CARNEGIE MELLON UNIVERSITY

School of Architecture
College of Fine Arts

Thesis

Submitted in Partial Fulfillment of the requirements for the degree of

Master of Science in Computational Design

TITLE:

Perceive the Spatial Flow of the Canvas —— Enhancing the Experience of Roaming Chinese Landscape Scrolls Using Augmented Audio Reality (AAR)

AUTHOR:

Qilin Wu

ACCEPTED BY ADVISORY COMMITTEE:

Daragh Byrne		May 6th, 2025
Professor NAME	Principal Advisor	DATE
Golan Levin		May 6, 2025
Professor NAME	Advisor	DATE

Perceive the Spatial Flow of the Canvas

—— Enhancing the Experience of Roaming Chinese Landscape Scrolls Using Augmented Audio Reality (AAR)

Qilin Wu

School of Architecture

Carnegie Mellon University

Pittsburgh, Pennsylvania, US

AndrewId: qilinwu

Adviser: Daragh Byrne, Golan Levin

Abstract

Traditional Chinese landscape scrolls embody the concept of ”Wandering/ Roaming”, inviting viewers to immerse themselves in compositions that convey time, movement, and spatial flow. While existing studies explore the integration of traditional aesthetic forms with novel technologies, they often treat these artworks as static, overlooking the spatial-temporal dimensions inherent in Chinese scroll paintings. This research explores how Augmented Audio Reality (AAR) can reinterpret and enhance the immersive experience of roaming scrolls.

The prototype utilizes Ultra-Wideband (UWB) sensors to track the position of the user's headphones within a defined space, capturing real-time movement. An Inertial Measurement Unit (IMU)

monitors tilt and orientation, enabling sound to respond spatially and adapt to user perspective. Unity serves as the development platform, managing spatialized soundscape via the Audio-Source components, corresponding with scroll elements and user interactions.

Spatial-temporal movement analyses and soundscape heatmaps revealed that three categories of spatialized audio prompts most effectively guided attention and deepened engagement with the roaming experience. By merging AAR with traditional art, this framework reimagines cultural artifacts as multisensory, interactive encounters—bridging heritage and contemporary audiences.

Contents

Glossary **6**

1 Introduction: Roaming Canvas **7**

1.1 Long and Narrow Painting Composition — Panoramic 8

1.2 Scattered Perspective — Shifting Viewing Angles 9

1.3 Prompt Path — Narrative 10

1.4 Summary: Needs and Merits 10

1.5 Situating This Thesis 11

1.5.1 Research Question 11

1.5.2 Approach 12

1.5.3 Contribution and Impact 12

1.5.4 Structure of the Thesis 13

2 Background **15**

2.1 Immersive Media and Interactive Design 15

2.2 Spatialized sound and movement in Augmented Audio Reality (AAR) 16

2.3 Mixed Reality, Augmented Reality in Cultural Heritage 20

2.4 Summary of Gaps 22

3 Method: From Scroll to Space **24**

3.1 Interpreting Thousand Li of Rivers and Mountains 26

3.1.1 Why this Scroll 26

3.1.2 Navigating the Scroll as an Experience 27

3.2 Enabling Infrastructure 27

3.2.1 Devices and Tools 27

3.2.2 UWB Sensors Setup 28

3.2.3 IMU Sensor Setup 29

3.2.4 Unity Integration 30

3.3 Scale and Spatial Considerations 35

3.3.1 Scale of Landscape Scroll 35

3.3.2 Scale of Physical Space 36

4 Prototype Development **38**
4.1 Criteria for Designing the Roaming Experience . . . 38
4.1.1 Embodied and Spatially Grounded . . . 38
4.1.2 Open-Ended Paths and Spatial Navigation . . . 38
4.1.3 Temporal Unfolding and Progressive Flow . . . 39
4.1.4 Layered Discovery: Depth, Surprise, and Attention . . . 39
4.2 Design Concept: Interaction Phases for Listening . . . 40
4.3 Interactional Phases . . . 41
4.4 Audio Content . . . 43
4.4.1 Localized Audio Anchors Based on Dynamic Proximity . . . 44
4.4.2 Ephemeral and Evolving Audio Prompts . . . 45
4.4.3 Layered Audio for Interactive Zones . . . 47
4.4.4 Nomadic Exploration Sounds . . . 49
4.4.5 Shifting Sonic Focus . . . 51
4.5 Evaluation . . . 55
4.5.1 Participants and Behavior Data Overview . . . 57
4.5.2 User Test and Evaluation 1 . . . 58
4.5.3 User Test and Evaluation 2 . . . 60
4.5.4 User Test and Evaluation 3 . . . 62
4.5.5 User Test and Evaluation 4 . . . 63
4.5.6 User Test and Evaluation 5 . . . 65
4.5.7 User Test and Evaluation 6 . . . 67
4.5.8 User Test and Evaluation 7 . . . 69
4.5.9 User Test and Evaluation 8 . . . 71
4.5.10 Second Observation and Conclusion . . . 73

5 Findings and Discussion **79**
5.1 Insights on “Flowing Space” . . . 79
5.2 Active Roaming vs. Passive Viewing . . . 80

6 Look Forward **81**
6.1 Limitations . . . 81
6.2 Future Work . . . 82
6.2.1 Multi-User Interaction and Emergent Soundscapes . . . 82
6.2.2 Cultural Specificity vs. Generalizability . . . 82

6.2.3 Expert vs. Lay Engagement 83

Acknowledgements **85**

Glossary

AAR (Augmented Audio Reality) – A technology that overlays spatialized audio onto the real-world environment, enhancing perception and interaction through sound.

Roaming/ Wandering – A mode of engagement in which viewers move fluidly through a space, either physically or perceptually, experiencing an artwork or environment dynamically rather than from a fixed perspective.

Immersive Media – Digital experiences that engage multiple senses, creating a sense of presence through technologies like VR, AR, and spatial audio.

Spatialized Sound – Audio technology that simulates the perception of sound in three-dimensional space, allowing listeners to perceive direction and distance.

Soundscape – The acoustic environment of a space, including natural and artificial sounds, which contributes to the perception and experience of a location.

UWB (Ultra-Wideband) – A wireless communication technology used for precise indoor positioning and tracking due to its high accuracy and low latency.

IMU (Inertial Measurement Unit) – A sensor system that measures motion, orientation, and acceleration using accelerometers and gyroscopes, commonly used in tracking and navigation.

1 Introduction: Roaming Canvas

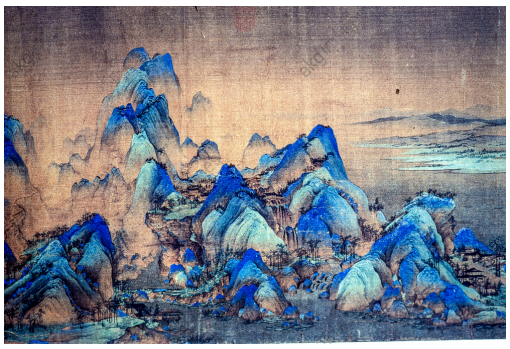


Figure 1: Thousand Miles of Mountains and Rivers, East Asian Scroll Painting [2]

Throughout art history, there has been a transformative shift from depicting patterns to creating immersive, three-dimensional world that invite audience participation, drawing them into the artwork's spatial and sensory dimensions. Flat painting starts to express space. While the perspective method of the Western Renaissance prioritizes individual perspective and a single vanishing point (Figure 2), which brings the picture into focus, in contrast, the Chinese painting tradition focuses on expressing vast and varied spaces and rich scenes in landscape paintings.

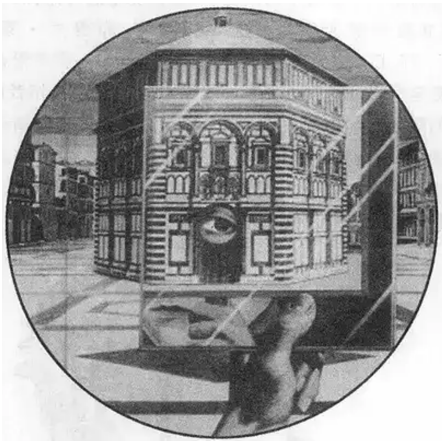


Figure 2: Brunelleschi's perspective device [8]

Eastern paintings, particularly Chinese landscape paintings, offer a distinct wandering spatial experience. Yan Guanghua Gong, a scholar of the Southern Song Dynasty, asserted, "Painting is not just a skill; it should achieve the same essence as the 'Ching' [43]." The Book Ching encapsulates a worldview where the dynamic energy of change is integral, even within the stillness of the hexagrams. By linking the concept of painting to the Ching, Yan underscores that art, like nature, should

transcend fixed representation, becoming a channel for expressing dynamic change and interaction. Therefore, the concepts of ”wandering” and “roaming” are deeply embedded in Chinese landscape painting, inviting the audience to delve into the narrative of the artwork. Instead of simply seeing a direct scene, unfolding the Chinese landscape scroll enables viewers to explore intricate details, akin to roaming within the landscape world itself.

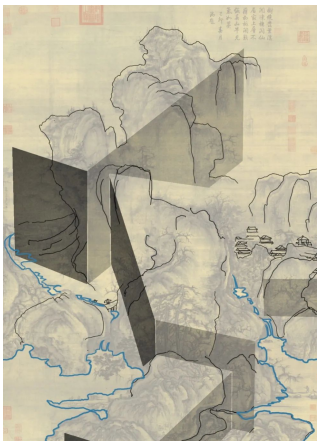


Figure 3: The twisted mountain in ”Early Spring” [42]

1.1 Long and Narrow Painting Composition — Panoramic

The panoramic scroll format embodies a lengthy and slender visual arrangement, creating a panoramic that unfolds gradually over time. Serving as a personal medium, scrolls invite active engagement, allowing the viewer to control the pace of exploration and the sequencing of imagery. This dynamic interaction transforms the act of viewing into an almost cinematic experience, where narrative and spatial depth emerge progressively.

For example, observing “Early Spring” (see Figure 4, left) from bottom to top along the vertical axis mirrors the audience’s gradual upward perspective. Throughout the viewing experience, it feels as though the viewer ascends slowly alongside the upward movement of the mountain. Meanwhile, the painting “A Thousand Li of Rivers and Mountains”(see Figure 4, right), spanning 12 meters in length, guides the viewer’s gaze along winding rivers, distant mountains, and interconnected scenes, as if journeying through an expansive terrain, representing a typical narrative painting and the concept of “travel.”

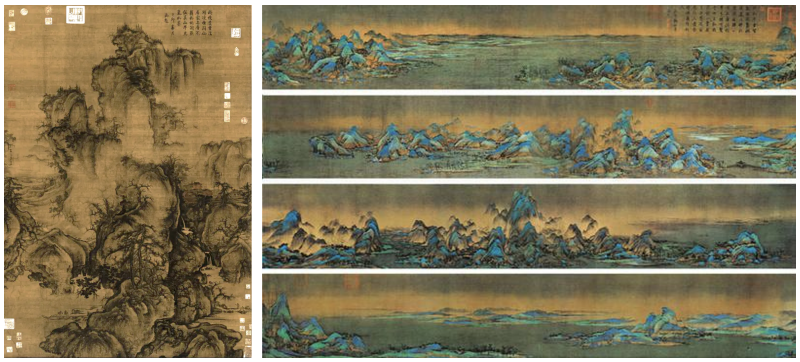


Figure 4: (left) “Early Spring”
(right) Horizontal scroll: A Thousand Li of Rivers and Mountains [1]

1.2 Scattered Perspective — Shifting Viewing Angles

Scattered perspective, also known as multiple perspectives or shifting viewing angles, is a fundamental technique in Chinese painting. Unlike the Western approach (Figure 5, left), which captures a single frozen moment, Chinese painters depict a dynamic accumulation of time and movement within a scene. Different viewpoints of an object are seamlessly integrated into a unified composition, and the spatial transformations are unfolded through lived experience.

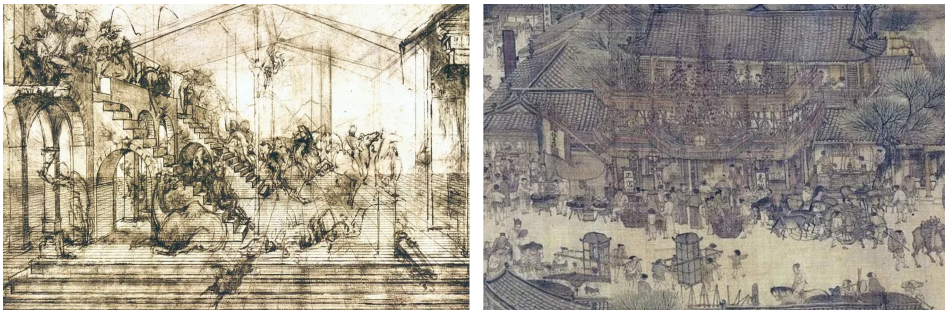


Figure 5: (left) Western focus perspective [5]
(right) Scattered perspective: simultaneous display indoors-outdoors

For instance, mountains might be depicted from a bird’s-eye view, while nearby buildings or figures are shown in an elevation view, blending different spatial depths seamlessly. Similarly, in architectural representations, both sections and elevations, indoor and outdoor scenes (Figure 5, right) can coexist within the same frame, merging different levels of space and organizing different spaces into an overall order.

1.3 Prompt Path — Narrative

Paths in Chinese landscape painting are carefully arranged to evoke the experience of ”travel activities,” linking various landscape nodes and immersing the viewer in the sensation of ”traversing in the painting”. By linking disparate elements, these paths create a journey through the scenes, evoking the tangible experience of ”traveling” in the artwork. Therefore, they not only serve as visual guides, directing the viewer's gaze across the large-scale composition, but also enhance the realism and accessibility [31] of potential explorations within the painting.

As discussed by Nan Zhang and Yuhang Kong [42] in their analysis of Northern Song vertical scrolls (Figure 6), mountain trails often appear as narrow, winding passages weaving through cliffs and rock formations, suggesting an upward journey, while waterways take on meandering, downward flows. This intricate path system weaves together both physical and visual journeys in two directions, creating a dynamic spatial rhythm that enriches the painting's overall structure.

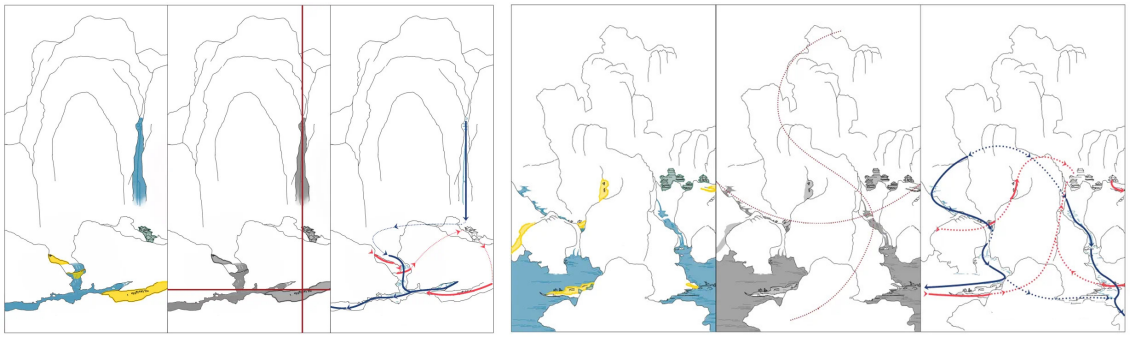


Figure 6: (left) The circular path of “Streams and Mountains”
(right) The double contrasting path of “Early Spring” [42]

1.4 Summary: Needs and Merits

Different media correspond to different painting languages, narrative structures, and viewing methods. The interactive manner in which viewers unfurl traditional Chinese landscape scrolls from left to right hints at the potential to traverse and wander through the space of a Chinese painting. Traditional Chinese landscape scrolls offer rich visual narratives, inviting viewers to immerse themselves in compositions that convey time, movement, and spatial flow. The objective of exploring landscape painting perspectives is to foster a more inclusive spatial discourse and weave thoughts on the way the body “sees” and “listens to” the world. Through this exploration, this research seeks to cultivate a fresh, immersive viewing encounter that transcends time and space.

