording.

[8] Commercial video of VCA from Foster Electric: `https://www.foster-electric.com/products/productdata/vibration_actuator_en.mp4`

[9] Demonstration video of the test setup: `https://youtu.be/YB4zvlhoZz8`

Figure 2.33: Recording setup of Playstation 5 gamepad DualSense, contact microphone Lom Geofon mounted with suction piece.

In the test, I played the first couple of minutes of the Astro's Playroom and simultaneously recorded DualSense vibrations with a microphone setup shown in the figure 2.33 and recorded a video of this process. The test focuses on analyzing haptic output for the main character, Astro's footsteps on different materials. In game audio practice, footsteps are designed with variations on different materials[78]. In Astro's Playroom, haptic feedback also varies according to the material that Astro is stepping.

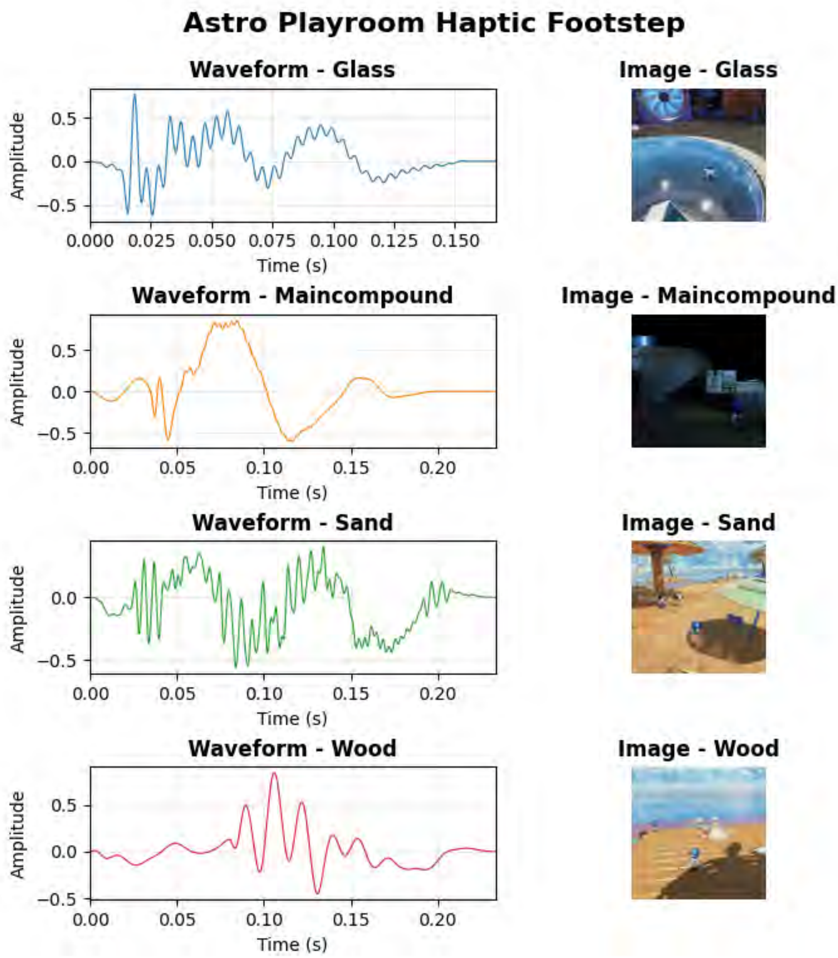


Figure 2.34: Plotting of recorded footstep haptic layers from Astro's Playroom. Recorded on 4 different material.

The above figure 2.34 shows the waveforms of 4 different haptic footstep signals on different materials. As we can see from the waveforms, glass and sand footsteps have higher frequency and specifically, sand footstep has a noisy texture. The test results show us the variations in haptic signals for different texture needs. Game audio designers deliberately designed another layer to enhance multi-modal stimulation.

Related Work: Racing Simulations

Throughout my research, I've explained many different inspirations that I had for the Embracing Sphere. Audio-tactile interfaces I have discovered include many artworks, interactive media or video games that might have been relatively linked with Embracing Sphere in concept or software/hardware side.

This related work investigation about racing simulations will cover current haptic technologies in sim-racing video games and hardware. Additionally, the section is going to be about my own experience with sim-racing and seated vibrotactile stimulation.

In 2021, when I was a bachelor's student and with my dear friend Can Memisogullari, we started to play a racing game called F1 2020. At first, we played through the main content, which is basically racing against AI. After a while, Can mentioned a community that organizes its online racing league in F1 2020.

The league was organized during the COVID times and every member of the league was an adult whom has a lot of free time to spend on practicing and preparing to race competitively. Looking back from now, it feels so absurd that we had set a schedule for a 10 week season, different race each week, that every driver practiced hard on weekly different tracks. There was live coverage with commentary for all races and even a referee committee that gets together after the race to check the incidents that happened in the race.

First season, Can and I raced with generic gamepads and there was a distinction between players who used gamepads and players who used force feedback (FFB) wheels. FFB wheel users were basically faster and more consistent in vehicle control. Of course, there is more than one parameter different between the gamepad and FFB wheel but there was a consensus in the community that the pace difference is coming from the feeling of the vehicle.

FFB is a technology used in racing simulation hardware, especially in steering wheels, to provide physical sensations that mimic real-world driving. FFB wheels use motors and sensors to generate resistance, vibrations, and subtle movements that reflect what is happening with the car in the game. This feedback allows the player to feel important information, such as:

- The grip level of tires.
- Road surface texture (bumps, kerbs, gravel traps)
- Weight transfer of the car while cornering, braking and accelerating.

When racing with a gamepad, the only feedback comes from visual and sometimes subtle vibration cues. This makes it harder to judge the car's behavior, especially at the limit. With an FFB wheel, the physical sensations are much richer and more detailed. This difference is why FFB wheel users tend to be faster and more consistent in a sim-racing league.

With that information, I decided to buy myself an FFB wheel. The wheel I bought was a belt-driven Thrustmaster T300 RS GT, as seen in the figure. 2.35.

Figure 2.35: A picture of my simulation rig, Thrustmaster T300 RS GT.

After I adapted my driving style and muscle memory to a wheel, I've started to see improvements in my peak pace and overall consistency around the track. I believe it is mostly because of the constant vehicle state information stream to my nervous system with FFB. I could feel the loosening grip on the back tires when I made a mistake out of the corner. The more interesting part, I started to make real life reactions to vehicle behavior such as counter steering and brake pressure managements [10] .

Sim-racing softwares (racing video games) are developed with realism taken into account and sim-racing games are a benchmark for immersive experiences in the gaming industry, especially through their advanced use of haptic feedback. Video games such as iRacing and BeamNG.drive are renowned for their realistic simulation of vehicle dynamics, which is later integrated into haptic hardware such as FFB wheels and bass shakers.

These simulation games are designed to take full advantage of haptic hardware. For example, iRacing provides highly detailed telemetry data that is used by force

[10] A short clip of my driving and counter steering to losing back of my car in the online sim-racing league: https://www.twitch.tv/mehmetcolak/clip/KnottySmoothJuiceCclamChamp-5y_nOCkwwXRh4bjq

feedback systems to replicate real-world car behavior. Third party softwares can receive this telemetry data, evaluate and output a haptic feedback signal stream. This integration makes the player feel every bump, skid and collision.

Figure 2.36: A visual of user interface of a third party telemetry software for iRacing.

Since 2021, I have followed the technological developments in sim-racing softwares and hardwares. As my experience directs, I believe that it is really immersive and most of the reason for immersion was coming from haptic feedback that was evaluated through simulation telemetry data. Because of that, in Embracing Sphere, I chose to utilize a similar setup to create a virtual environment for navigating and exploring.

Related Work: Feelies in Brave New World by Aldous Huxley

Brave New World is a dystopian novel by English author Aldous Huxley. He wrote Brave New World in 1931 and published it in 1932[43]. The story begins in a distant future London, where people are engineered and conditioned from birth into fixed classes: Alpha, Beta, Gamma, Delta, and Epsilon. Embryos are chemically altered to suit their future roles. Higher classes are genetically enhanced, while lower-class citizens are deliberately impaired. Each class is trained/conditioned to accept its status, wears distinct clothing and fulfills specific societal functions such as leadership or manual labor.