My research delves into how the body perceives the artwork in a digital age, where the boundaries between physical space and virtual space are increasingly fluid. This allows viewers to "roam" the artwork spaces in an embodied way, exploring the essence of Chinese landscape painting. Mixed Reality (MR), an interactive technology that seamlessly blends physical and digital environments, introduces novel possibilities. Leveraging MR, viewers can actively "roam" within the artwork, deepening their engagement and interaction with these landscapes. To support this, a strong display canvas with the viewer positioned at its center will be established, aiming to revive the flowing spatial immersion inherent in these works. This canvas grants audiences the liberty to wander, roam, and master the operation and rhythm of reading the painting.

Augmented Audio Reality (AAR), a specialized form of MR, further enhances this immersive engagement and introduces auditory layering, shaping a dynamic sonic environment that deepens spatial perception and interaction beyond traditional visual augmentation. By integrating responsive audio elements that adapt to movement and spatial positioning, AAR reimagines traditional Chinese landscape scrolls into participatory experiences, where sound reinforces narrative depth and enriches the sense of presence. This study explores how AAR and MR can elevate the spatial dynamics and narrative depth of Chinese landscape scrolls, creating an immersive medium that bridges tradition and modernity.

1.5 Situating This Thesis

1.5.1 Research Question

This research asks: **How can Augmented Audio Reality (AAR) enhance the spatial flow and immersive experience of Chinese landscape scrolls?** Traditional Chinese scroll paintings are designed for an interactive, unfolding engagement, creating a sense of movement and spatial depth. However, existing AR technologies that support digital museums and immersive exhibitions explore restoring the original proportions of paintings, projecting works in space through digital media, or reconstructing spatial scenes via 3D modeling. These methods diverge significantly from traditional paper media. While these approaches extend accessibility and broaden audience reach, they are limited in spatial solidity. In other words, conveying a deeper, more nuanced sensory engagement inherent in traditional landscape scrolls remains a challenge. This research seeks to bridge that gap by leveraging AAR to create an immersive, participatory experience that aligns with the fundamental principles of Chinese scroll painting.

1.5.2 Approach

This study presents an experimental prototype to explore how Augmented Audio Reality (AAR) enhances the spatial flow and immersive experience of Chinese landscape scrolls. Rather than merely focusing on system implementation, this research seeks to define an interactive roaming model for embodied, nonlinear engagement with the scroll.

Firstly, applying Ultra-Wideband (UWB) and Inertial Measurement Unit (IMU) sensors and the Unity platform to integrate Mixed Reality and spatialized audio introduces three-dimensional landscape representations with a perceptible spatial soundscape, accommodating shifts in the viewer's physical presence. The prototype enables real-time spatial tracking and dynamic auditory interaction; while the user is traversing spatial landscapes, a multisensory sense of continuity should be evoked instead of simply flattened and fragmented frames.

To systematically assess the influence of spatialized audio and interactive movement on user perception, this study establishes four key criteria for design and evaluation, grounded in the spatio-temporal dimensions of the landscape scroll and the "roaming" experience.

Moving beyond system development, this work adopts an empirical approach to record and analyze participants' motion and engagement patterns as they navigate the scroll space. User studies were conducted, capturing spatial trajectories and interaction behaviors within the AAR-enhanced environment. The analysis compares patterns across participants to uncover shared tendencies and divergent behaviors. These findings are interpreted through the lens of the established criteria, which provide a structured framework for understanding how spatialized audio shapes interpretation, movement, and engagement within the roaming experience.

1.5.3 Contribution and Impact

This research redefines the viewing experience of Chinese landscape scrolls by shifting from passive observation to an immersive, multisensory engagement using Mixed Reality (MR) and Augmented Audio Reality (AAR). It challenges conventional approaches to cultural heritage interpretation by developing a structured method for navigating and experiencing long, detailed, and nonlinear artifacts, such as scroll paintings, through spatialized interaction. Key contributions include:

(1) **A Case Study on Spatialized Audio in Scroll Interaction.** This research provides a case study on how spatialized audio enhances the perception of spatial depth and movement within Chinese landscape scrolls. Through AAR-guided navigation, users develop a heightened

sense of scale, continuity, and unfolding composition in an auditory journey, offering insights into how sound can reinforce visual storytelling in nonlinear cultural artifacts. This contribution may be particularly relevant to immersive media designers and interactive storytelling developers seeking new techniques to create dynamic, embodied experiences.

(2) **A Model for Wandering, Embodied Art Interpretation.** By allowing users to physically roam and navigate a virtual Chinese landscape scroll, this project demonstrates a new paradigm for engaging with multi-sensor narratives. The findings contribute to research in immersive media and interactive storytelling by showcasing how sensor-based spatial mapping can create a fluid, dynamic relationship between viewer movement and artistic composition, reinforcing the way to experiencing elongated and sequential artworks, such as handscrolls, manuscripts, and architectural friezes, etc. This approach has the potential to offer new pathways for museum professionals and digital exhibition curators to rethink how audiences engage with large-scale or sequential artworks.

(3) **An Exploration of Cross-Cultural Art Accessibility.** This research highlights how immersive technologies can bridge historical art forms with contemporary audiences, particularly those unfamiliar with the viewing logic of traditional Chinese scrolls. By integrating digital methods with ancient artistic practices, this project examines a new direction for cross-cultural appreciation and accessibility in art interpretation. This may be particularly valuable for cultural heritage institutions and educators looking to broaden engagement with traditional art forms through innovative, interactive platforms.

1.5.4 Structure of the Thesis

After this introduction, Chapter 2 (Background) reviews relevant work in immersive media, spatialized sound, and cultural heritage technologies, particularly in the context of artistic interpretation. It identifies key research gaps and situates this study within existing discourse.

Chapter 3 (Methods) details the technical and conceptual framework, including devices and tools such as Ultra-Wideband (UWB) sensors, Inertial Measurement Units (IMU), and Unity for spatialized audio rendering. It also outlines key design considerations, including scale, landscape dimensions, and spatial positioning.

Chapter 4 (Prototype Development) defines the core design principles shaping the interactive experience, introduces the Criteria for Designing the Roaming Experience, and describes the

conceptual development, interaction phases, audio content, and execution of the prototype. The chapter concludes with an evaluation of its effectiveness.

Chapter 5 (Findings and Discussion) reflects on insights gained through prototyping and testing, exploring how the project reinterprets the concept of "flowing space" and differentiates active versus passive engagement.

2 Background

In recent years, advancements in immersive technologies, such as Mixed Reality (MR) and Augmented Audio Reality (AAR), have opened avenues for enhancing cultural heritage experiences. Traditional Chinese landscape scrolls, featuring the intricate spatial flow and layered visual narratives, provide a ground for such technological interventions. My research focuses on how MR and AAR can be applied to transform the way viewers engage with the ancient art forms—shifting from passive observation to active interaction.

Drawing from fields such as immersive media, cultural heritage technologies, mixed reality, and augmented audio reality, this research will explore the trends within existing research, identifying key gaps and potential contributions.

2.1 Immersive Media and Interactive Design

In the field of immersive media, a significant theme is the use of interactive technologies to enable dynamic real-time responses to user input. Jaimes and Sebe [21] note that multimodal fusion for Human-Computer Interaction allows users to interact with virtual spaces, thereby improving user engagement. However, past research focused on bimodal systems that integrated two modalities, such as gesture and speech (Chun et al.,) [10]. In addition, Chatterjee et al. [9], Pfeuffer et al. [28], and Mohan et al. [24] all explored gaze and gesture techniques, and other research [27,30] encourages users to explore spatial, touchless interactions with a large screen at a distance. However, these systems require users to manually coordinate input modalities, often detracting from the natural flow of interaction, while innovative, are limited in their ability to deliver truly immersive interactions and fall short in seamlessly integrating multiple input modalities to create a fluid user experience.

Daniel Vogel [37] introduced the concept of "implicit interaction" within the framework of interaction phases (Figure 7), and contextual cues such as body orientation or proximity are automatically sensed without the need for conscious coordination. More recent studies have focused on integrating multimodal fusion methods in a seamless and intuitive way. For instance, Hansberger et al. [17] discussed tailored, real-time interaction in virtual reality. Herruzo and Pashenkov [19] explored AI in live performances, delving into the use of AI-driven real-time graphics to transform audience interaction during live performances. These examples demonstrate a broader movement toward active audience participation, reshaping narrative spaces, and creating more engaging, participatory experiences in immersive media.

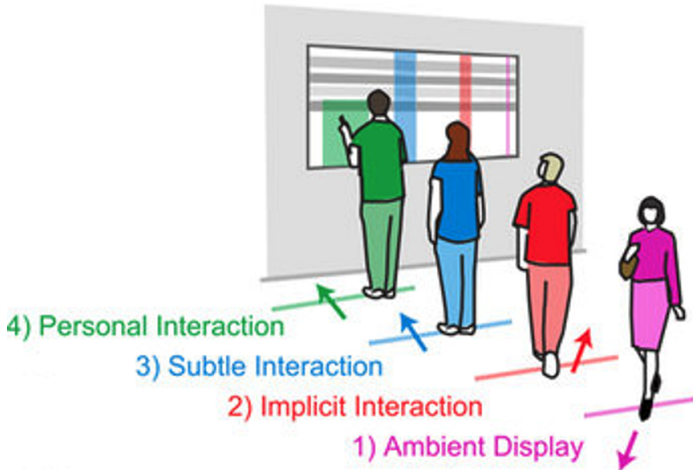


Figure 7: Vogel's interaction phases for ambient public displays, facilitating transitions from implicit to explicit, public to personal interaction [37]

2.2 Spatialized sound and movement in Augmented Audio Reality (AAR)

Augmented Audio Reality (AAR) introduces new dimensions to immersive experiences by blending locative and binaural audio with real-world movement, enhancing the depth of spatial interactions. The development of spatial sound and its conceptual implications signifies a new phase in media history.

While traditional auditory experiences in media have been linear and passive, AAR disrupts this model by transforming sound into an active agent that reshapes perception and interaction within a space. A number of of projects and research efforts explore spatialized sound to reconfigure presence and spatial continuity, altering the listener's perception of place and time.

In the domain of **locative audio walks and binaural experiences**, sound design and spatial interaction converge to construct immersive realities. Janet Cardiff's [34] *Her Long Black Hair* (Figure 8), exemplifies this by entwining spatialized soundscapes with physical movement, reconstructing the listener's sensory relationship with Central Park through layered narratives of history, memory, and fiction, offering participants a sense of mobility as well as a deeper locational belonging. Similarly, in *Alter Bahnhof Video Walk* (Figure 9), Janet Cardiff and George Bures Miller employ binaural audio, ambient sounds, narration (voice) and video to guide participants through an old train station, creating a layered experience where past and present blur into one another. In the project *Rough Mile* [18], audio storytelling fosters a heightened sense of spatial awareness. In Sikora et al.'s archaeological site AAR experience (Figure 10), leveraging spatial audio, the concept is applied

to historical sites to engage users in an auditory reimagining of the past. These works collectively emphasize the potential of locative audio to transform familiar spaces into dynamic, narrative-driven environments, through auditory cues.

Within **institutional and cultural contexts**, previous experiments have showcased the potential for AAR to enrich museum and gallery experiences and audio preservation initiatives. Montan's [25] demonstrates how positioning binaural audio content relative to photographic prints can create immersive and reflective art experiences, rather than static images (Figure 11). Zimmermann and Lorenz's [44] *LISTEN system* (Figure 12) introduces the concept of the "attractor sound", enabling long-range indoor positioning. Building on this, Laurence Cliffe [12] illustrates how spatial audio acts as both a contextual layer and an active agent, enhancing cultural institution interactions. The preservation-oriented project like Unlocking Our Sound Heritage (British Library, 2018-2021, Figure 13) explores how digitizing analog audio materials not only safeguards cultural artifacts but also democratizes access. The evolving role of audio in cultural heritage, showcases its capacity to transform passive spectatorship into participatory engagement.

AAR applications have also extended to the **augmentation of everyday objects and spaces**. As Yang and Mattern [41] and Jing, Frank et al [40] suggests augmenting everyday items with virtual audio sources, which are accurately localized with portable headphones. *Interfacing the Multiplexer Room* [13] (Figure 14) examines the spatial audio in immersive experiences within domestic settings. It underscoring non-visual tracking technologies of ultra-wideband networks to create interactive sound systems respond to the domestic environmental fluctuations. These consider spatial audio as a feasible tool for both functional augmentation and creative expression.

Interactive and serendipitous audio experiences are another focus of the AAR experiments. In Sounding Out the City: Personal Stereos and the Management of Everyday Life [6], Michael Bull challenges visually dominated cultural analyses by proposing an auditory perspective, using ethnographic research on personal stereo use to explore how sensory engagement with mobile audio technology transforms mundane urban experiences into cinematic, serendipitous, and deeply personal narratives. Afterwards, Theodore Watson's Audio Space (Figure 15) explores how user movement actively shapes soundscapes, fostering a playful relationship with auditory environments. Similarly, *Horror-Fi Me*, initially a site-specific installation, was adapted for home use during the COVID-19 pandemic, using mobile AR and binaural spatialized audio to trigger non-looping sounds with continuous ambient elements for a personalized and immersive auditory experience.

Theoretical explorations have further enriched the understanding of AAR's potential. Gaver's [16] work in auditory interfaces and affordances laid early groundwork for non-visual interaction models, while Cardiff and Miller's audio walks elevated locative narrative into an art form. More recently, Cliffe [11] conceptualizes spatial audio as a "third reality," where digital and physical soundscapes merge to enhance immersion. This aligns with theories of embodied interaction, where movement-driven sound design fosters deeper engagement by synchronizing auditory feedback with user actions.

Collectively, these projects and studies underscore the transformative potential of spatial audio in reshaping user interactions and redefining spatial perception. When sound dynamically responds to movement, the AAR transition from a mere augmentation of reality to an active reconfiguration of space, not just extending auditory experience but also critically redefining spatial perception and interaction. AAR enhances sensory awareness and opens new possibilities for immersive embodied engagements.



Figure 8: Her Long Black Hair, Audio Walk with photographs, Janet Cardiff, 2004 [7]



Figure 9: Alter Bahnhof Video Walk 26 min, Cardiff and Miller, 2012 [22]

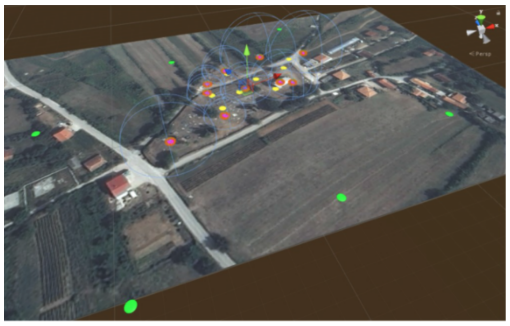


Figure 10: 3D scene of an archaeological site in Soundscaper, 2018 [32]

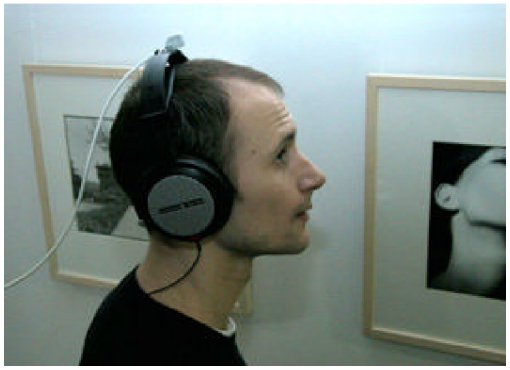


Figure 11: Transforming static images into multisensory wearing AAR, Nils Montan [25]



Figure 12: The LISTEN System Applied at the Kunstmuseum in Bonn [44]



Figure 13: Studio image of unlock our sound heritage, UOSH project, 2018-2021 [20]



Figure 14: Interfacing the Multiplexer Room [13]

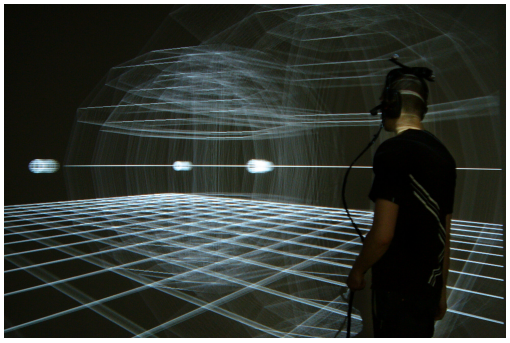


Figure 15: Audio Space is a 3D augmented Installation [38]

2.3 Mixed Reality, Augmented Reality in Cultural Heritage

Cultural heritage encompasses a wide range of artifacts, spaces, and practices, including historical art forms, architectural sites, landscapes, documents, and community traditions. Digital tools, particularly Mixed Reality (MR) and Augmented Reality (AR), have been increasingly employed to preserve, interpret, and engage with these diverse cultural assets. This research specifically focuses

on the role of immersive media in interpreting traditional art forms within the context of cultural heritage.

In the realm of Mixed Reality (MR) and Augmented Reality (AR) applied to historical art forms, the goal is to make them more accessible and engaging for modern audiences. Thiel (2014) [33] asserts that "the institutional walls of the white cube are no longer solid," highlighting how digital interventions challenge the static authority of traditional exhibition spaces - institutional solidity of the "white cube". Media theorist Sabine Breitsameter expands on this, introducing a "sonic understanding of space" [4], where audio-augmented objects extend beyond visual constraints, creating more permeable and interactive environments. These perspectives suggest that MR and AR have the potential to redefine cultural heritage experiences by dissolving spatial and institutional boundaries.

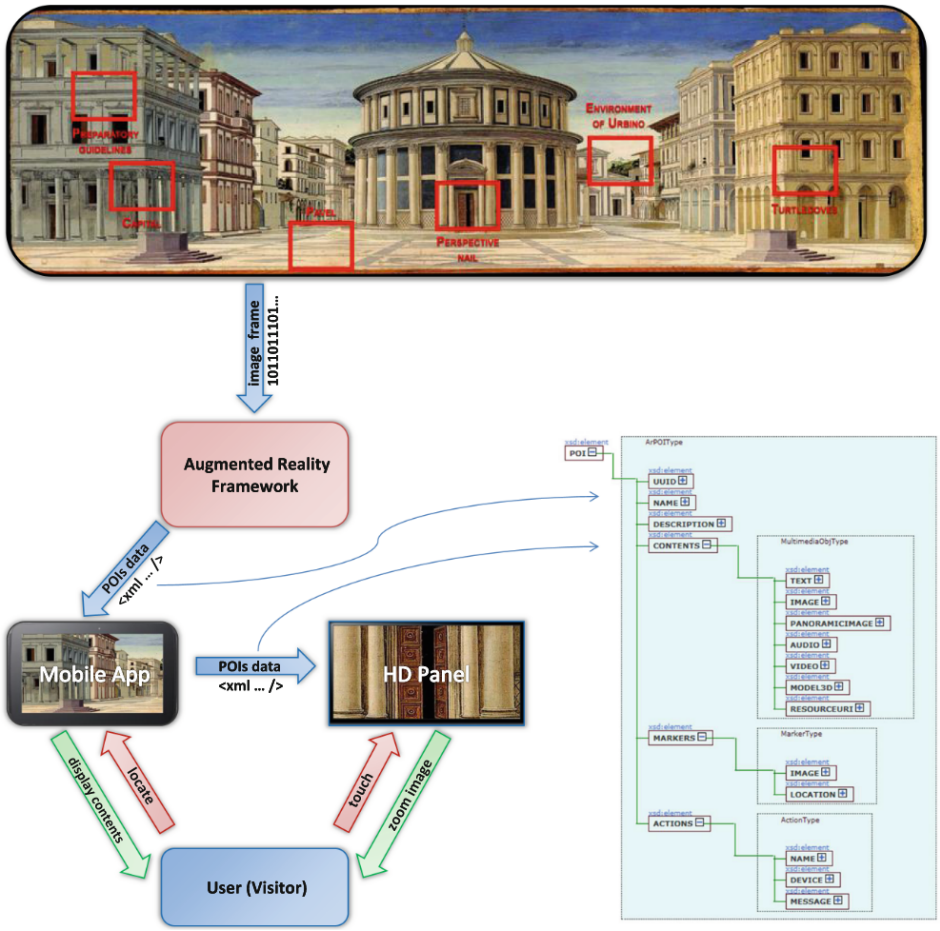


Figure 16: "Citt' a Ideale AR" application(2015, Magnani Palace, Reggio Emilia), matching keypoints to trackables for virtually floating [29]

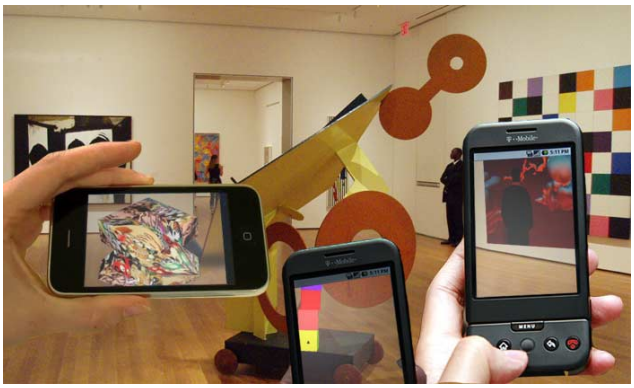


Figure 17: We AR in MoMA (2010) by Manifest.AR at the Museum of Modern Art [39]

In cultural heritage, AR has been widely employed to enhance engagement, overlaying traditional displays with 3D models and interactive elements. Pierdicca et al. (2015) [29] demonstrate this in Figure 16, where AR enriches museum exhibitions through digital augmentations. Similarly, Manifest.AR's 2010 intervention at MoMA (Figure 17) superimposed virtual artworks onto the gallery space via smartphone applications, using GPS and internet connectivity [39] to expand curatorial reach beyond physical boundaries. Building on this, Cliffe (2022) [11] highlights how spatially rendered audio enhances visitor interaction, enabling a deeper and more context-sensitive engagement with exhibits in the future.

However, these innovations pose challenges for curators and museum professionals. Despite these advancements, existing frameworks for implementing Audio Augmented Reality (AAR) are limited and scarce. Unlike visual AR, which has established applications in museums, AAR offers a unique opportunity for multisensory engagement, fostering a dynamic relationship between sound, space, and movement. It's affirmed that AAR interventions in visual art is becoming widespread and has demonstrated its potential to transform cultural heritage interpretation, the integration of Audio Augmented Reality (AAR) remains underdeveloped. There is a need for structured methodologies to integrate spatialized sound into museum settings. Establishing such frameworks would enable cohesive narratives and foster meaningful and deeper visitor interactions.

2.4 Summary of Gaps

(1) Firstly, **Kinesthetic Qualities of Chinese Landscape Scrolls.** A significant gap within the literature is the limited attention given to the kinesthetic qualities of Chinese landscape scrolls, which are meant to be experienced progressively as the scroll unrolls. Unlike Western paintings, where the entire composition is visible at once, Chinese scrolls invite viewers to

journey through their narrative landscapes over time. This requires exploring appropriate mediums that capture the essence of artworks during the process of appreciation. Framing Chinese landscapes in ornate gold frames, as seen in Western galleries, may not be the most suitable presentation method. Although new approaches have been explored, they often fail to originate from the inherent characteristics of the artwork itself, as most current applications treat artworks as static, missing the unique spatial-temporal dimension that Chinese scrolls offer. This research seeks to bridge this gap by utilizing mixed reality (MR) to recreate the sense of movement inherent in Chinese scrolls, enabling viewers to experience the artwork dynamically, moving through its narrative with spatial cues embedded both visually and aurally.

(2) Secondly, **Underexplored Role of Augmented Audio Reality (AAR).** Augmented Audio Reality (AAR) is relatively underexplored in the context of traditional art forms. While visual augmentation through AR is prevalent, the use of spatial audio as a means of enhancing user interaction has not received enough attention. By incorporating AAR into the viewing experience of Chinese scrolls, this research offers an immersive and holistic sensory engagement. This research incorporates AAR, addressing this unexplored aspect of augmented reality in cultural heritage. The convergence of sound and movement with AAR provides additional layers of meaning through spatialized sound, allowing the viewer to not only observe but also ”hear.” This research focuses on spatialized sound and movement in AAR to enhance the immersion of spatial experiences. The concept of spatial flow is central to experiencing Chinese scrolls, where the viewer moves through the depicted landscape, journeying and unfolding the scrolls. Contextualizing auditory cues aligns with the spatial flow to deepen the cultural narrative.

3 Method: From Scroll to Space

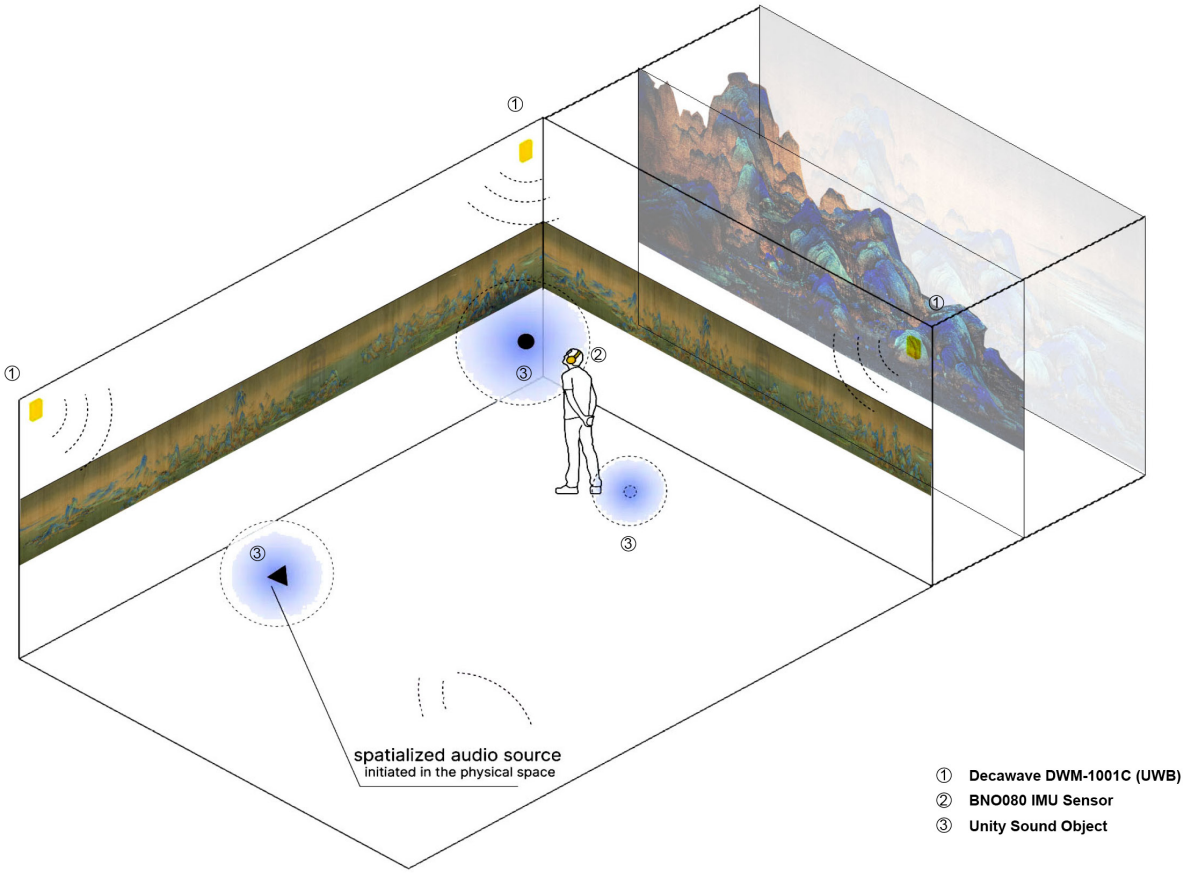


Figure 18: Prototype Interaction Diagram, drawn by the author

By integrating spatialized audio cues and sensor-based movement tracking, this research seeks to extend the traditional experience of scroll viewing by transforming it into an immersive, interactive engagement. Through real-time positional shifts, the system links environmental sounds to visual elements, and the auditory dimension enhances spatial perception, reinforcing the sense of movement and presence within the scroll. This chapter will detail the selection of the panoramic landscape, the implementation of sensor-based tracking systems, and the methods used to map the physical scale of the scroll into Mixed Reality (MR) space (overviewed in Figure 18).

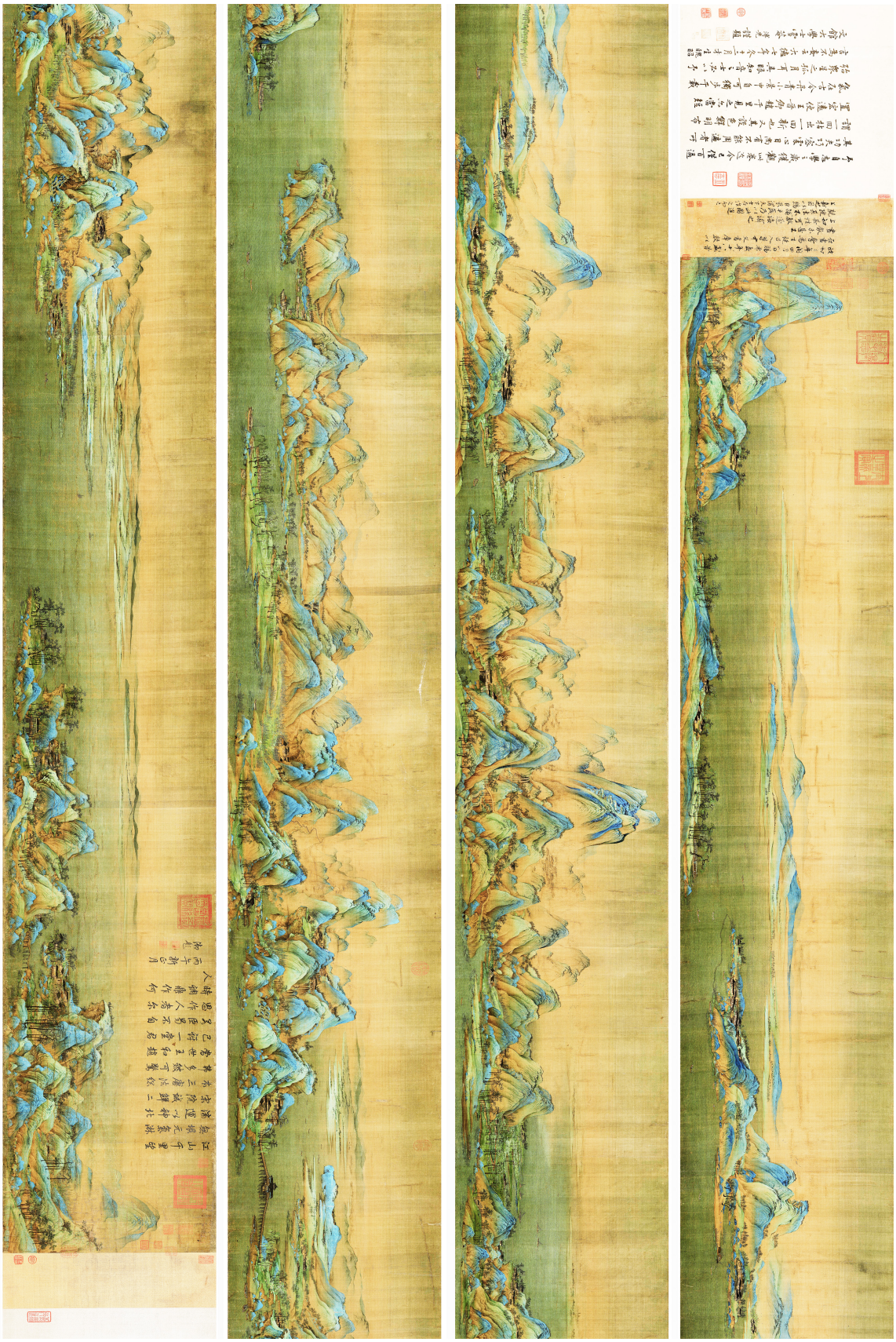


Figure 19: Thousand Li of Rivers and Mountains (1113 C.E., 20.3 × 469.1 in. [1]

3.1 Interpreting Thousand Li of Rivers and Mountains

3.1.1 Why this Scroll

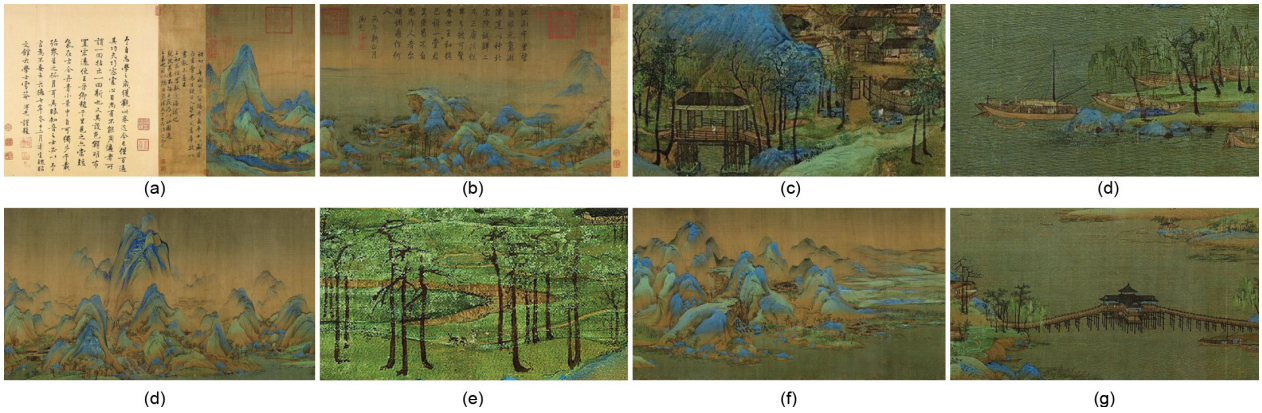


Figure 20: *A Thousand Li of Rivers and Mountains*, a silk scroll painting (20.3 × 469.1 in.). [36]

Thousand Li of Rivers and Mountains (1113 C.E., 20.3 × 469.1 in., Figure 19) is a seminal masterpiece of landscape painting from the Northern Song dynasty, characterized by mineral pigments of azurite and malachite, created by the court painter Wang Ximeng at the age of 18 years and is currently housed at the Palace Museum in Beijing.