In Brave New World, a distinctive imaginary multi-sensory entertainment medium called the ”feelies” is introduced. This sensory apparatus serves as a cinema-like experience but unlike conventional cinema, it incorporates not only sight and sound but also smell and touch. The feelies are depicted as one of the society’s control mechanisms, alongside drugs like ”soma” and sleep-learning techniques, ”hypnopaedia.”[31]. The masses kept under control and satisfied using feelies as an amusement tool within the theme of sensory colonization[36].

According to Grossi[36], the fantasy of total sensory cinema is deeply tied to the configuration of a mass society and the programming of the human psyche. The pejorative descriptions of these technologies aside, the idea of an enhanced multi-sensory medium and imagining this during the golden age of real-world cinema fascinated me. Even today, the development of immersive media continues to involve the direct ”colonization” of human sensory fields such as virtual reality and augmented reality technology advancements.

Eventually, Embracing Sphere can be seen as a re-evaluation or re-imagination of many of these examples discussed in the section ”Haptics in Media Arts and Video Games”. As the artist behind this project, I have tried my best to be honest and true to my own creation. Practicing artistic/scientific research about my work was the main control system of my artistic outputs, providing me with many references in my domain and my pursuit. With the inspirations covered in detail in this section, the next chapter, ”Personal Project, System Design and Methodology” will detail concept and design features of Embracing Sphere.

Chapter 3

Personal Project, System Design and Methodology

3.1 Conceptual Framework of Embracing Sphere

This section details the concept and design features of the Embracing Sphere installation. It describes the main conceptual components chosen and explains how they integrate to create an audio-tactile environmental storytelling experience.

Embracing Sphere aims to combine key theoretical constructs that have been covered in Chapter 2 (environmental storytelling, multi-modal perception, and audio-tactile interaction). This vision directs the design and the implementation of the Embracing Sphere.

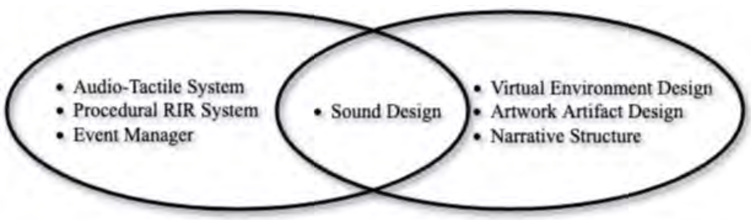


Figure 3.1: Venn diagram of conceptual elements of the Embracing Sphere.

The above diagram shows conceptual elements of the Embracing Sphere with ”technical” and ”artistic” categorizations. This diagram shows my consideration points when creating the experience, such as ”How audio-tactile system is going to be displayed to the audience?” and ”How much narrative should I enforce in the experience?”. These considerations basically constrain my technical and artistic decisions according to the requirements of the exhibition thematically and technically.

For example, the audio-tactile display equipment is chosen with consideration of the collaborative exhibition setting of the Ars Electronica Festival Campus Exhibition. Therefore, instead of a loudspeaker array, a headphone and 3 bass shakers are chosen for the audio-tactile play system. In this setup, the headphone is displaying a binaural mix of audio channels composed in ambisonic format for 3D audio.

3.2 Procedural RIR Generation System

This section will detail the procedural RIR generation process in Embracing Sphere. Every different surrounding of ours has different acoustical parameters and in the time domain, has different characteristics. Some enclosures have more dull and short responses, while others can be bright and long. These characteristic differences are covered in section 2.2.2 Room Impulse Response Measurement Methods.

Utilizing these characteristic differences may apply contrast and environmental shifting abilities to the auditory face of Embracing Sphere. Room impulse responses can be archived and used on demand with an RIR bank system that changes the RIR file in convolution reverb, but artistically, I chose a more aleatoric way of generating RIR files with some parameters introduced in advance (procedural generation).

Sabine's formula, which is covered in 2.2.1 Definitions for Room Acoustics, is simple enough to implement and outputs a big enough range of values to hear the difference. The RT60 parameter generated with the Sabine formula is used in the RIR generation module.

$$T_{60} = \frac{0.161 \cdot V}{A}$$

In detail, the Sabine formula needs effective surface area (A), volume (V) variables to output the RT60 value. Randomly generating numbers for an imaginary room's dimensions can enable us to calculate an effective surface area and volume.

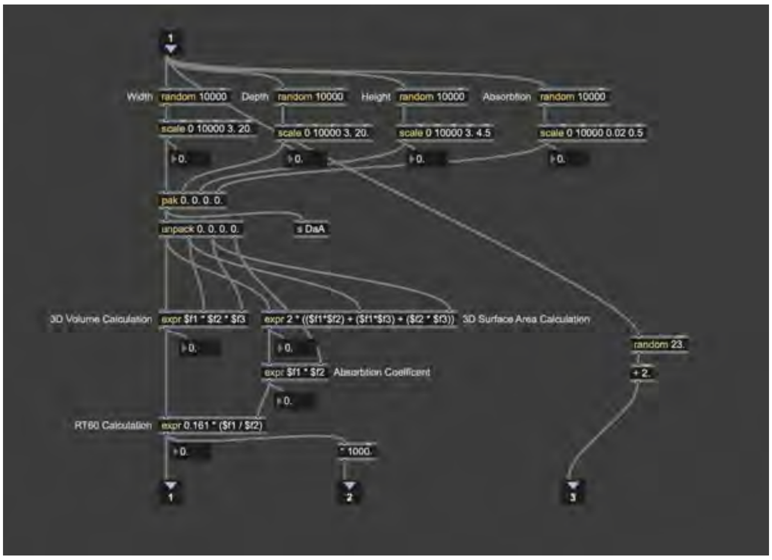


Figure 3.2: RT60 value generator with randomized room dimensions and absorption parameter. Developed in Max/MSP visual scripting software.

As seen in the figure 3.2 4 random number (3 for dimensions of the room, 1 for overall absorption coefficient) generated using Max/MSP random number generator object, with the scale object parameters the output numbers limited between 3 meter to 20 meter in width and depth and 3 meter to 4.5 meter for height of the imaginary room. The rest of the expression objects are implementing the Sabine formula into the module. The output defines the length of the RIR file.

RT60 alone is not enough to create an RIR. At the same time, we need to generate a gradual fade and filter out envelopes to imitate an absorption slope. Subtractive synthesis came into formula at this point, basically gradually filtering and turning down the volume of a white noise, with the subtractive synthesis approach can be utilized in procedural RIR generation.

A flowchart shown in the figure 3.3 shows a high level process flow in the procedural RIR generation. The output of this process is not a simulation level or mentioned RIR capturing methods level but this primitive approach generates fast, realtime RIR files to use in convolution reverb.

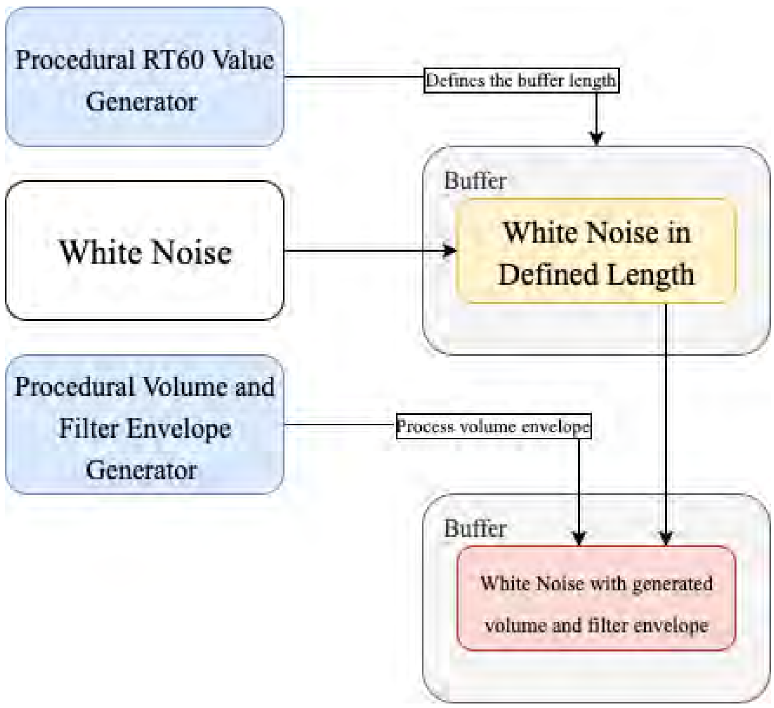


Figure 3.3: Flowchart of procedural RIR generation module.

With a high level explanation of procedural RIR generation, the next section will cover haptic and audio content generation and playback system details.

3.3 Audio-Tactile Content Design and Playback System

Embracing Sphere is developed as an interface rather than artistic content. Although it is going to be exhibited in Ars Electronica Festival - Campus Exhibition, the system was developed to have adaptability so that audio and haptic content of the system can be authored and updated.

The artwork can be seen as a theme park train ride or commercially named ”4D Cinema” experiences. Nonetheless, this research is investigating audio-tactile interfaces as the main focus. There is a narration to follow that is going to be detailed in the next section but the main drive here for the Embracing Sphere is feeling the environment and making the audience interpret their own environmental storytelling cues.

3.3.1 Audio Content and Playback System Technologies

Given collaborative exhibition situations taken into account, we decided to choose a regular headphone so as not to interfere with the other artworks. Following the headphone decision, multi-channel playback system requirements reduced dramatically.

In this plan 2 channel audio and 2 channel haptic signals are composed for the experience. For the length of the experience, the average audience attention span without visual motion graphics considered and according to the research around this topic[12], approximately 8 minute experience is planned.

Stereophonic audio was first patented by Alan Blumlein, a renowned electronic engineer, with his inventions in telecommunications and sound technologies[84]. The headphones we are still using now utilize the same principle as the first stereophonic system. In the vision of creating audio-tactile virtual environments, stereophonic compositions are not enough to meet the spatial needs of the auditory domain. Thus, ”binaural” mixing has been chosen to leverage spatial audio technologies such as ambisonics.

Binaural mixing, unlike traditional stereo, recreates the natural listening experience by profiling how our ears receive sound in three-dimensional space. This technique utilizes Head-Related Transfer Functions (HRTFs) to simulate how sound waves interact with our head and ear cone before reaching our eardrums. Binaural audio mixing processes audio with these psychoacoustic filters to create convincing spatial perception when played through headphones. With binaural audio, listeners perceive sounds as coming from specific locations around them[32][70].

For being format-agnostic in future projects, the main content will be composed in ambisonic format. Unlike channel-based surround sound systems that assign specific loudspeakers to discrete audio channels, ambisonics captures and reproduces the complete spatial sound field with spherical harmonics mathematics. Ambisonic recordings can be decoded for any loudspeaker configuration or downgraded into a binaural format for headphone playback to maintain spatial accuracy across different playback systems[88].

Figure 3.4: Trademark logo of Ambisonics.

In practical terms, an ambisonic format audio file appears as a multichannel audio file in audio software, just like any other channel-based format but it requires a decoder to correctly route the information in the multichannel audio file to specific loudspeakers in order to play the audio file as desired.

For the Ars Electronica exhibition, a Pure Data patch was developed for binaural decoding and convolution DSP processes. The mentioned pure-data patch will be explained further in the next chapter ”4.2 Software Development” section.

3.3.2 Vibrotactile Content and Playback System Technologies

Vibrotactile content creation in Embracing Sphere combines 2 different approaches as sample based content creation and synthesized content creation. The same approaches derived from traditional sound design practices but focused on low frequencies, constrained by vibrotactile display and human haptic perception limits.

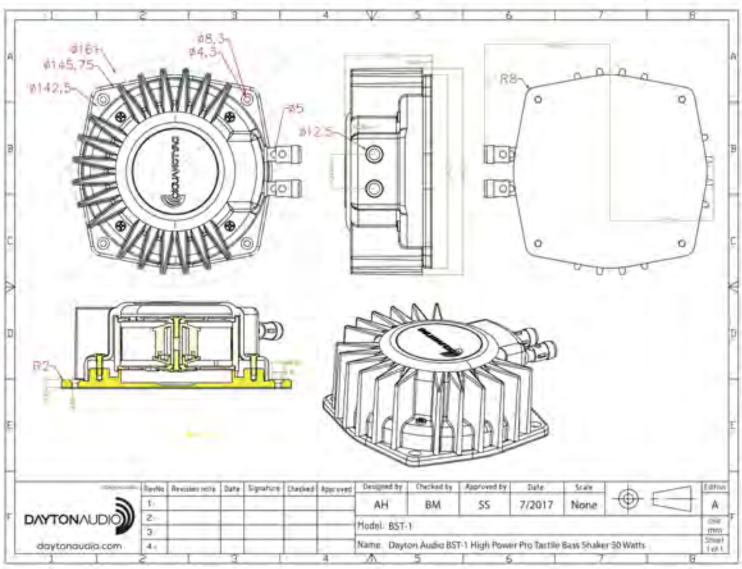


Figure 3.5: Mechanical drawing of a voice coil actuator, Dayton Audio BST-1.

Main vibrotactile actuator shown in figure 2.18, 50W 4ohm voice coil actuator specified in the spec sheet as output frequency between 10Hz and 80Hz with a resonant frequency specified as 30Hz. A signal that will be displayed through this actuator should be in that frequency range. Therefore, a sensitive geophone (Lom Geofon) is used for sample based approach in vibrotactile content creation.

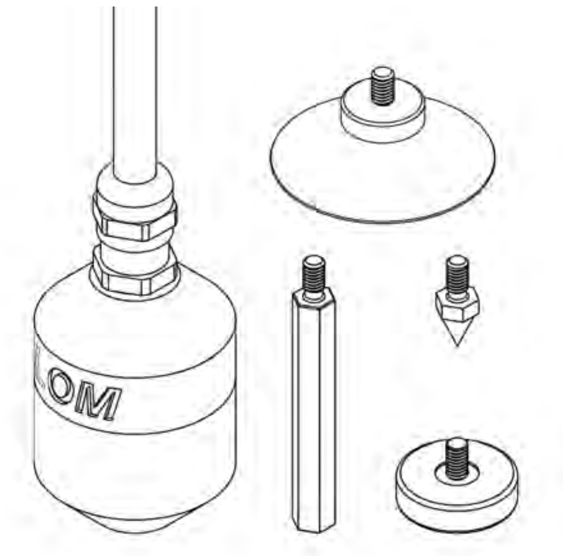


Figure 3.6: Wireframe drawing of a geophone microphone, Lom Geofon.

Lom Geofon product spec sheet indicates frequency response between 10Hz and 1000Hz with resonant frequency at 14Hz. This frequency response specifications meets our requirements. The specific microphone mentioned in preceding chapters is this microphone, which was used in sampling and documenting low frequency vibrations on solid objects, such as a metal bridge, shown in the figures 2.31, 2.30 and PlayStation 5 gamepad DualSense shown in the figure 2.33.

The plan is to collect as many solid vibration textures as possible, edit and display these vibrations through a voice coil actuator array, which in the proposed setup, 3 voice coil actuators are placed around different positions of the audience to achieve spatial perception in the vibrotactile display as an audio playback system.

Within this plan, utilized softwares and more details on hardwares will be covered in the next chapter ”Implementation”.

3.4 Narrative Structure

To effectively communicate in this section, generative adversarial networks (GANs) are utilized to create storyboard visuals that help readers understand the narrative structure and progression of the installation (the used text prompts can be seen in the appendices). These AI-generated illustrations serve as visual references that can convey the abstract concepts and environmental transitions that occur during the Embracing Sphere experience.

Figure 3.7: The poster cover of Ars Electronica Festival 2025, Panic: Yes/No?.

The Ars Electronica Festival theme in 2025 is selected as ”Panic: Yes/No?”. To match the theme, narrative content is shaped around claustrophobia and drowning.

The experience plays a looping sound of bridge rumble, noticeable in close proximity to the artwork. When the audience decides to sit in the seat, the main narrative experience starts.

In the first seconds of the experience, opening with first person character perspective, an individual sits on a bench on the metal constructed bridge, listening to crows in the cold, cloudy winter afternoon.

Image 1

Then a car approaches, opens the window next to the passenger seat, and honks once. That imminent honk starts flash scene changes, and with each scene transition, we hear the honk sound in a new environment. After this short flash scene sequence, our character decides to get into the car.