Unlike traditional handscrolls that serve as static depictions of nature, Thousand Li of Rivers and Mountains is designed to be "traveled" rather than merely viewed. The monumental scroll, measuring over 12 meters in length, is structured into six interconnected landscape segments [1], each dominated by towering mountain forms. These sections are seamlessly linked through bridges, waterways, and winding pathways, fostering an unfolding, roaming experience that encourages immersive, nonlinear engagement.

Beyond its visual grandeur, Thousand Li of Rivers and Mountains is imbued with a sense of movement and narrative continuity, as highlighted in historical commentaries [36]. The Qing dynasty Emperor Qianlong, in his poetic inscription, praised the work's ability to capture the boundless spirit of nature, while Song dynasty calligrapher Cai Jing noted how the painting was a direct product of the imperial painting academy's refined techniques. Later scholars, including Yuan dynasty monk Pu Guang, emphasized how each viewing reveals new layers of discovery. These inscriptions and colophons, borne by the scroll itself (Figure 20 (a)(b)), reinforce the idea that the scroll is meant to be explored repeatedly.

3.1.2 Navigating the Scroll as an Experience

The painting employs "scattered perspective" (see section 1.2), allowing multiple vantage points to coexist within a continuous landscape, enabling a fluid transition between various scenes. Rather than observing from a fixed position, the viewer moves through the painting, guided by its visual cues. Several *structural elements* contribute to this immersive quality. As illustrated in Figure 20 (c)-(g), pathways and bridges serve as navigational guides, directing the viewer's gaze and creating rhythmic spatial transitions between mountains, rivers, and human settlements. Water and mist enhance atmospheric depth, softening the boundaries between land and sky while reinforcing a sense of distance and gradual discovery. Meanwhile, mountain forms and layering, achieved through high distance, deep distance, and level distance compositions, heighten the illusion of vast terrain, further encouraging the sensation of continuous exploration.

This expansive spatial composition makes Thousand Li of Rivers and Mountains particularly suitable for Mixed Reality (MR) reinterpretation, as it inherently aligns with the concept of navigable digital spaces. Through the integration of real-time spatial tracking and augmented audio layers, this research explores how the painting's intrinsic sense of movement can be preserved and enhanced in a multisensory digital environment.

3.2 Enabling Infrastructure

Building on the exploration of *Thousand Li of Rivers and Mountains* as a roaming experience, the next step is to establish the technological foundation that brings the artwork to life in an interactive form. This section describes the technological framework required to create an immersive augmented audio-visual experience. It covers the devices and tools used for spatialized audio rendering, the technical pipeline, the setup and integration of Ultra-Wideband (UWB) and Inertial Measurement Unit (IMU) sensors, and the Unity-based implementation. These components work together to adapt audio elements to the user's movement and orientation, ensuring a fluid and responsive mixed-reality environment.

3.2.1 Devices and Tools

To achieve this responsive mixed-reality environment, the research combines Ultra-Wideband (UWB) and Inertial Measurement Unit (IMU) sensors, integrated into Unity for real-time spatial tracking and spatialized soundscape rendering. Specifically, the UWB is used for precise spatial

positioning, the IMU for detecting head tilt and orientation, and Unity to create spatialized sound objects that correspond to elements in the landscape scroll (refer to Figure 18).

- **Ultra-Wideband (UWB)** uses high-frequency radio signals to calculate the distance between two points with an accuracy of centimeters. In this prototype, UWB sensors are used to track the position of the user's headphones or headgear within a defined space, detecting the user's movement in real-time. This allows the system to adapt the audio experience by adjusting the spatial positioning and intensity of the sound objects based on the user's location within the workspace. For example, as the user approaches a river scene or a distant mountain in the scroll, the system can adjust the volume, pitch, and direction of corresponding sounds for immersion.

- **Inertial Measurement Unit (IMU)** measures the tilt and orientation of the user's head or headphones, thus enabling more dynamic audio experiences. For example, when the user tilts their head to the left or right, the system can simulate the audio perspective of that movement, as though they are looking directly at a sound source, such as a waterfall. The IMU data complements the spatial positioning provided by the UWB sensor, allowing the sound to be both spatially accurate and orientation-dependent.

- **Unity** serves as the platform for creating and managing the mixed reality environment and integrating sound. Using Unity's AudioSource component, spatialized sound objects are created to correspond to the scroll elements (e.g. wind, water, birds, and distant human activities). Unity's 3D audio features allow for positional audio, where the volume, pitch, and direction of sound objects are changeable, providing the opportunity for spatial sound interaction with the user's location and orientation. The audio objects are placed within the digital recreation of the landscape in Unity, each linked to specific locations that correspond to landscape elements and cues within the scroll.

3.2.2 UWB Sensors Setup

To create a spatially aware system for detecting user position and movement within a designated physical space, the prototype must prepare Ultra-Wideband (UWB) sensors.

Setting up the UWB system needs eight UWB devices:

- 6 Beacons — These will be placed at specific positions on the walls of the space to establish a reliable spatial grid as a coordinate system. One beacon must be designated as the initiator, acting as the reference point (0, 0, 0).

- 1 Sender — This is placed on the user's equipment (e.g., the headphones or headgear). It transmits the coordinates of its current position within the space to the receiver in real time.
- 1 Receiver — This is used to receive the data from the sender, proceeding it to the PC with Arduino system to post-process the data, and feed to the Unity-based sound system for adjusting the audio and virtual elements accordingly.

Additionally, the DRTLS app (used with Decawave UWB devices) on a tablet is used to configure and manage this system (Figure 21). It helps set up the UWBs, visualize the coordinate system, and verify communication between all devices.

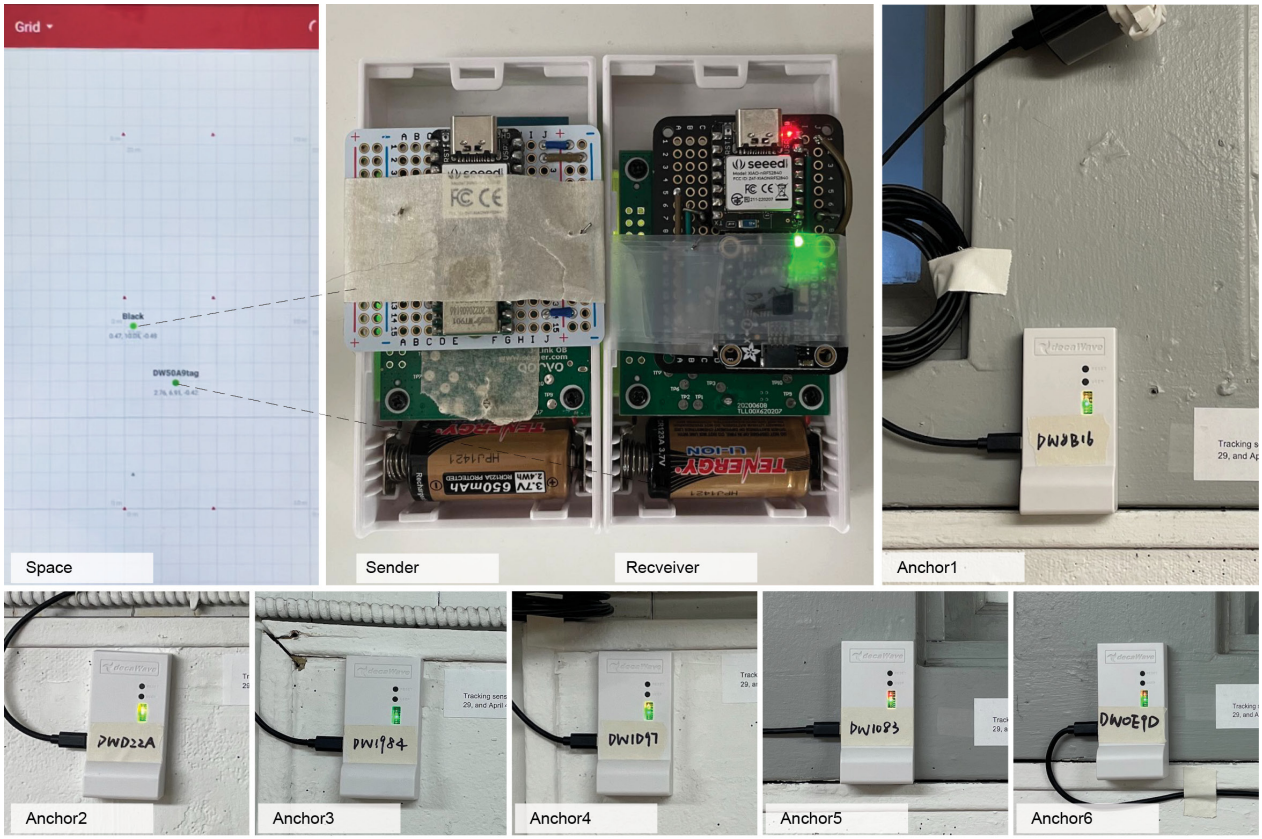


Figure 21: UWB and IMU Sensor Setup Diagram, photo by the author

3.2.3 IMU Sensor Setup

In addition to UWB, Inertial Measurement Unit (IMU) sensors allow the system to respond to more micro-movements of the user's head. The accelerometer detects the tilt of the head along different axes (x, y, z), while the gyroscope measures rotational movements (angular velocity).

- Calibration: Before use, the IMU sensor must set the initial orientation of the user's head as a reference point and ensure that it aligns correctly with the UWB network coordinate system.
- Placement: To ensure it remains in a fixed position relative to the user's head, the IMU sensor is attached to the user's headgear or headphones.
- Data Capture: The IMU continuously captures orientation data in real-time, which is processed by the platform (e.g., Arduino or PC) and integrated into the Unity environment.

3.2.4 Unity Integration

This section explores the integration of physical sensor data processing with virtual audio generation in Unity, detailing how user movement and orientation dynamically influence the spatial sound environment. It studies the interaction between the streaming tracking data from the UWB and IMU sensors and the Unity audio rendering system.

- **Mapping Real-World Space and Real-Time Data to Unity**
  A 1:1 scale virtual room is created in Unity to match the physical environment, ensuring accurate spatial mapping. In this setup:
  - The **X-axis** corresponds to the scroll's *lengthwise direction* (unrolling direction).
  - The **Y-axis** represents the *vertical spatial height* in the physical space.
  - The **Z-axis** maps the *distance between the user and the scroll* (depth).

  A python script is used to parse, filter, and fuse the raw data received from sensors. This script processes the incoming data stream from the sensor via the PC port and sends the cleaned positioning and orientation data to Unity via **OSC (Open Sound Control) messages** [15]. For instance, the UWB system continuously tracks the user's real-world position, transmitting physical X, Y, and Z coordinates, which Unity then interprets to place the user within the 3D virtual space.

  Afterwards, the Real-Time Visualization includes:

  (i) Unity receives parsed OSC messages through an OSC Receiver.

  (ii) A **GameObject (user avatar)** is assigned to represent the tracked user in the Unity.

  (iii) The UWB and IMU sensor data are applied to dynamically update the user's position and orientation in real time.

- **Sound Object and Virtual Soundscape Creation**

AudioSource Property	Description
Audio Clip	Reference to the sound clip file that will be played.
Play On Awake	When enabled, the sound starts playing automatically when the scene loads. If disabled, playback must be manually triggered using the *Play()* function in scripting.
Loop	Determines whether the audio clip repeats continuously when it reaches the end.
Spatial Blend	Sets how much the 3D engine has an effect on the audio source.
Volume	Controls the loudness of the audio at a distance of one world unit (one meter) from the *AudioListener*.
Pitch	Adjusts playback speed and pitch. A value of 1 represents normal playback speed, while values above or below 1 increase or decrease the pitch, respectively.
Min Distance	Defines the range within which the sound is heard at full volume. Outside this distance, attenuation begins based on the rolloff settings.
Max Distance	Sets the distance beyond which the sound will no longer attenuate, maintaining a constant volume.

Table 1: Sample Audio Source Settings Description [35]



Figure 22: Left: Adjust curves on audio sources to impact the amount of volume, spatial blend, spread, or reverb based on the distance between the object's center and the player. Unity allows these same curves to control its low-pass filter frequency if you add the low-pass effect; Right: A large "room" of reverb [http://www.tcwav.com/ARAudioGuide.pdf]

Each feature element is paired with a sound object in Unity. Each sound object is linked to an **Unity AudioSource** [35] component, where attributes such as 3D spatialization, attenuation, and rolloff are controlled. The initial positions of the sound objects are configured based on the relative positions of elements in the landscape painting, respecting the authentic spatial layout.

Unity AudioSource component provides several key properties (see Table 1) that influence how sound is played and perceived in the virtual environment. These properties provide precise control over audio behavior, ensuring a responsive and immersive sound environment that dynamically adapts to the user's movement and spatial positioning.

Within this framework, landscape elements such as waterfalls, birds, and wind interact seamlessly with the user's real-time actions and head orientation. For example, as the user approaches the edge of a cliff, the rush of wind intensifies, accompanied by the distant call of an eagle, reinforcing the sensation of height and openness. Similarly, turning toward a bamboo grove reveals the gentle creaking of swaying stems, while moving closer to a riverbank unveils the subtle splashes of fish jumping, encouraging deeper engagement. By integrating spatialized audio techniques, this dynamic soundscape enhances the perception of depth, movement, and presence, transforming the scroll into an interactive and multisensory experience.

As part of the sound asset collection, this project gathered a total of 74 audio clips (Table 2), featuring both life activities and ambient sounds. The collection includes over ten different categories, such as bird calls, livestock, footsteps, insects, waterfalls, mountains, wind sounds, etc., with durations ranging from 2 seconds to more than 1 hour.

These sound assets were primarily sourced from three open-access sound effect platforms: Artlist [3], Mixkit [23], Nature Soundmap [14]. Among them, Nature Soundmap is a geo-based audio library, allowing the download of environmental recordings captured in the mountainous regions of Northwestern China. However, due to the limited availability of recordings from this region, additional nature field recordings from Artlist and Mixkit were supplemented. This set of audio assets lays a foundation for integrating realistic sound into the Unity Audio Source.

	Audio Clip	**Number and Duration**
Life Activity	- Birds singing and chirping	- 13 clips, 28s–5m
	- Footstep	- 8 clips, 24s–32s
	- Cow dry moo	- 5 clips, 2s–6s
	- Water wheel	- 2 clips, 52s–2m32s
	- Insects and frog calling	- 6 clips 48s-1m43s
	- Ancient school	- 2 clips, 27s–1m35s
	- People talking and laughing	- 4 clips, 1min15s–1m44s
	- etc.	
Ambiences	- Waterfall and water over rocks	- 12 clips, 1m38s–3m34s
	- Breaking twigs and leaves	- 4 clips, 3s–23s
	- Swamp loop and water lapping	- 3 clips, 43s–1m
	- Wind and mountain ridge	- 11 clips, 2m57s–1h4m32s
	- River and Lake	- 7 clips, 51s–1h3m25s
	- Distance flute	- 3 clips 2m28s-3m2s
	- etc.	