Image 2

They close the door and immediately feel isolated from the outside world, losing the vibration of the bridge as the car starts to accelerate.

Image 3

After a short acceleration, the car gradually starts to drive out of the lane. Eventually, the car jumps off the bridge and falls into the river.

Image 4

The car starts to fill with water and is swollen by the river. Claustrophobic scene cues with struggling to breathe, heavy pressure from the bottom to the head. After a short time of struggling, a state of tranquility arrives.

Image 5

In this tranquility, the first and only voice line is introduced: ”I’ve seen things, you people wouldn’t believe.” This line is inspired by the famous monologue in the Blade Runner movie directed by Ridley Scott. Just like the horn sound, the word ”thing” in the voice line drives flash scene sequences with louder and faster transitions.

Image 6

Image 7

Image 8

Our character witnesses memories or scenes of their life and accepts their fate, falling into abstract space.

Image 9

Eventually, they regain consciousness, floating on the river with heavy breathing and waves splashing on them. The experience initiates an end loop that plays looping wave and breathing sound cues in the river until the audience stands up and leaves the experience.

Image 10

Within this story, I, as the author, aimed to convey as many environments as possible. Creating a narrative around the "Panic: Yes/No?" theme, I chose to explore my own instinctive panics and fears. The story is written with little to no voice-line interruptions, with the only voice line elevating new flash scene introductions.

For interactivity and playback structure, I planned two looping states and one fixed state for the playback system. This system enables the audience to first feel a glimpse of the experience through their feet when they approach the artwork. Later, the experience starts intuitively, and after the fixed state ends, the system transitions seamlessly to an ending state loop.

Chapter 4

Implementation

The Embracing Sphere installation is built into a physical car seat. When a user is seated, the system plays a pre-authored experience composed of synchronized audio and vibrotactile events. The following subsections detail the hardware and software components that set up the installation.

4.1 Hardware Setup

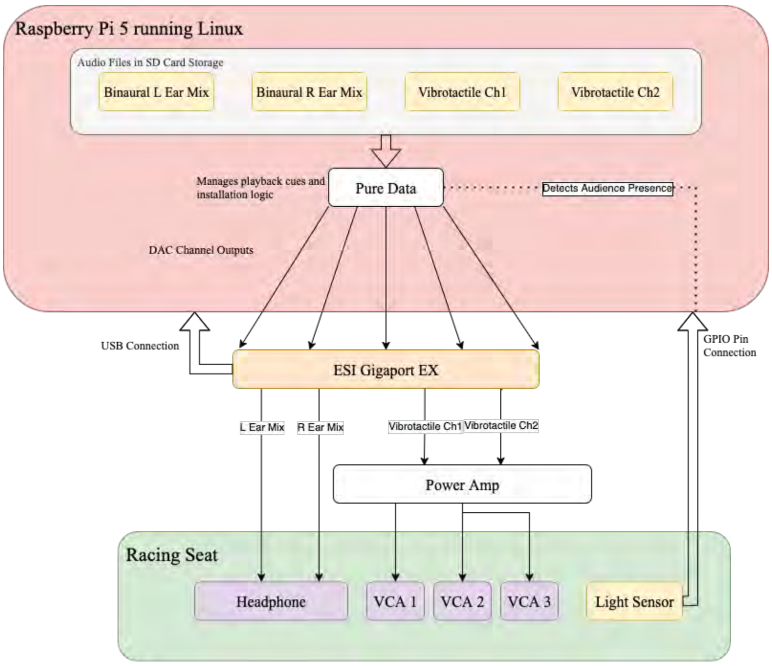


Figure 4.1: Planned flowchart of the Embracing Sphere installation.

As the main processing unit of the playback system, a Raspberry Pi 5 has been chosen. Raspberry Pi is a small processing unit that can run traditional personal computer operating systems such as Linux. This small computer has been chosen for its small form factor and GPIO pin usage.

Figure 4.2: Processing and digital audio converter units of the Embracing Sphere installation. Raspberry Pi 5 and ESI GigaPort EX.

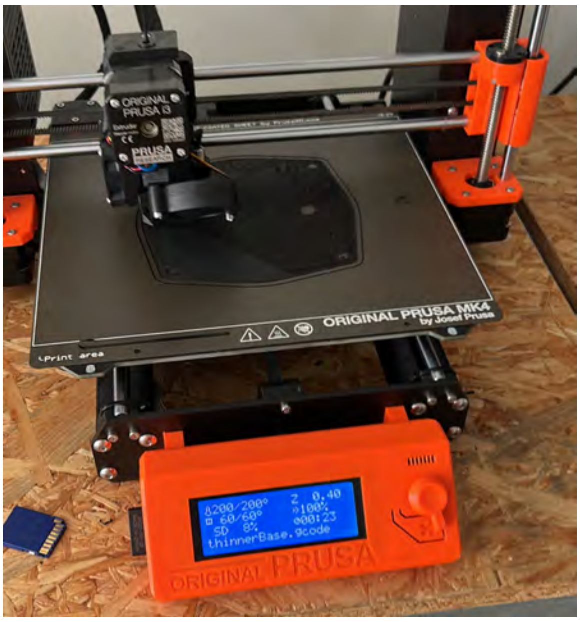


Figure 4.3: 3D printing of VCA mounting brackets.

The playback system is interactive according to audience behavior, with limits and some sensors required to use as a detection point. In that case, a photoresistor is used to detect if any user is seated or not by checking the light level coming to the sensor. If a user is seated, the idling loop stops and the main experience initiates.

The audio channels that drive headphones require 2 channels to play binaural audio, which isolates the listener within the sonic environment and another 2 channels are required to drive voice coil actuators.

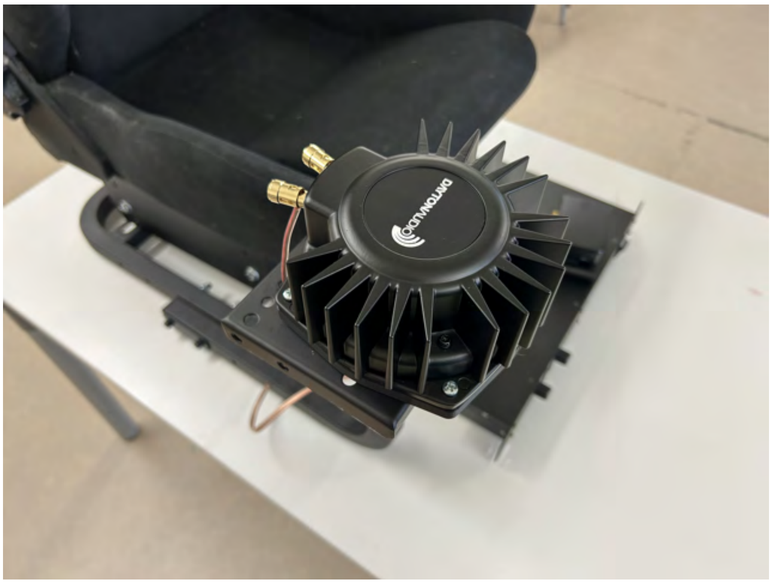


Figure 4.4: First bass shaker positioned on the right hand.

In total, at least 4 channels are needed to create this playback system that drives both auditory and vibrotactile parts of the installation. For 4 independent channels of audio output, the ESI GigaPort EX sound device has been chosen. Sound devices are hardwares used in recording and monitoring efficiently through ADDA (analog to digital, digital to analog) converters[20]. This sound device has 8 individual audio outputs and it is class compliant, available to use in Linux systems like Raspberry Pi 5.

These 8 analog outputs are line level -10 dBV RCA outputs, which are enough to drive headphones but lack the power to drive bass shakers that need 50W of energy to work. To amplify the haptic channels, the Thomann t.amp S-100 MKII power amplifier was selected.

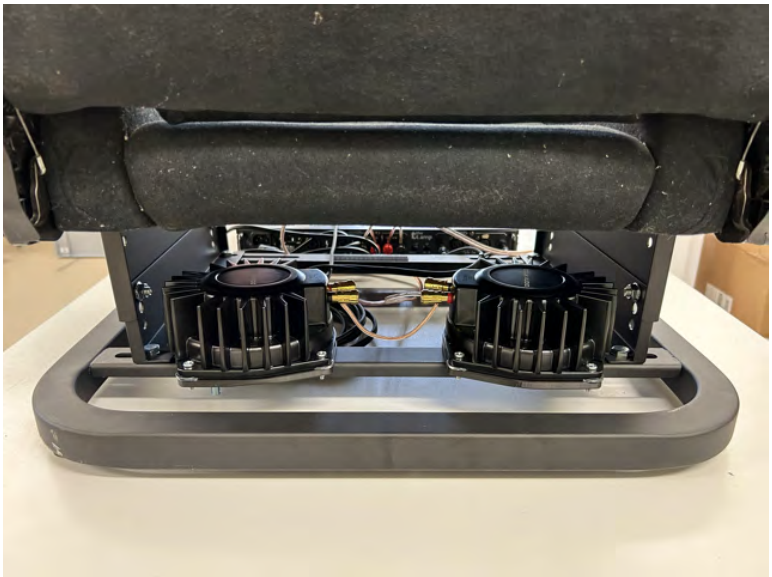


Figure 4.5: 2 bass shakers positioned underneath the seat, screwed into the metal chassis.

Figure 4.6: Fully assembled playback system with seat of Embracing Sphere installation. Power amplifiers for VCA's are visible in front of the seat.

Mounting VCA's into the seat and connecting cables into the sound device, then the processing unit was very intuitive. A sim-racing metal frame used for fixing everything into the seat. The metal frame also helped in the transmission of vibrations through the chassis, which elevated the whole body vibration feeling.

Hardware setup and tests, such as sine and noise signals and transient signals for testing the lag between the headphone and the actuators, have been done. After setting up the signal flow, a stress test was made where every actuator was tested in the loudest state without clipping in any part of the system. I was quite impressed by the results of the first tests. Overall, the loudness was sufficient and the vibrotactile feeling was distinctive.

4.2 Software Development

After assembling every instrument together, software development and practicing sound design started. The Embracing Sphere includes fixed media in the main content (fixed time and event sequences) and a procedural effect chain. Therefore, I utilized a Digital Audio Workstation (DAW) called Reaper, for sound design and authoring of the sound design. To achieve user interaction and procedural effect chain, I used Pure Data, a visual scripting software for interactive audio-visual projects.

In Reaper, I have structured a hierarchy in the sound layers of the experience. In that structure, audio signals and vibrotactile signals are separated into 2 different effect chains. Audio signals/events in the audio channels are first spatialized with the IEM Suite plugins (Institute of Electronic Music), an ambisonic plugin toolkit that has utilities for spatial encoding and decoding, then decoded for binaural monitoring and that monitored output is exported to use in the Pure Data patch.

Vibrotactile signal just filtered with a low pass filter, a filter that filters the audio signal above certain frequencies. This filtering was necessary to display audio signals with the bass shakers, as mentioned before, bass shakers have a ceiling on frequency to display. Any signal that has frequency content above 80Hz creates unwanted artifacts and issues on the bass shaker.

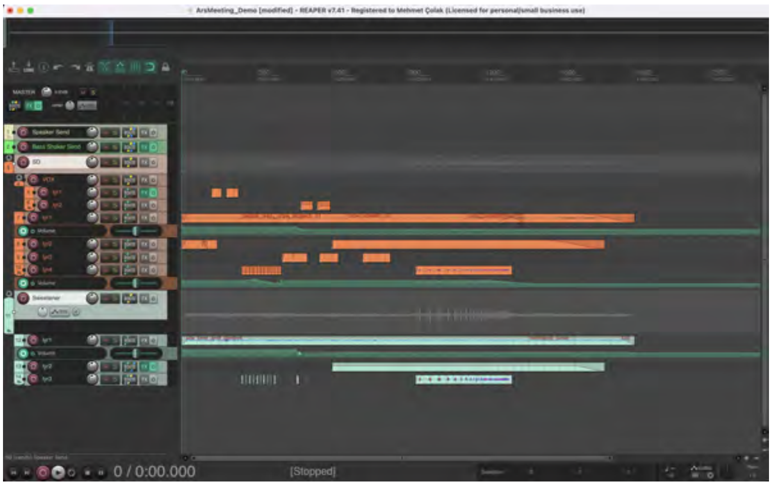


Figure 4.7: Screenshot of the Reaper project for Embracing Sphere sound design.

After the sound design process finished, a test patch was created on Pure Data. This patch allows for simulating an exhibition environment by changing play states on mouse click. Image of the patch can be seen in the figure 4.8 below.

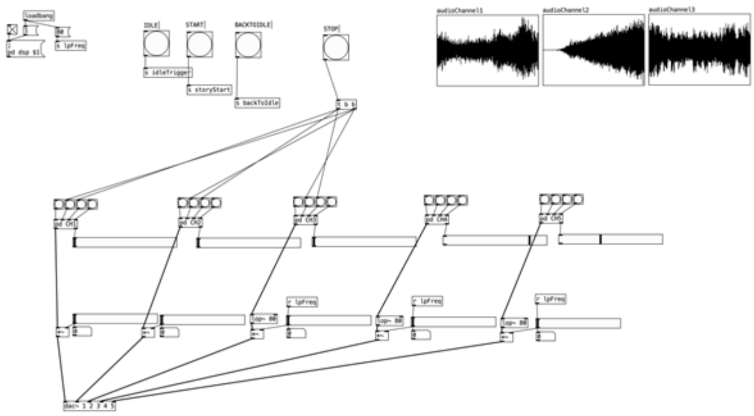


Figure 4.8: Screenshot of the Pure Data patch for testing audio files and play states in the exhibition scenario.

Further developments on the patch include translating the procedural RIR patch from Max/MSP to Pure Data and setting the patch as a startup command in the Raspberry Pi 5 to ensure seamless integration of the artwork in the exhibition.

The command was made by editing the autostart file in the Raspberry Pi 5 by adding a line shown below.

```
sudo pd -nogui -alsa /home/pi/embracing_sphere/main.pd
```

This specific command starts the patch on startup with given instructions such as ”-nogui”, which directs no graphic user interface while playing and audio engine instructions for Linux.

4.3 Challenges and Solutions

One of the biggest challenges I have seen while I was developing Embracing Sphere was finding good sources for vibrotactile content creation. Despite many written and video sources available for sound design and audio content creation, tutorials on vibrotactile content creation or fundamental knowledge were not enough in comparison with the audio domain.

I have solved this issue with reverse engineering already established interactive applications, such as video games and racing simulations. By sampling these applications, I had a chance to examine vibrotactile content. Some samples I recorded are used in Embracing Sphere’s vibrotactile layer, such as the tire and engine rumble of the vehicle.

Figure 4.9: A picture from a simulation rumble signal sampling session. Assetto Corsa (a vehicle simulation game, developed by Kunos in 2013) and sim-hub (an add-on tool for telemetry logging and generating rumble signal with telemetry data accessed from simulation) were used for vehicle simulation, recorded through a sound card.

Another challenge I've faced is that the trial-and-error pattern was rather slower than desired. To solve this challenge, I chose the same approach as the Emoti-Chair project [8]. I used the seat itself to author the audio and vibrotactile content. I sat on the seat while editing and designing and monitored the overall feeling immediately. This approach reminded me of theater mixing in the cinema industry, where tailored content is directly displayed and evaluated in the corresponding display system.

Figure 4.10: A picture from my in-display mixing process. It shows my testing practice after rumble signal captures from simulation applications.

In design and storyboard preparations, I've faced a challenge of finding the right multi-modal sensory scenarios to successfully create an environmental storytelling. After long thinking about real-life experiences that affected me directly with auditory and tactile modalities, I came up with a story that happens mostly inside a car. This decision is derived from both my own experiences and research about presence, "At the Heart of It All: The Concept of Presence"[59]. According to this research about presence, the experience of being inside a moving vehicle represents a perfect example of "presence as transportation," where multiple sensory modalities work in harmony to create the illusion of moving through space while remaining physically stationary relative to the vehicle interior.