Table 2: Sample audio assets collected from Artlist [3], Mixkit [23], Nature Soundmap [14]

- **Audio Control and Synchronization**

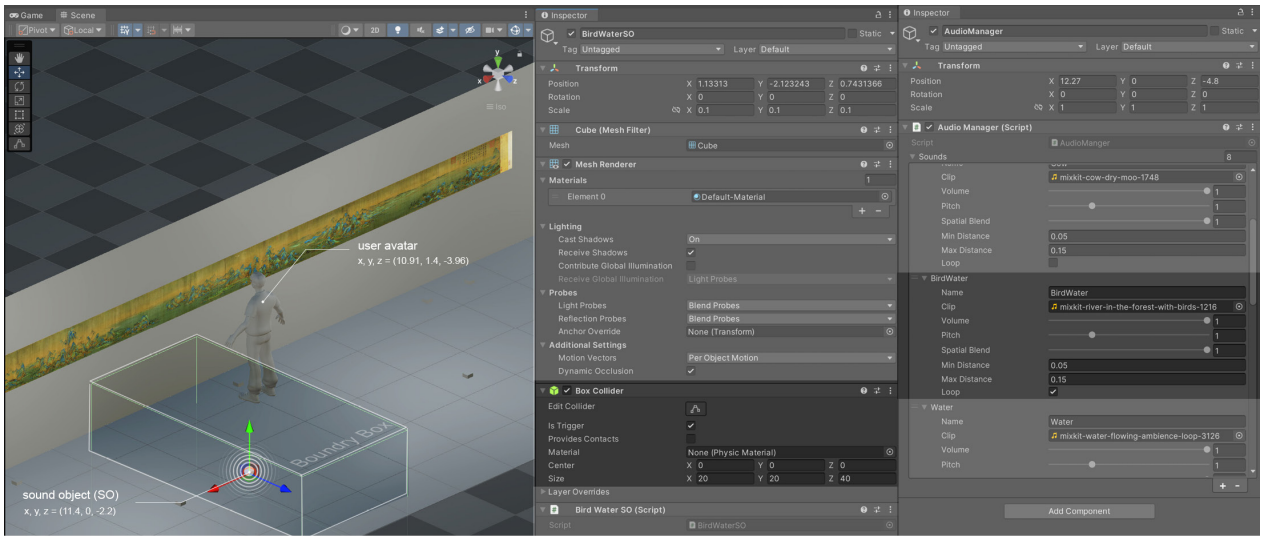


Figure 23: Sound object collision property to trigger audio playing, drawn by the author

Unity's coordinate system is used to map the user's real-time position to specific landscape elements within the digital scroll environment. Key features such as waterfalls, birds, wind, and people are positioned within Unity's 3D world space to align with their corresponding locations in the original landscape painting.

For example, if a bird SO (Sound Object) in the painting is assigned a specific 3D coordinate (e.g., X=11.4, Y=0, Z=2.2) in Unity, the system ensures that as the user moves through the space, the soundscape responds dynamically, reinforcing the sense of presence within the interactive environment.

In Unity, each Sound Object (SO) is assigned a collision property (Figure 23), enabling interactive audio triggering. When the user avatar enters the sound object's boundary box, the audio source is activated, and the corresponding sound begins to play. Conversely, as the user moves outside the boundary, the sound stops, creating a responsive auditory experience.

Additionally, the boundary box leverages Unity's 3D audio properties (Table 1), particularly Spatial Blend, to enhance realistic sound positioning. Each sound's radius and attenuation settings are carefully calibrated to create a natural sound decay effect, ensuring that the

perceived volume adjusts dynamically with the user's position and movement. The system uses Min Distance and Max Distance parameters to define the audible range, allowing spatialized audio to fade as the user moves farther from the source.

3.3 Scale and Spatial Considerations

3.3.1 Scale of Landscape Scroll

Translating the scroll's real-world dimensions into the Unity virtual environment requires establishing a consistent unit of measurement in Unity that corresponds to the scroll's physical proportions. The original Chinese landscape scroll, *A Thousand Li of Rivers and Mountains*, selected for this prototype measures 20 inches in height and 469 inches in width. Two approaches were considered:

- Option 1: Projecting a section of the scroll across four entire walls to create a fully immersive experience. This approach magnifies the artwork, enhancing visual detail and creating a striking, large-scale presentation.
- Option 2: Displaying the scroll at its original dimensions, printed and placed around the room. This option recreates the traditional viewing experience, allowing viewers to move around the scroll as they would in a gallery.

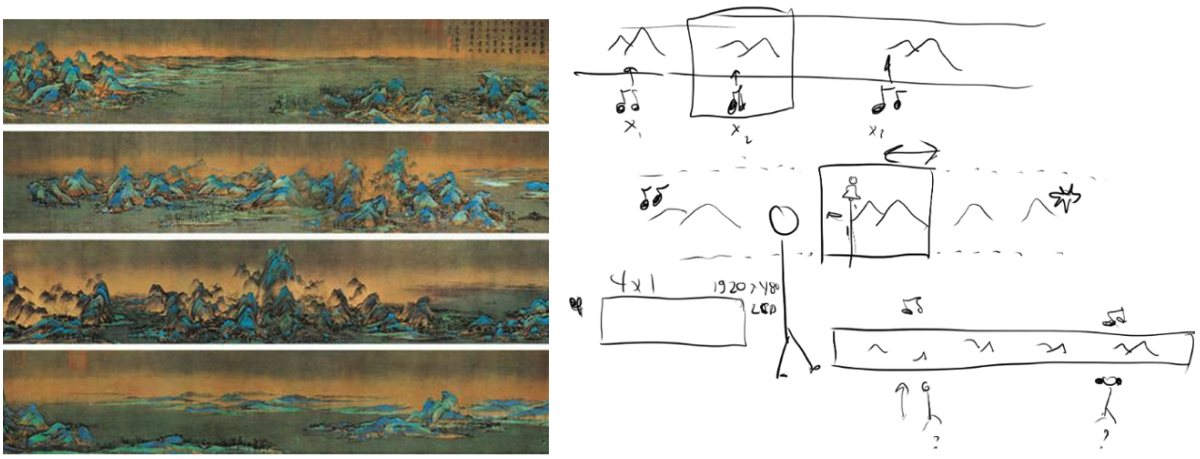


Figure 24: Landscape scale design,
A Thousand Li of Rivers and Mountains, 20 inches in height and 469 inches, The Palace Museum [26]

After consulting with professors and experts, I opted for Option 2, prioritizing respect for the artwork's original dimensions and its alignment with how the scroll would traditionally be exhibited.

This decision ensures that the prototype experience closely mirrors real-world gallery settings, where the scale of the artwork and movement-based engagement are integral to its appreciation. By maintaining the authentic proportions of the scroll, this research focuses on enhancing spatial flow and auditory immersion within traditional Chinese landscape paintings. This approach allows users to engage with the soundscape of the scroll in an environment that honors its historical scale, reflecting on the critical practice of recontextualizing traditional art and improving realism and preserving cultural integrity.

3.3.2 Scale of Physical Space



Figure 25: *Loge space*, College of Fine Arts floor 4, photo by the author

The selected physical space, *Loge Space* (Figure 25), measuring 24.54 × 4.8 × 3 meters, provides a sufficiently long continuous wall to accommodate the landscape scroll, allowing users to gradually unfold and experience the painting in a seamless and immersive manner. Located on the top floor of the Carnegie Mellon University College of Fine Arts, this linear exhibition space is typically used for student critiques, installations, and exhibitions. Its elongated structure naturally guides the

movement along the scroll, reinforcing the intended navigational viewing experience. Additionally, the gallery-like lighting and display-oriented design ensure optimal visual clarity and spatial immersion. The controlled indoor environment minimizes external lighting inconsistencies, preventing glare or uneven illumination that could disrupt the interactive experience. Given its dedicated function as a presentation space, the room serves as an ideal setting for integrating Mixed Reality (MR) elements, transforming traditional scroll viewing into an interactive, user-driven encounter.

To enable precise spatial tracking, a total of six Ultra-Wideband (UWB) sensors are deployed, with three positioned along each of the two long walls. UWB is well suited for this project due to its centimeter-level accuracy and low latency, ensuring seamless synchronization between user movement and spatialized audio. Unlike traditional positioning systems, UWB excels in dynamic MR environments, where precise localization directly enhances the immersive soundscape. Furthermore, UWB is highly effective in indoor spaces, particularly in small to mid-sized environments, making it an ideal choice for the dimensions of this room. The deployment of six sensors provides sufficient coverage and tracking accuracy, and its mobility and scalability make it a practical solution for crafting a highly interactive and immersive audio-visual experience.

4 Prototype Development

This chapter describes the development of our Augmented Audio Reality (AAR) prototype and the design of the interactive roaming experience. We first establish the core criteria guiding our approach (Section 4.1), then outline the conceptual framework (Section 4.2) and detail the defined interactional phases (Section 4.3). Next, we explain the audio content design and soundscape mapping strategies (Section 4.4), and finally present our evaluation methodology and behavioral analyses (Section 4.5). Together, these sections translate the theoretical insights from earlier chapters into a concrete, embodied prototype.

4.1 Criteria for Designing the Roaming Experience

Drawing from the spatial and temporal dynamics of Chinese landscape scrolls and insights from past work in Mixed Reality (MR) and Augmented Audio Reality (AAR) within media and cultural contexts, I have developed a set of criteria for shaping the ”roaming” experience within a flowing canvas. These criteria synthesize core strategies inherent in the scroll format while aligning with my approach to interaction design and embodied engagement.

Rather than serving as rigid guidelines, these criteria function as a fluid framework for evaluation — a structure that shapes design decisions while remaining open to iteration. By applying them to my interventions, I continuously assess their effectiveness, refining them to ensure they align with the project's goals and the evolving nature of the experience.

4.1.1 Embodied and Spatially Grounded

The experience must be felt through the body, reinforcing the act of wandering. Movement should not be an abstract input but a direct interaction with space and sound. The system should dynamically respond to the act of walking, pausing, turning, and tilting — the body's position relative to the scroll and the ears' relationship to spatialized sound sources, allowing allowing users to navigate the artwork as an extension of the body's presence. Success is measured by whether users feel physically immersed in the landscape, rather than detached observers.

4.1.2 Open-Ended Paths and Spatial Navigation

Soundscapes and spatial cues should dynamically respond to movement, allowing actively shaping the user's journey rather than passively following a predetermined course. By integrating sound

elements that emerge or evolve based on proximity and motion, the interactive space fosters organic, non-linear navigation, allowing users to craft their own auditory experience rather than following a fixed route. Success is measured by path diversity and meandering patterns — whether users explore unpredictably, engage with different areas instead of moving in a straight line.

4.1.3 Temporal Unfolding and Progressive Flow

Much like unrolling a scroll, the experience should evolve not only spatially but also temporally. The soundscape should reveal itself progressively, with momentary pauses, redirections, and transitions shaping the pacing and rhythm of the canvas flow in a natural and intuitive manner, enhancing continuity rather than causing disruption. Success is measured by whether the system enables fluid movement, avoiding abrupt or disconnected jumps while compelling users to stay within the soundscape, where details emerge gradually and encourage prolonged engagement.

4.1.4 Layered Discovery: Depth, Surprise, and Attention

Much like the shifting perspectives and layered narratives within a scroll, the experience should encourage exploration through depth and perspective shifts. Certain interactions should remain hidden until the user pauses, reorients, or actively listens, while some discoveries should be perceived under specific conditions, creating an encounter of surprise and depth. Success is measured by repeat engagement, exploratory behavior, and moments of wonder — how effectively the system sustains attention, sparks curiosity, and encourages users to revisit, linger, and uncover new nuances rather than experiencing everything at a surface level in a single pass.

4.2 Design Concept: Interaction Phases for Listening

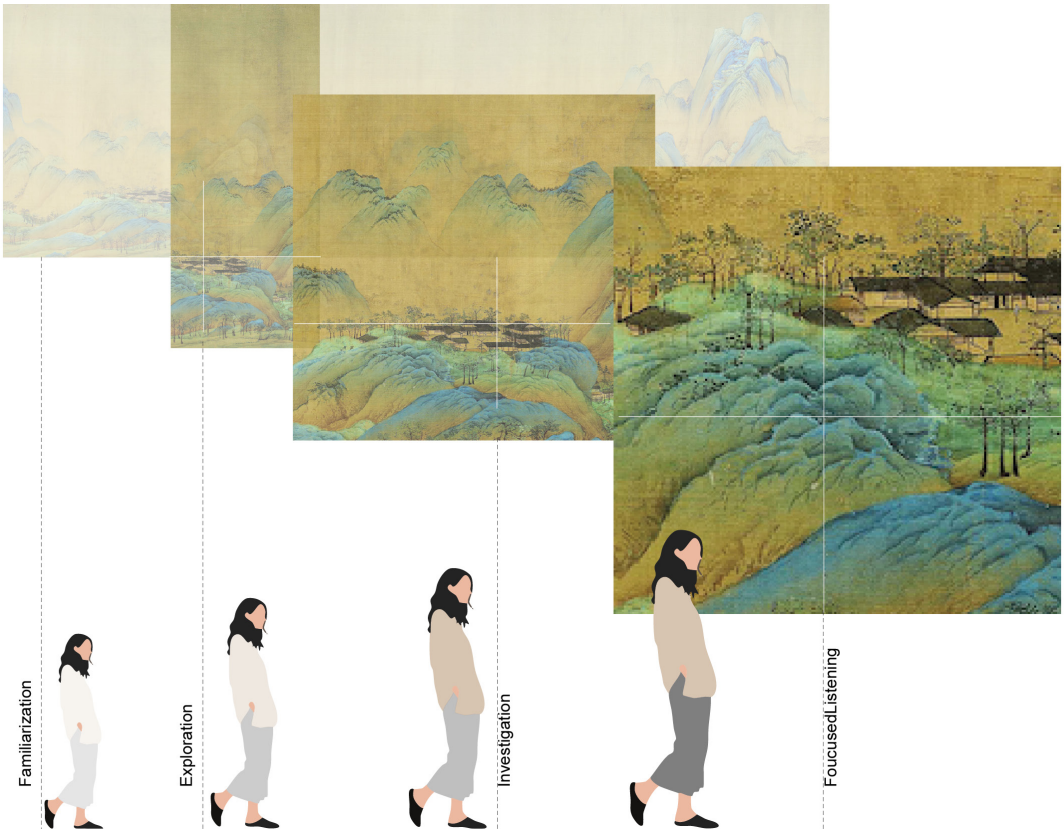


Figure 26: The distance influences interactional phases, drawn by the author

Each interaction phase of the MR experience will incorporate distinct audio elements to enhance immersion and engagement. The early phase with spatialized audio, help users understand and get familiar with how movement affects auditory experiences. Investigative sounds will provide contextual cues that prompt users to explore and uncover hidden elements within the painting. Spectacle audio will feature engaging sounds designed to inspire curiosity and invite participation from onlookers. Finally, environmental and micro sounds, such as birdsong or the gentle flow of a river, will create an ambient soundscape, enriched by subtle micro-sound layers that deepen the sense of immersion as users focus on intricate details. As time progresses, the soundscape evolves, and the user's interaction with it changes autonomously.

This AAR interaction framework builds upon Daniel Vogel's [37] concept that distance influences interactional phases (Figure7). This principle is further developed in the AAR interaction scene (Figure 26), demonstrating how spatial proximity and movement shape the evolving auditory experience in the MR environment.