The research shows that presence works best when visual, audio, and tactile feedback are combined. Vibrotactile modality highlighted the importance, especially for creating realistic environmental experiences. This multi-modal approach aligns perfectly with the aims of Embracing Sphere, as the car environment naturally provides rich opportunities for meaningful vibrotactile content, such as the subtle rumble of the engine transmitted through the seat to the varying textures of different road surfaces felt through the vehicle chassis.

Secondly, from my own experiences, I think that creating contrast situations and implementing these two distinctive situations enhances both ends of the spectrum, such as a subtle vibrating bridge to saturated stimulations of being inside a car that is falling into the sea.

Embedding environmental storytelling cues in this story was the biggest challenge I've faced. Because of the short length of the story, I placed short term environmental cues such as rumble strips on the road and a semi-open window next to the passenger seat, which results in catastrophic results. For more long term environmental cues, I used a more abstract approach by using procedural RIR to enhance flash memory sequences more vibrantly and distinctly.

Ensuring the user experience will be the same as my intentions within the story was hardest, but this is the challenge of many creative practices. For now, I cling to believing in my creative craft.

Chapter 5

Conclusion

5.1 Summary of Findings

The experience of researching and experimenting was no less than great for me. Throughout the project realization, I learned and tested many different techniques. This long marathon of writing about my research and Embracing Sphere while simultaneously assembling Embracing Sphere for the Ars Electronica Festival expanded my skillset.

Embracing Sphere, an Environmental Storytelling with Audio-Tactile Playback System, started with the question of "how can environmental storytelling be conveyed through non-visual mediums?". Later, it became more focused research about audio-tactile interfaces and their usage in creative works such as narration and environmental storytelling with the pursuit of a multi-modal sensory medium. My background in technical audio design helped me a lot when structuring this multi-modal playback system and deciding the technologies that I want to use in the Embracing Sphere.

The considerations about the technical requirements of the work eventually shaped the results. Investigating the potential of audio-tactile systems for environmental storytelling, this thesis created an interdisciplinary grounding, mentioning game audio, sound design, human-computer interaction and artistic context this project is situated within.

The theoretical foundation and examples in the Chapter 2, written to establish and support the argument that environmental storytelling can be efficiently communicated through multi-modal, especially audio-tactile, channels. For the novelty of the research, both of the modalities investigated with specific techniques, such as procedural room impulse response generation in the auditory domain and reverse engineered sampling of haptic signals from racing simulation games and solid surface vibrotactile recordings from games and real life structures.

I hope the use of all the technologies and techniques that I investigated was satisfying when translated into my artistic output, Embracing Sphere. Most of the technical execution (the hardware and software integration) involved trials and errors, resulting in self-observatory decisions except for the headphone usage as an auditory display. Although artistic content (sound design, virtual environment, narrative structure) decisions are affected by the theme and the concept of the Ars Electronica Festival. One of my aims was to reach a format or theme agnostic system on both the hardware and software side of the project. Personally, I believe the current version of Embracing Sphere can be adapted and developed according to different needs and themes in the future.

5.2 Observations from Ars Electronica Festival

This section will cover the Embracing Sphere installation exhibited at the Ars Electronica Festival and will show a collection of qualitative feedback through informal interviews and conversations with visitors of the Ars Electronica Festival. The opportunity to present the work to a public audience was earned via Interface Cultures master's programme and the installation exhibited in the Post City building in Linz.

The Ars Electronica Festival in Linz is an annual international festival for digital arts and new media. The festival focuses on art, technology and society, making it a relevant exhibition for presenting my audio-tactile installation Embracing Sphere and Post City, the exhibition venue, is a converted postal center in Linz.

The collaborative exhibition environment at Post City meant sharing space with multiple other installations, creating practical constraints that significantly impacted the Embracing Sphere's design and implementation. This shared context demanded that the installation communicate its interaction model quickly and clearly while competing for attention within a dense, multi-sensory environment filled with other technological artworks.

During the Campus Exhibition, I worked as a production coordinator, which gave me more time to set up and test Embracing Sphere properly. Being able to do time management better, I could adjust details such as installing lights to set the mood, attaching the sensor to the seat for more accurate measuring. By utilizing this extra time, I refined the experience and felt more confident before the audience arrived.

Figure 5.1: Picture of the Embracing Sphere in the Ars Electronica Festival, Campus Exhibition in Post City Linz.

To gather systematic feedback during the exhibition, I employed a combination of on-site observation, informal visitor interviews and conversations. Key audience reactions and behavioral patterns observed throughout the festival's duration, highlighting initial engagement and notable user responses to the audio-tactile system.

One very distinct observation that I have is, the average audience athleticism was worse than I expected. As the installation has a racing seat that requires sitting on to feel the experience, the metal frame and seat are placed close to ground level, about 35 cm high at sitting level. Many people struggled to stand up while grounded that low without any bars to hold. This is my subjective interpretation but I also think many people didn't bother to try the experience for the same reason.

Another physical and positional observation that requires adjustments for future developments is the position of the bass shaker for the hand is not optimum for people who are shorter than average.

An observation that I had expected was the spatial definition of the binaural mix in the audio playback. It was definitely a downside whilst I could have used multichannel speaker system for a more precise spatial feeling but as I stated in the previous chapters, the Ars Electronica Festival exhibition was a collaborative exhibition, so I had neighbours that I have to be careful to not to be distracting.

The sound piece of the story had a 2 second silence between the car driving off the cliff and hitting the river to emphasize an ungrounded feeling. Many people misinterpret this silent part as ending, remove the headset and stand up and eventually couldn't experience the rest of the story. In my authoring process, I viewed that silence is marginally acceptable but according to my observations, it was the opposite.

Many visitors approached me and asked about the vibration mechanism, such as "What is under the plush toy that vibrates?", "It felt like I was in a car. How did you do it?". I get these questions as an accomplishment for my design decisions because on the outer shell of the Embracing Sphere, I aimed to hide every technical component to emphasize the immersion and audio-tactile interface as an artifact. Immersed visitors were eager to learn more about the technical practices behind Embracing Sphere.

5.3 Answering Research Questions

- To what extent does multi-modality play a role in the environmental narration?
- How can multi-modal stimulation be utilized to enhance auditory environmental narration?
- How effectively can an interactive audio-tactile system convey distinct environmental characteristics (space size, material properties, etc.) and narrative cues to an audience?

This section will cover the three central research questions of this thesis, as indicated above. Beginning with an evaluation of how multi-modality informs environmental storytelling in the context of audio-tactile integration.

The influence of multi-modal stimulation on auditory environmental storytelling is examined by reflecting on the system's design and the observed effects of combining haptic and auditory feedback on immersion and spatial perception. According to personal observations, Embracing Sphere as a playback system was quite efficient in describing a virtual environment with audio-tactile channels. Although the environmental storytelling side was either hard to examine, since each audience is unique and their focus or critical listening capabilities may affect their interpretation of the story.

The second research question is answered throughout the thesis in detail. I looked for the concepts that I can derive from real world immersive experiences, such as vibrating enclosures, surfaces and surroundings that we feel regularly in real-life (vehicle, bridge, underwater).

Embedding story cues into these surroundings was the main challenge in that subject and from my point of view, I managed to embed audio-tactile information into the virtual environment. That information could have been missed or neglected by the audience but environmental storytelling, in essence, lives in this neglected domain that demands intentional exploration.

In the 3rd research question, I can self direct the biggest criticism, which is that before this research, I was assuming that the space perception in the auditory domain is really intuitively connected with the reverb and reflection character of the space. One aspect I missed was that the time difference between the initial sonic event and the first reflected sound also gives highly distinctive cues about the size of the space. This concept is already known as "early reflections" or "pre-delay" in digital audio processing vocabulary. Further developments of this project will consider adding this feature to the procedural room impulse response generation. The rest of the features developed in the auditory channel worked well and managed to convey narrative cues to the audience, since it was inheriting most of the methods from cinema and game studies.

According to the noticed issues in the system, further developments will be focused on spatial precision and ergonomics.

Ultimately, a multi-modal system developed under the research of environmental storytelling, audio and haptics. With the support of many exterior sources, I pursued the potential of this combination. From an academic standpoint, potential examination still requires more qualitative and quantitative evaluation methods, while from an artistic researcher standpoint, I hope to establish a new interface, transmitting the information to humans in a novel way. Embracing Sphere as an interface for a newborn and it is going to be developed further with many new content and technological improvements.

Bibliography

[1] Jonathan Abel and Wieslaw Woszczyk et al. Recreation of the acoustics of hagia sophia in stanford's bing concert hall for the concert performance and recording of cappella romana. In ?, editor, *International Symposium on Room Acoustics*, 2013.

[2] Omar Al-Tabbaa and Zaheer Khan. Unravelling the complexity of the video game industry: An integrative framework and future research directions. *Telematics and Informatics Reports*, 12:100, 2023.

[3] Ercan M. Altinsoy. *Auditory-Tactile Interaction In Virtual Environments*. PhD thesis, TU Dresden, 2005.

[4] Ercan M. Altinsoy and Sebastian Merchel. Brtf (body related transfer function) and whole-body vibration reproduction systems. *AES*, 2009.

[5] Ercan M. Altinsoy and Maik Stamm. Touch the sound: The role of audio-tactile and audio-proprioceptive interaction on the spatial orientation in virtual scenes. In ?, editor, *Proceedings of Meetings on Acoustics*, volume 19, 2013.

[6] Nurul Amin. The evolution of hagia sophia: A historical analysis. *Journal of the Department of Islamic History and Culture*, 1(1), 2024.

[7] Christoffer Andreas and Jens Holger et al. The acoustical history of hagia sophia revived through computer simulation. *?*, 43(55), 2003.

[8] Anant Baijal and Julia Kim et.al. Composing vibrotactile music: A multisensory experience with the emoti-chair. In *Haptics*. IEEE, 2012.

[9] Christopher L. Bennett. *Digital Audio Theory*. Routledge, 2021.

[10] betterexplained. Intuitive guide to convolution. `https://betterexplained.com/articles/intuitive-convolution/`, 2020.

[11] Barry Blesser and Linda-Ruth Salter. *Spaces Speak, Are You Listening?, Experiencing Aural Architecture*. MIT Press, 2007.

[12] Nico Irawan Bondan Wahyu Pamekas. The influence of motion graphics on user experience in interactive media. In Bondan Wahyu Pamekas, editor, *International Conference of Health, Science and Technology*, 2024.

[13] Natalia A. Bracikowska. Environmental storytelling: The liminal space between embedded and emergent narrative. *Young English Philology Thought and Review*, 6, 2020.

[14] Rob Bridgett. *Leading with Sound: Proactive Sound Practices in Video Game Development*. Routledge, 2021.

[15] Lorna M. Brown, Stephen A. Brewster, and Helen C. Purchase. Tactile crescendos and sforzandos: Applying musical techniques to tactile icon design. In *Human Factors in Computing Systems*, 2006.

[16] Don Carson. Environmental storytelling: Creating immersive 3d worlds using lessons learned from the theme park industry. `https://www.gamedeveloper.com/design/environmental-storytelling-creating-immersive-3d-worlds-using-lessons-learned-from-the-theme-park-industry/`, 2000.

[17] Seungmoon Choi, Yongjae Yoo, and Inwook Hwang. Consonance of vibrotactile chords. *Transactions On Haptics*, 7(1), 2014.

[18] Brent Cowan and Bill Kapralos. Spatial sound for video games and virtual environments utilizing real-time gpu-based convolution. *Future Play*, 2008.

[19] Sarah H. Creem and Dennis R. Proffitt. Defining the cortical visual systems: ”what”, ”where” and ”how”. In *Acta Pychologica*, 2001.

[20] Gary Davis and Ralph Jones. *The Sound Reinforcement Book*. Hal Leonard Publishing, 1987.

[21] Tim Dettmers. Deep learning in a nutshell: Core concepts. `https://developer.nvidia.com/blog/deep-learning-nutshell-core-concepts/`, 2015.

[22] Jacob J. Diffen, Andrew Johnston, and Robert Sazdov. Aural architects: Exploring professional practice in videogame audio. *Explorations in Sonic Cultures*, 2024.

[23] Derek DiFilippo and Dinesh K. Pai. The ahi: An audio and haptic interface for contact interactions. *User Interface Software and Technology*, pages 149–158, 2000.

[24] Brandon Dolinsky. Environmental storytelling in video games. `https://gamedesignskills.com/game-design/environmental-storytelling/`, 2025. Accessed: 2025-05-11.

[25] Actronika Editors. What is under dualsense. https://www.actronika.com/post/whats-under-the-hood-of-the-dualsense, 2024.

[26] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *AES*, 2000.

[27] Michael Filimowicz, editor. *Foundations in Sound Design for Interactive Media: A Multidisciplinary Approach*. Routledge, 2020.

[28] Daniel Fox. *The Modes of Intervention in Alvin Lucier's I Am Sitting in a Room*. PhD thesis, City University of New York, 2020.

[29] Karmen Franinović and Stefania Serafin, editors. *Sonic Interaction Design*. MIT Press, 2013.

[30] Emma Frid and Hans Lindetorp. Haptic music - exploring whole-body vibrations and tactile sound for a multisensory music installation. In *nternational Sound and Music Computing Conference*, 2020.

[31] Laura Frost. *The Problem with Pleasure Modernism and Its Discontents*. Columbia University Press, 2013.

[32] William G. Gardner. 3d audio and acoustic environmental modeling. *Wave Arts*, ?(?), 1999.

[33] Michele Geronazzo and Stefania Serafin, editors. *Sonic Interactionsin Virtual Environments*. Springer, 2023.

[34] Marcello Giordano and Marcelo M. Wanderley. Perceptual and technological issues in the design of vibrotactile-augmented interfaces for music technology and media. *Haptic and Audio Interaction Design*, 8, 2013.

[35] Mark Grimshaw. *The Acoustic Ecology of the First-Person Shooter.* PhD thesis, University of Bolton, 2007.

[36] Giancarlo Grossi. Epistemology of the feelies. *Cinéma and Cie*, 23(40), 2023.

[37] Eric Gunther, Glorianna Davenport, and Sile O'Modhrain. Cutaneous grooves: Composing for the sense of touch. In *New Instruments for Musical Expression*, 2002.

[38] Sabine Haerri and Yvonne Weber. Movement and impact. https://webarchive.ars.electronica.art/festival/2009/humannature/en/index.html%3Fp=2289.html, 2009.

[39] Thomas Hermann, Andy Hunt, and John G. Neuhoff. *The Sonification Handbook*. Logos Verlag Berlin, 2011.

[40] Mark Hollins and Ryan S. Risner. Evidence for the duplex theory of tactile texture perception. *Perception and Psychophysics*, 62(4), 2000.

[41] David Howard and Jamie Angus. *Acoustics and Psychoacoustics*. Focal Press, 2009.

[42] David M. Huber and Philip Williams. *Professional Microphone Techniques*. Primedia, 1998.

[43] Aldous Huxley. *Brave New World*. Chatto and Windus, 1932.

[44] Ali Israr, Seungmoon Choi, and Hong Z. Tan. Mechanical impedance of the hand holding a spherical tool at threshold and suprathreshold stimulation levels. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2007.

[45] Henry Jenkins. Shall we play? `https://henryjenkins.org/blog/2011/05/shall_we_play.html`, 2011.

[46] Henry Jenkins. From transmedia to immersive worlds: An interview with dr. carlos a. scolari on the evolution of media and the future of storytelling. `https://henryjenkins.org/blog/2025/3/2/from-transmedia-to-immersive-worlds-an-interview-with-dr-carlos-a-scolari/`, 2025.

[47] Neil J.Mansfield. *Human Response to Vibration*. CRC Press LLC, 2005.

[48] Martha Joseph. Collecting alvin lucier's i am sitting in a room. `https://www.moma.org/explore/inside_out/2015/01/20/collecting-alvin-luciers-i-am-sitting-in-a-room/`, 2015.

[49] Hanna Järveläinen, Stefano Papetti, and Eric Larrieux. Exploring the effects of additional vibration on the perceived quality of an electric cello. *Vibration*, 7, 2024.

[50] Jochen Kaiser and Marcus J. Naumer, editors. *Multisensory Object Perception in the Primate Brain*. Springer, 2010.

[51] Maria Karam and Carmen Branje. The emoti-chair: An interactive tactile music exhibit. In ?, editor, *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 2010.

[52] Jin Kim and Yung Choi. Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence*, 10(3), 2001.