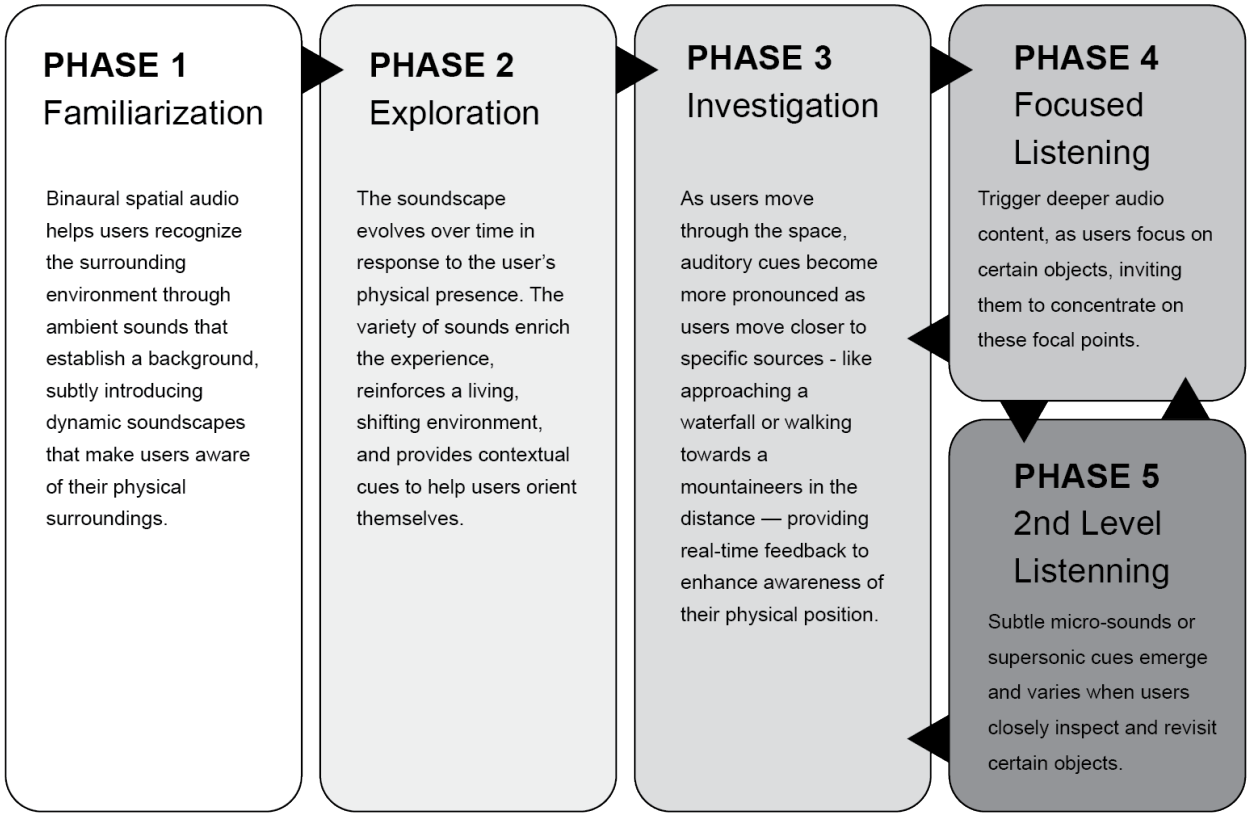


Figure 27: The identified sequence of interactional phases (refer to the McMichael Experiment [11]), drawn by the author

4.3 Interactional Phases

To inform the design further, the concept of interactional phases was introduced—an approach inspired by insights from the McMichael Experiment sound installation, as documented in Cliffe's study [11]. It demonstrated how spatialized audio can structure user engagement through distinct phases. In this context, audio serves both as a guide and an invitation, aligning with the artistic intentions of the scroll. It encourages a sensory journey that reflects the dynamic interplay of nature and perspective depicted in the artwork. By bridging passive observation with active participation, the sound design enhances the immersive experience.

A1 Familiarization: Environmental Sound Awareness

Audio Content: Binaural spatial audio helps users recognize the surrounding environment through ambient sounds that establish a background, subtly introducing dynamic soundscapes that make users aware of their physical surroundings.

Experience: As users move through the space, they hear sounds like music or wind in the

mountains, grounding them in the environment without drawing immediate attention to specific sources.

Key Feature: Background audio lays the foundation for deeper exploration, enhancing immersion and creating a seamless auditory experience that draws users further into the environment.

A2 Exploration: Recognizing Dynamic Soundscapes

Audio Content: The soundscape evolves over time in response to the user's physical presence. The variety of sounds enrich the experience, reinforces a living, shifting environment, and provides contextual cues to help users orient themselves.

Experience: As users move through the space, sounds change in relation to their position (e.g., water growing louder near a river or leaves rustling more in a forest).

Key Feature: Dynamic audio enhances immersion, reinforcing the sense of a vibrant space while offering contextual cues to help users navigate surroundings.

A3 Investigation: Noticing Sound Changes with Proximity

Audio Content: As users move through the space, auditory cues become more pronounced as users move closer to specific sources - like approaching a waterfall or walking towards a mountaineers in the distance — providing real-time feedback to enhance awareness of their physical position.

Experience: As users investigate further, they notice how sound evolves with proximity, creating a sense of discovery and guiding their focus (e.g., the sound of water intensifies as they approach its source).

Key Feature: Proximity-based sound cues guide users' exploration, adding a layer of interaction to the experience.

A4 Focused Listening: Detailed Evolution of Sound

Audio Content: Sounds become richer and more intricate as users focus on specific areas, revealing hidden details like subtle birdsong or the sound of wind rustling through leaves.

Experience: By pausing and focusing on a particular object or scene, users are rewarded with a more detailed and nuanced auditory experience (e.g., detailed natural sounds, such as specific footsteps).

Key Feature: Focused listening enhances the experience, encouraging users to linger and engage more deeply with the sound environment.

A5 Second-Level Focused Listening: Micro-Immersion Through Sound

Audio Content: Subtle, barely perceptible sounds—like slight variations in each interaction or minor shifts in the sound source—emerge when users closely examine certain objects,

reflecting the randomness and complexity of the environment and capturing more details.
Experience: With prolonged attention to a specific scene or object, sounds evolve, such as the intermittent sound of someone whistling in the distance, followed by the sudden, playful moo of a cow nearby, creating an evolving soundscape that responds to the user's focus.
Key Feature: The evolving micro sounds, driven by the randomness of nature, reward sustained attention and encourage active engagement by providing a fresh auditory experience with each interaction.

4.4 Audio Content

Auditory cues should align with the visual elements in the painting. The soundscape must be mapped to specific sound sources as they appear in the scroll's composition. In this framework, sound objects are strategically placed to correspond with key features in A Thousand Li of Rivers and Mountains. Users can only hear certain sounds when they physically move into specific spatial zones, ensuring an interactive location-based audio experience.

Building on the interaction concept, the collected sound clips are categorized into five primary types (Figure 28), leveraging spatial audio, adaptive soundscapes, and interactive triggers to create a dynamic and immersive auditory experience. A detailed analysis of the sound clips used in this project is provided in the following tables.

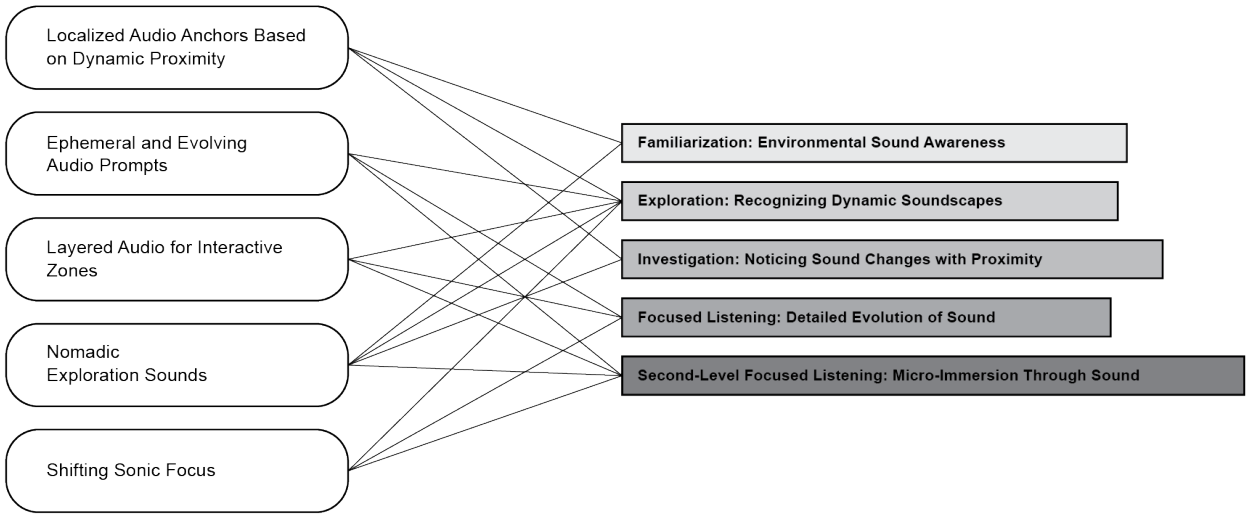


Figure 28: Audio Content Development, drawn by the author

4.4.1 Localized Audio Anchors Based on Dynamic Proximity

Spatial audio techniques will associate specific sound sources with different regions of the virtual scroll, such as streams, water wheels, or distant birds. These sounds act as static anchors, providing auditory connections to visually significant areas.

- This effect will utilize Unity's audio zones (within the defined boundary box) and attenuation curves to achieve precise sound scaling.
- Adaptive audio will respond to user proximity. For instance, as a user approaches a waterwheel or waterfall, the volume will gradually increase, drawing attention naturally.
- A flock of birds on the left side of the scroll will produce an approaching sound from the left ear channel, shifting gradually to the right as the user moves.

Category	**ID**	**Name**	**Duration**	**Description**
Localized Audio Anchors Based on Dynamic Proximity	A1	*Water Wheel Clip*	53s	- The sound of water flowing over a wooden waterwheel, with crisp rotations of the wheel.
	A2	*Waterfall Clip*	51s	- The sound of rushing water with natural reverberation.
	A3	*Bird Group Clip*	1m35s	- Multiple birds chirping in the sky, interwoven with the sound of flowing water.
	A4	*River Stream Clip*	41s	- The open lake surface is blown by the wind, making continuous and rhythmic waves.
	A5	*Ancient Chime Clip*	57s	- The sound of a watchman striking a gong, beginning with two sharp, resonant strikes that cut through the air, then fading into the distance.

Table 3: Samples of Localized Audio Anchors Audio Contents

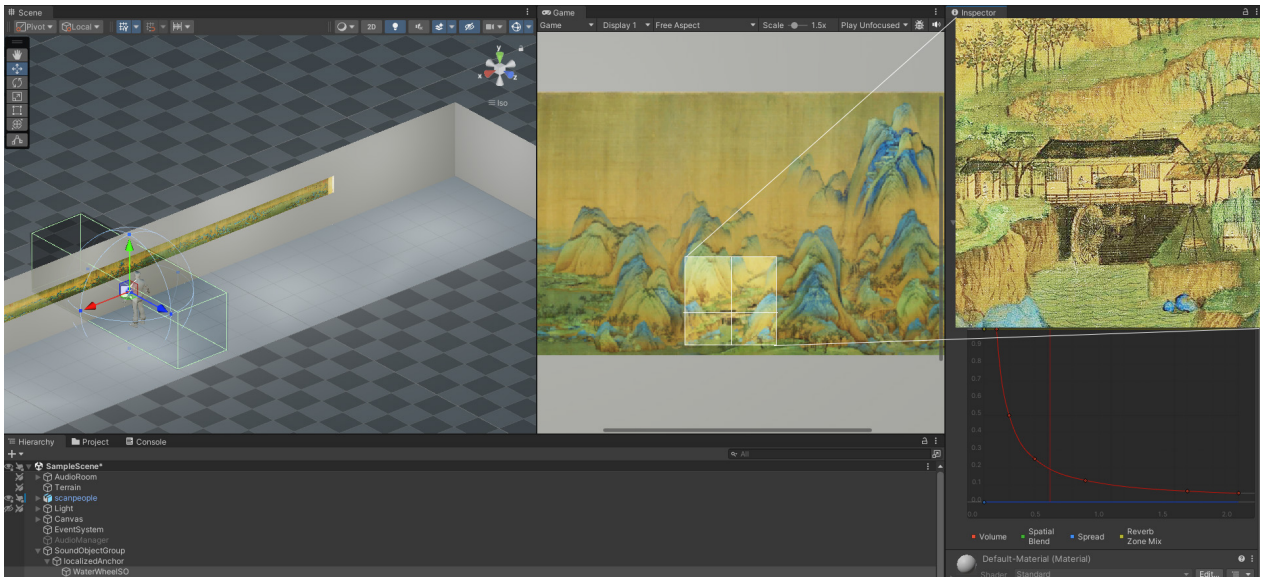


Figure 29: Samples of Localized Audio Anchors Audio Contents: WaterWheel, drawn by the author

4.4.2 Ephemeral and Evolving Audio Prompts

Short-lived, spontaneous sounds, such as the rustling of leaves, a bird's fleeting call or a traveler's whistle, will be used to create a sense of serendipity and discovery, ensuring diversity while maintaining thematic consistency.

Unity's Timeline tool and Random class will be used to (i) re-trigger sounds at irregular intervals, (ii) play audio of random durations, or (iii) play different variations of the same sound category from a predefined playlist at different times.

- **Example:**

  (i) As users enter a trigger zone, a walker whistling in the forest may play at random intervals and stop when they leave, avoiding predictable repetition.

  (ii) Footsteps in the mountains will follow a randomized playback schedule, simulating the irregular pacing of a traveler.

  (iii) In the mountains, users may hear different bird calls at different times — a sparrow chirping, a canary singing, or a flock of geese soaring overhead — creating a dynamic and evolving soundscape.

Category	ID	Name	Duration	Description
Ephemeral and Evolving Audio Prompts	B1	*Whistling in Forest Clip*	16s	- A person whistling while walking, with rhythmic footsteps.
	B2	*Footstep Clip*	32s	- Light footsteps crunching on fallen leaves, continuous and soft.
	B3-1	*Bird Singing Clip of Canary*	2s	- A brief, sharp, and melodic chirp, bright and fleeting, resembling a delicate musical note.
	B3-2	*Bird Singing Clip of Mahjong*	26s	- A layered soundscape where birdsong intertwines with the distant clattering of mahjong tiles.
	B3-3	*Bird Singing Clip of Goose*	2min4s	- (Longer due to the natural spacing between each call) Unlike continuous birdsong, honks and murmurs are distinct, with noticeable pauses in between. The tone and intensity vary, creating a non-repetitive pattern that mirrors how geese communicate in a natural setting.

Table 4: Samples of Ephemeral and Evolving Audio Contents

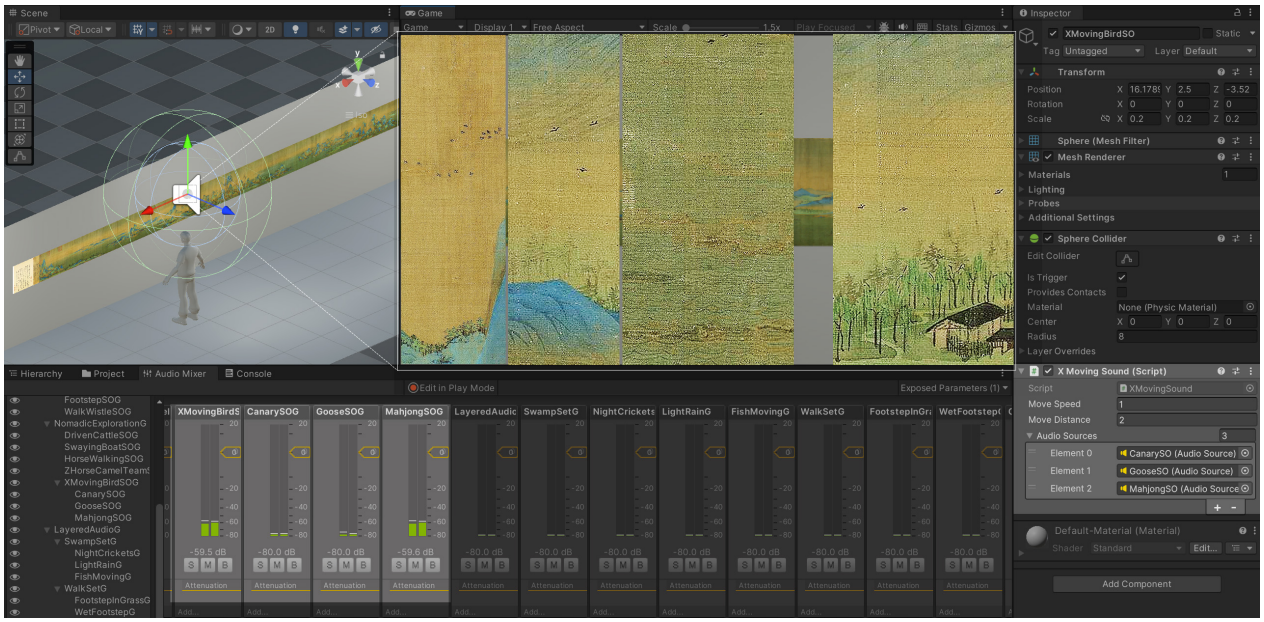


Figure 30: Samples of Ephemeral and Evolving Audio Prompts: bird singing playlist, drawn by the author

4.4.3 Layered Audio for Interactive Zones

To reflect the complexity and temporality of natural environments, interactive zones will feature multi-layered soundscapes that evolve over time (Figure31). This means that as users remain in the same area, the soundscape will shift unpredictably, capturing the dynamic nature of the environment.