[53] G. Kress and T. van Leeuwen. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. Oxford University Press, 2001.

[54] Heinrich Kuttruff. *Room Acoustics*. CRC Press, 2016.

[55] Susan J. Lederman. Auditory texture perception. *Perception*, 8:93–103, 1977.

[56] Susan J. Lederman and Roberta L. Klatzky. Haptic perception: A tutorial. *Attention, Perception and Psychophysics*, 71(7):1439–1459, 2009.

[57] Ming C. Lin and Miguel A. Otaduy, editors. *Haptic Rendering Foundations, Algorithms, and Applications*. A K Peters, Ltd., 2008.

[58] Xiulong Liu, Anurag Kumar, and Paul Calamia. Hearing anywhere in any environment. *arXiv*, 2025.

[59] Matthew Lombard and Theresa Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997.

[60] Jack Lowe. Environmental storytelling: A digital frontier for narrative geography. In *Geographies of Interactive Digital Narratives session at RGS-IBG*, 2019.

[61] Alvin Lucier. I am sitting in a room. `http://www.alvin-lucier-film.com/i_am_sitting.html`, 1969.

[62] Alvin Lucier. I am sitting in a room. `https://alvinlucierlovely.bandcamp.com/album/i-am-sitting-in-a-room`, 1990.

[63] Kelly McErlean. *Interactive Narratives and Transmedia Storytelling: Creating Immersive Stories Across New Media Platforms*. Routledge, 1st edition, 2018.

[64] Steven McSeveney and Monica Tamariz et.al. Auditory occlusion based on the human body in the direct sound path: Measured and perceivable effects. *International Audio Mostly Conference: Explorations in Sonic Cultures*, 2024.

[65] David Murphy and Flaithrí Neff. Spatial sound for computer games and virtual reality. *IGI Global*, 2010.

[66] Janet H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. The Free Press, 2016.

[67] Florence S. Nicholls and Michael Cook. Archaeological gameworld affordances: A grounded theory of how players interpret environmental storytelling. In ?, editor, *ACM Conference on Human Factors in Computing System*, 2025.

[68] Ken C. Parsons and Michael J. Griffin. Whole-body vibration perception thresholds. *Journal of Sound and Vibration*, 121(2), 1988.

[69] Bissera V. Pentcheva. Hagia sophia and multisensory aesthetics. *International Center of Medieval Art*, 50(2), 2011.

[70] Justin Peterson and Hyunkook Lee. *3D Audio*. Routledge, 2022.

[71] Dale Purves, George J. Augustine, and David Fitzpatrick. *Neuroscience: The Somatic Sensory System*. Sinauer Associates, 2 edition, 2001.

[72] quoteresearch. Quote origin: Between stimulus and response there is a space. in that space is our power to choose our response. https://quoteinvestigator.com/2018/02/18/response/, 2018.

[73] Catherine L. Reed, Roberta L. Klatzky, and Eric Halgrenc. What vs. where in touch: an fmri study. *NeuroImage*, 25, 2005.

[74] Byron Remache-Vinueza and Andrés Trujillo-León et.al. Audio-tactile rendering: A review on technology and methods to convey musical information through the sense of touch. *Sensors*, 21(6575), 2021.

[75] Manfred Robert Schroeder. New method of measuring reverberation time. *The Journal of the Acoustical Society of America*, 37, 1964.

[76] Russell Leigh Sharman. Mise-en-scene. https://uark.pressbooks.pub/movingpictures/chapter/mise-en-scene/#:~:text=If%20mise,the%20vision%20of%20the%20director/, ?

[77] Yotam Shibolet. Game movement as enactive focalization. *Press Start*, 4(2), 2018.

[78] Jean-Luc Sinclair. *Principles of Game Audio and Sound Design*. Routledge, 2020.

[79] Odeon Room Acoustics Software. Acoustics in ancient church, hagia sophia. https://odeon.dk/the-church-hagia-sophia, 2000.

[80] Axel Stockburger. *The Rendered Arena: Modalities of Space In Video And Computer Games*. PhD thesis, University of the Arts, London, 2006.

[81] Axel Stockburger. Playing the third place: Spatial modalities in contemporary game environments. *International Journal of Performance Arts and Digital Media*, 3(2 - 3), 2007.

[82] JCW Acoustic Supplies. Absorption coefficients of common building materials and finishes. https://www.acoustic-supplies.com/absorption-coefficient-chart/, ?

[83] Lindsay Tarnowetzki. Environmental storytelling and bioshock infinite: Moving from game design to game studies. Master's thesis, Concordia University, 2015.

[84] EMI Archive Trust. Alan blumlein and the invention of stereo. https://www.emiarchivetrust.org/alan-blumlein-and-the-invention-of-stereo/, ?

[85] Jeff Tyson. How playstation 2 works. https://electronics.howstuffworks.com/ps23.htm, October 2000. Accessed: 2025-05-11.

[86] Yon Visell and Jeremy R. Cooperstock. A vibrotactile device for display of virtual ground materials in walking. *Haptics: Perception, Devices and Scenarios*, 5024, 2008.

[87] William A. Yost, Arthur N. Popper, and Richard R. Fay, editors. *Auditory Perception of Sound Sources*. Springer Open, 2008.

[88] Franz Zotter and Matthias Frank. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Open, 2019.

List of Figures

List of Tables

Appendix A

AI Image Generation Model Prompts

Used AI Model: Midjourney 7 Standart

Style Prompt:

- Style: Cinematic realism, 35mm film grain texture, desaturated color palette
- Lighting: Natural overcast lighting with cool temperature bias (5000K-5500K)
- Color Grading: Teal and orange color scheme, lifted shadows, crushed blacks
- Aspect Ratio: 16:9 widescreen format Camera: Medium format aesthetic, shallow depth of field where applicable.

Image 1 Prompt: Wide establishing shot, camera capturing from behind of the young man in dark green fur-lined parka, face not visible, sitting alone on wooden bench placed on the pedestrian lane of massive steel truss bridge architecture with asphalt double lane, overcast gray sky with dramatic cloud formations, murder of crows perched on bridge cables and flying overhead, cold mist rising from water below, 35mm film aesthetic, desaturated colors, shallow depth of field on character, industrial gothic atmosphere.

Image 2 Prompt: Medium shot, sleek dark sedan approaching from left frame, passenger window rolled down showing pitch-black interior, driver's silhouette barely visible, young man in green parka visible in background on bench, steel bridge structure overhead, overcast lighting, cinematic composition, teal-orange color grading, film grain texture.

Image 3 Prompt: Close-up profile shot, young man in green parka leaning into dark car interior, car door open, one foot inside vehicle, hesitant body language, dramatic chiaroscuro lighting, dark interior contrasting with overcast exterior, steel bridge girders visible above, handheld camera feel, shallow focus on character, moody atmosphere.

Image 4 Prompt: Close-up, analog speedometer dashboard with red needle pointing to high RPM zone 5-6, needle in motion blur, reflection of rushing bridge scenery in chrome bezel, dramatic lighting from dashboard illumination, film grain, high contrast, sense of urgency and speed.

Image 5 Prompt: Wide dynamic shot, dark sedan airborne mid-fall from metal constructed bridge, river surface below, dramatic composition, bridge structure silhouetted against gray sky, car captured at moment of weightlessness, cinematic lighting, epic scale showing height and danger, film grain aesthetic.

Image 6 Prompt: Struggling man underwater, drowning.

Image 7 Prompt: Wide shot, vast empty aircraft hangar interior, single figure in distance, harsh industrial lighting, dramatic shadows and highlights, sense of isolation and possibility, warm amber lighting contrasting cool narrative palette, cinematic scale, film texture.

Image 8 Prompt: Top down shot inside boxing ring, warm arena lighting, golden hour color temperature, shallow focus on ring center, energetic and celebratory mood, young man is on the ring as a fighter, arms up, winning pose.

Image 9 Prompt: Surreal wide shot, cosmic wormhole vortex, swirling galaxies and light trails converging toward center, character's silhouette dissolving into infinite spiral, ethereal lighting transition from underwater blue to cosmic purples and golds, peaceful transcendent atmosphere, otherworldly beauty, seamless blend of realism and fantasy.

Image 10 Prompt: A young man floating in the river, only head outside water, breathing mouth open.