- Multiple dynamic boundary boxes will expand and contract over time within the same location, altering the combination of layered sounds users experience.
- **Example 1:** In a swamp, the soundscape is rich and ever-changing, blending the gentle ripple of water, the steady hum of cicadas, and the occasional splash of a fish breaking the surface, allowing users to hear distinct elements at different moments.
- **Example 2:** As a user walks through a forest, their footsteps create dynamic interactions with the terrain. Step onto soft grass with a gentle rustle, then onto damp soil with a deep, muddy texture, enhancing the sense of presence and materiality.

Category	ID	Name	Duration	Description
Layered Audio for Interactive **Zones of Swamp**	C1-1	*Jumping Fish Clip*	2s	- A fish leaps and lands back into the water.
	C1-2	*Insects and Water Clip*	15s	- The continuous hum of insects fills the air, blending with the faint, persistent hum of cicadas.
	C1-3	*Swamp Ambient Audio Clip*	34s	- In the quiet environment, the faint sound of cicadas becomes more distinct, accompanied by subtle water movements.
Layered Audio for Interactive **Set of Foot-Steps**	C2-1	*Walk on Grass Clip*	10s	- The soft, rhythmic crunch of footsteps on thick grass, with occasional slight shifts in pressure.
	C2-2	*Wet Footstep Clip*	16s	- The muted, squelching sound of footsteps in damp soil or waterlogged ground.
	C2-3	*Grass Rustling Clip*	1s	- A brief rustle of grass as something moves through it, creating a quick, subtle disruption.

Table 5: Samples of Layered Audio Contents

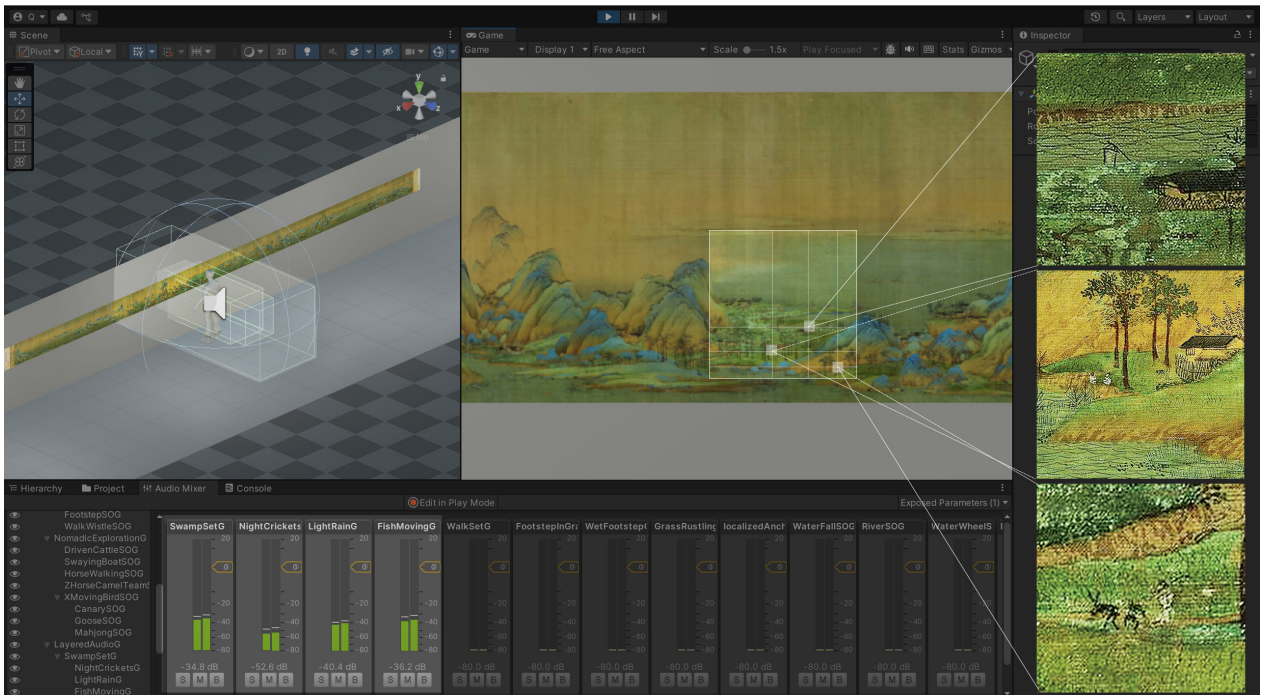


Figure 31: Samples of Layered Audio for Interactive Zones: swamp set, drawn by the author (Three dynamic-resized sound objects overlap illustrated by transparency)

4.4.4 Nomadic Exploration Sounds

To create a sense of mystery and enrich the soundscape, some sounds will exhibit nomadic movement, shifting dynamically within the scene at varying speeds.

There are three kinds: (i) **Horizontal Axis**: Sounds moving along the scroll's x-axis enhance the sense of visual expanse and motion, reinforcing the panoramic nature of the painting. (ii) **Vertical Axis**: Sounds traveling vertically along the scroll (z-axis), from distant to near, serve as an auditory guide, subtly drawing users closer to the painting. (iii) **Collision Triggers**: Sounds triggered only during close-range collisions encourage the user to capture vivid audio as the sound source moves within a small, nearby area. Overall, these add an element of intrigue, dynamic, and complexity, inviting deeper exploration.

- **Example:**

  (i) A slow-moving, drifting boat on a lake follows a rectangular trajectory, gently rocking back and forth between the shores. As the ancient saying goes, "The fisherman's bamboo flute at twilight," with the sound of the bamboo flute from the boat echoing through the

valleys. The perception of distance subtly shifts as the melody ebbs and flows with the boat's movement.

(ii) If the user is too far from the scroll painting, a dynamic auditory cue will encourage them to move closer. For example, the sound of a horse caravan emerges from behind, with rhythmic hoofbeats gradually approaching before passing by and fading into the distance, subtly prompting the user to follow the auditory path and step deeper into the world of the painting.

(iii) When a user unexpectedly bumps into a scene where a herder is driving cattle, the sudden startled moo of the cow, previously unnoticed by the user, draws attention to the hidden presence of the cow in the landscape. As the user and the cow maneuver around each other in the virtual space, it sparks a series of continuous interactions, further immersing the user in the environment.

Category	ID	Name	Duration	Description
Nomadic Exploration Sounds	D1	*Fisherman's Bamboo Flute Clip*	3m1s	- The soft, melodic tones of a bamboo flute echo across the water and drifts in the breeze.
	D2	*Z-Moving Horse Walking*	14s	The rhythmic clatter of horse hooves rising and falling in succession, evoking the dynamic motion of a traveling horse.
	D3	*Z-Moving Camel Team*	25s	A caravan of camels moving steadily, with the occasional jingle of bells hanging from their harnesses.
	D4	*Cow Moo Clip*	2s	A single, deep and resonant moo from a cow, powerfully cutting through the ambient soundscape. The sound has a clear, penetrating quality that immediately grabs attention.

Table 6: Samples of Nomadic Exploration Audio Contents

4.4.5 Shifting Sonic Focus

These proximity-based audio cues create a fluid shift in attention—from ambient background layers to vivid, localized sounds—enhancing narrative immersion. As users approach specific objects, foreground sounds emerge, unveiling hidden narrative elements while ambient distractions fade.

Rather than isolating fixed sounds, this approach emphasizes transitions: the soundscape continuously rebalances as users navigate the space, allowing key sonic details to surface naturally. This dynamic interplay between background and foreground enhances spatial presence and encourages deeper narrative discovery embedded within the scene.

- Triggered only when the user is extremely close, revealing narrative scenes that align with the depicted environment.
- Using the Audio Mixer manager, all background noise is lowered by 30 dB, ensuring the foreground sound remains dominant.
- **Example:** When a user approaches a village, the sound of friends clinking glasses in celebration becomes audible. As the user nears a private school, the teacher's instructions to students emerge. Upon reaching a pavilion by the water, soft conversations between women become discernible—each sound immersing the user deeper into the environment and narrative.

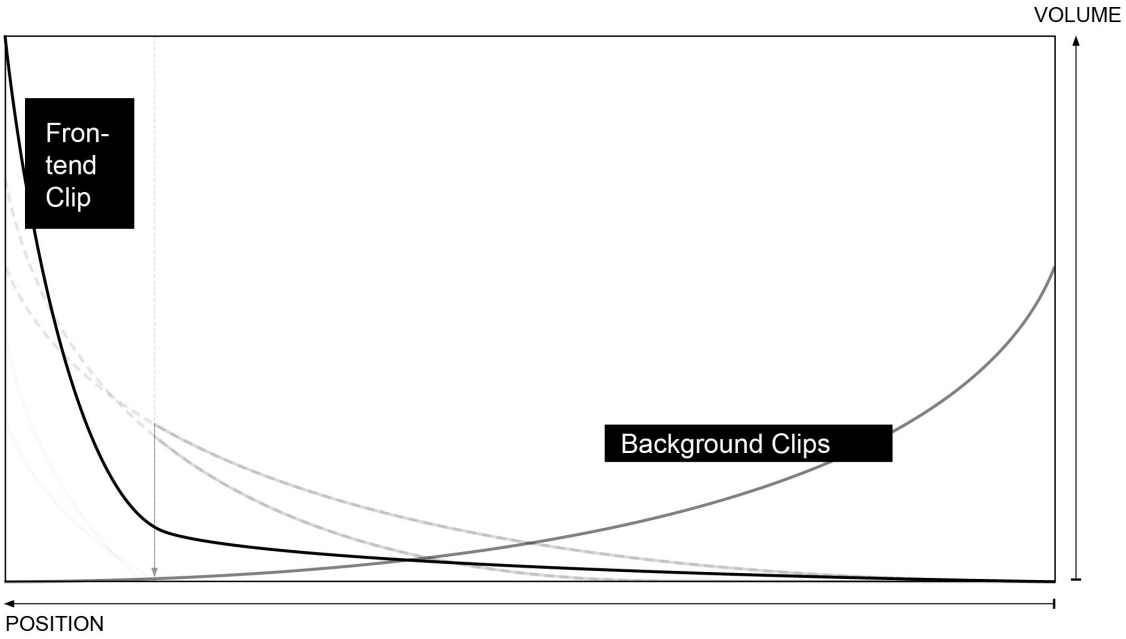


Figure 32: Foreground-Background Sound Transition Curve, drawn by the author
(Only One Focus sound, All Other Sounds Dim)

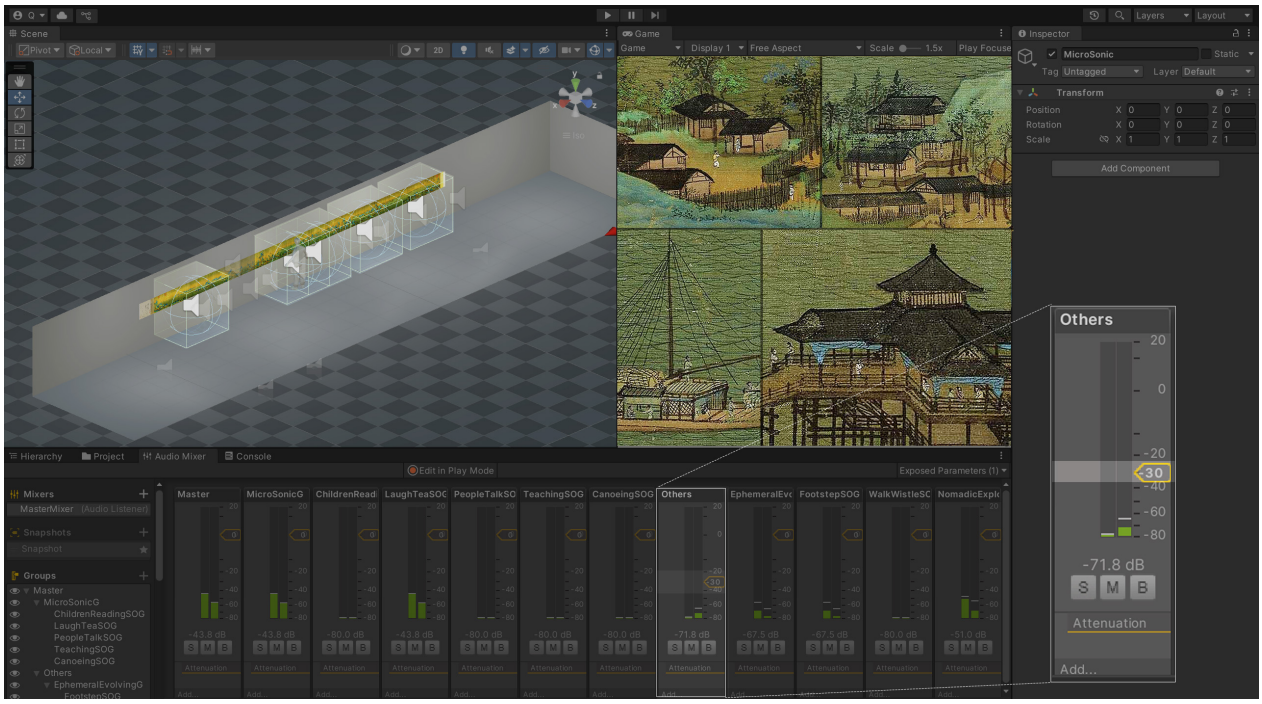


Figure 33: Samples of Shifting Sonic Focus, drawn by the author

Category	ID	Name	Duration	Description
Shifting Sonic Focus	F1	*Children Reading Clip*	27s	- Features the soft sounds of children reading aloud, possibly in a village private school setting.
	F2	*Crowd Speaking Clip*	1min44s	- Captures the murmur of a crowd in conversation. The voices overlap, perhaps from a market, gathering.
	F3	*Laugh and Drank Tead Clip*	1min15s	- The sound of laughter fills the background while people sip tea, possibly in a relaxed social setting.
	F4	*Teaching Poems Clip*	1min35s	- A teacher instructing students on poetry. The clear, measured tones of the teacher's voice dominate, occasionally punctuated by the students' responses.
	F5	*Canoeing Clip*	21s	- The sound of paddles dipping into the water, combined with the rhythmic motion of the canoe.

Table 7: Samples of Shifting Sonic Focus Contents

4.5 Evaluation

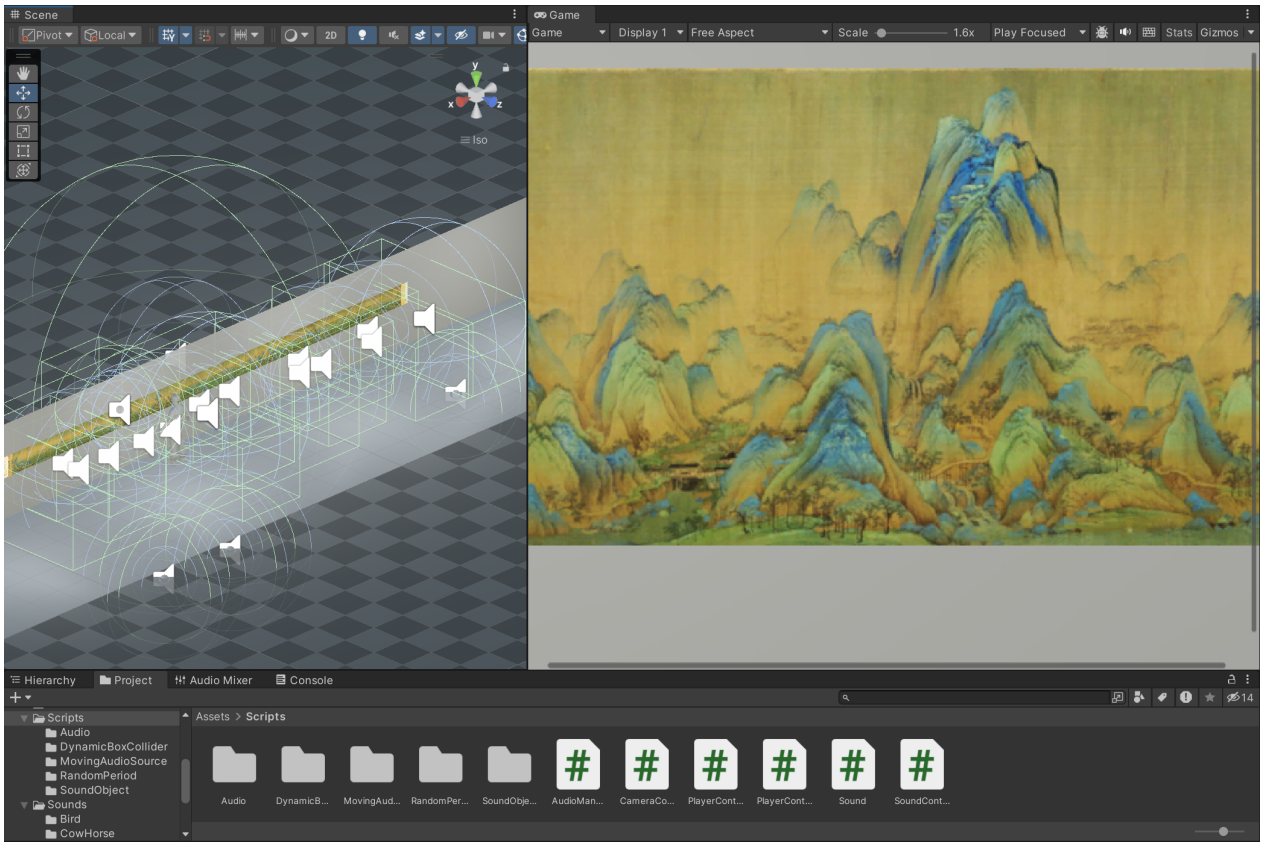


Figure 34: Unity Prototype Development, screenshot by the author

The goal of this user study was to examine how dynamic, spatialized soundscapes influence user attention, immersion, and exploratory behavior in Augmented Audio Reality (AAR). Building on the soundscape framework introduced earlier—**(A) localized audio anchors, (B) ephemeral and evolving audio prompts, (C) layered audio for interactive zones, (D) nomadic exploration sounds, and (E) shifting sonic focus** - this testing phase evaluates how effectively these auditory strategies shape the user experience. Specifically, it investigates whether sound can guide spatial engagement and narrative discovery without relying on visual prompts alone.

Grounded in the set of criteria established earlier, the study assesses how well the experience supports embodied interaction, open-ended navigation, temporal unfolding, and layered discovery. These criteria serve as a flexible framework for evaluating whether users move intuitively, feel physically present in the space, and uncover content through depth, surprise, and curiosity.

Eight participants (Figure 35, Table 8) were invited to freely explore the AR scroll installation,

experiencing its visual and auditory components without scripted instructions. Their movement, viewing angles, and pacing were self-determined—they could linger, roam, or revisit areas as they chose. This open-ended structure allowed naturalistic behavior to emerge, enabling a more authentic evaluation of how sound elements influenced exploration.

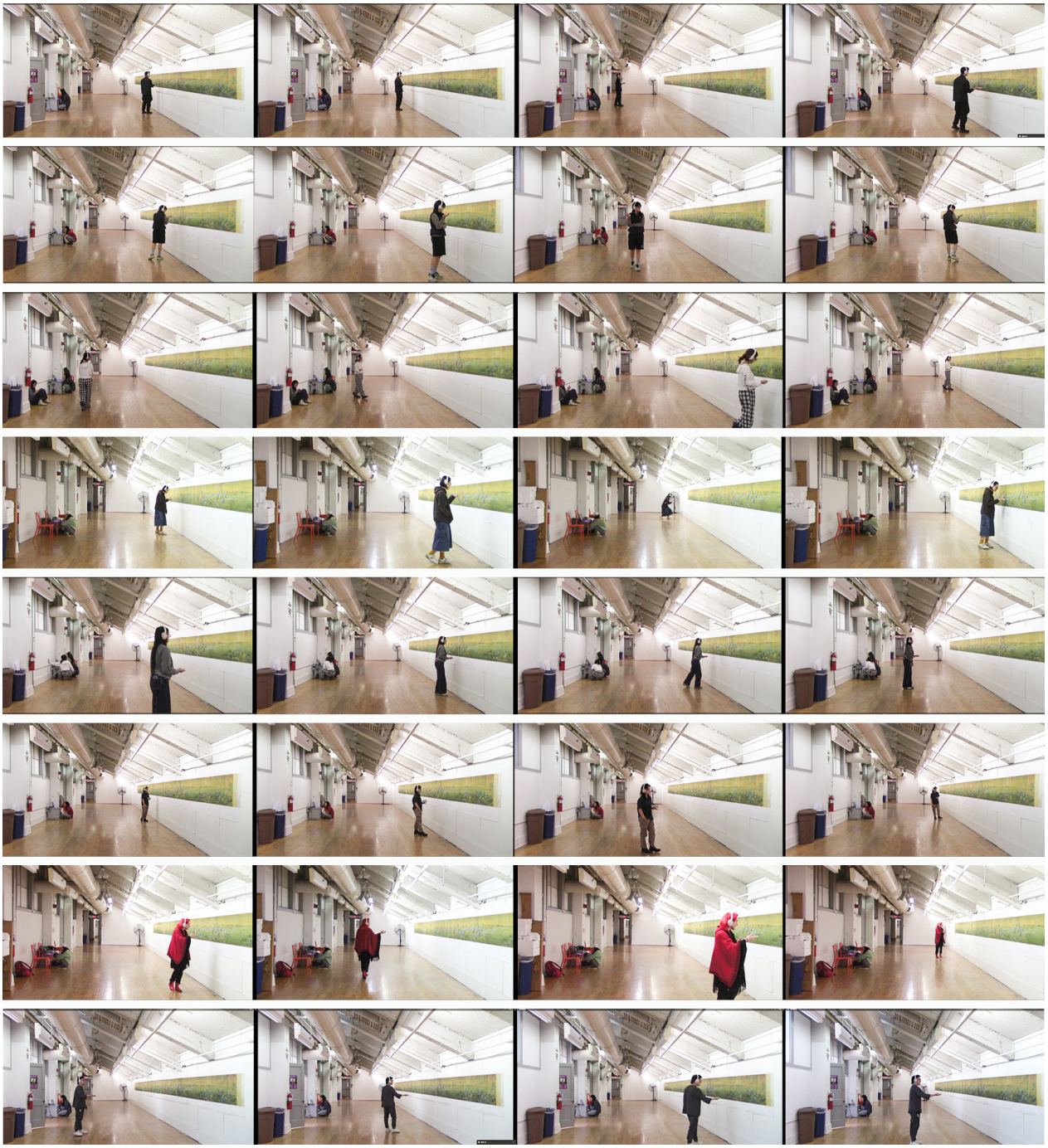


Figure 35: Participants Experience Photos, by the author

4.5.1 Participants and Behavior Data Overview

Name (Alias)	Mandarin Speaker	Nationality	Experience Time	Seen Painting Before
User1	Yes	China	12min17s	Yes
User2	No	South Korea	11min57s	No
User3	Yes	China	7min3s	No
User4	Yes	China	9min9s	Yes
User5	Yes	China	4min9s	No
User6	No	Vietnam	8min32s	No
User7	No	United States	4min58s	No
User8	Yes	China	9min57s	Yes

Table 8: Overview of participant information and viewing experience

As shown in Table 8, eight participants were recruited from the local campus community, comprising Mandarin and non-Mandarin speakers from diverse national backgrounds, including China, South Korea, Vietnam and the United States. The experience times varied, ranging from just over 4 minutes to more than 12 minutes, with respect to the open-ended and self-directed nature of the interaction. Half of the participants had previously seen the original painting.

To analyze their spatial and temporal behavior, sensor data capturing user coordinates were recorded at 10-ms intervals. These data provided insights into movement patterns, dwell times, and interaction zones. Simultaneously, Unity's dynamic soundscape adjustments were fully logged, allowing for a review of how audio cues influenced user navigation and perception.

The data collected will be used to evaluate how the design of the soundscape impacts user engagement, spatial awareness, and immersion within the AAR environment. Future analysis will focus on identifying behavioral trends, preferred exploration paths, and the effectiveness of different sonic elements in guiding user attention. There are two data visualizations that provide insight into user behavior in the AAR:

- **Spatial-Temporal Movement Analysis** (Figure 36, 38, 40, 42, 44, 46, 48, 50) presents a spatial-temporal analysis of user movement within the augmented soundscape. On the left, a time-series path illustrates the user's trajectory, where the X-axis represents spatial

position and the Y-axis denotes time progression. The black line traces movement, while red-highlighted points indicate moments of prolonged engagement. These lingering points suggest areas where users experienced heightened interest, likely due to auditory or visual stimuli. On the right, a reconstructed spatial sequence overlays key sound events, such as ”Water Wheel,” ”Laugh and Drink Crowd,” ”Flying Bird,” and ”Cow Moo,” onto a panoramic representation of the scene. A comparison of pause points with these auditory landmarks reveals that sound cues significantly influence movement patterns, encouraging users to stop, listen, and explore specific areas in greater detail.

- **User Movement in Soundscape** (Figure 37, 39, 41, 43, 45, 47, 49, 51) further examines the relationship between sound and spatial interaction through a user movement heatmap. The top section maps the user's trajectory in a 2D plane, with transparency levels indicating the duration spent at each location—darker regions signify extended stays, marking key engagement hotspots. Below, a histogram and fitted curve quantify these dwell times, identifying peaks where users paused the longest. The three most prominent maxima indicate locations of repeated interest and deeper engagement, which can be further mapped onto the soundscape plan to analyze overlapping sonic layers at these focal points.

4.5.2 User Test and Evaluation 1

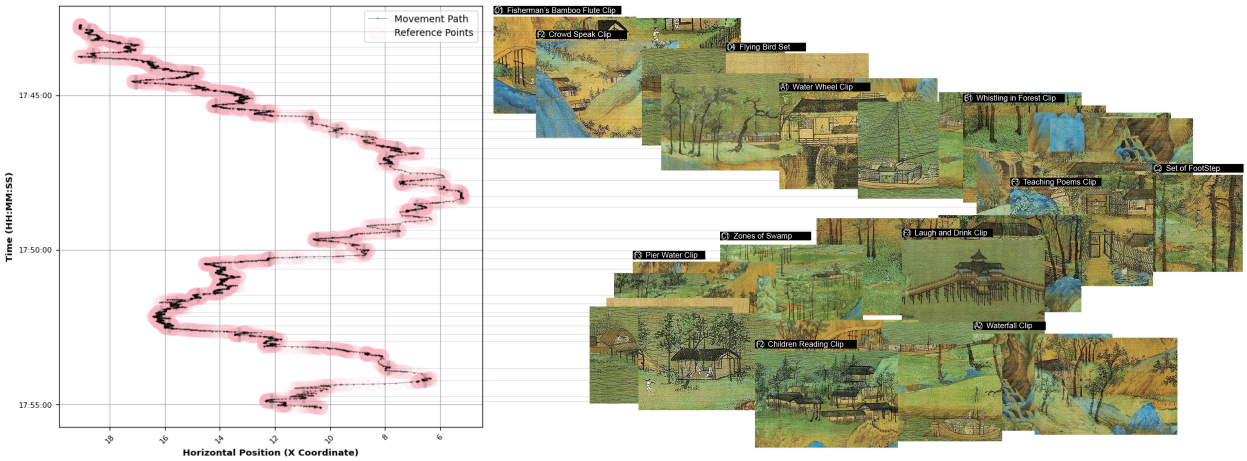


Figure 36: User Movement Over Time: Spatial-Temporal Visualization User 1, drawn by the author

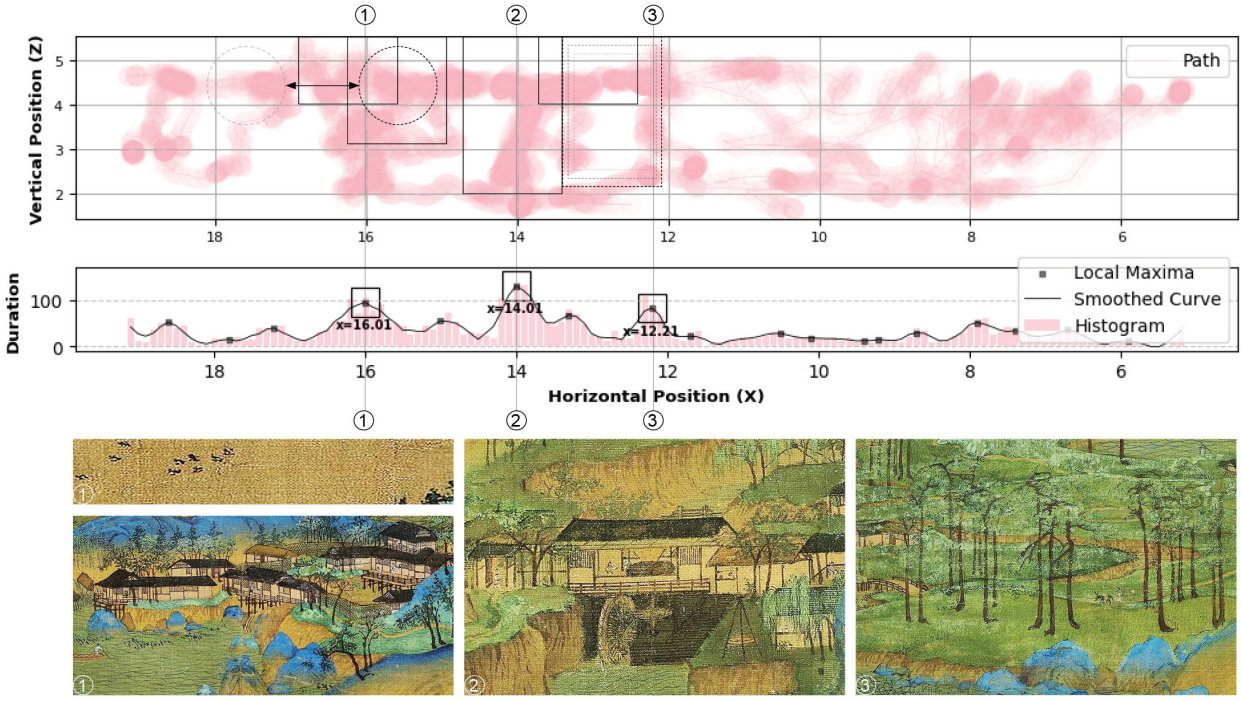


Figure 37: User Movement in Soundscape: X-Z Plane User 1, drawn by the author

The participant moves slowly through the space, stopping at specific points where sounds emerge. He lingers in front of the water wheel, standing still for several moments as if trying to catch the rhythm of its creaking and the water flowing beneath it. After a while, he shifts toward a pavilion where faint laughter drifts through the air, pausing near the entrance before stepping closer.

His pace varies — sometimes steady, sometimes hesitant — especially when nearing distinct sounds such as the rustling of trees or the distant chime of a bell. He stops in front of a school, listening intently, as if trying to separate the overlapping voices within. His movement is not random; it follows a deliberate pattern of stopping, listening, and adjusting, engaging with each sound as it connects to the scene before him.

"I have studied this painting before and am quite familiar with its composition, but this time, the experience felt different. Normally, I would glance over the scroll, taking in the overall scene without focusing too much on specific details. However, the presence of sound made me slow down and observe more closely. I found myself lingering in front of the water wheel, drawn to its rhythmic creaking and the sound of flowing water—something I had never paid much attention to before."

"As I moved, I realized that the sounds weren't just present—they shifted in intensity. This subtle transition in volume created a stronger sense of space, making the environment feel more immersive.

Instead of just seeing the painting, I was now experiencing it in layers, uncovering details I had previously overlooked."

4.5.3 User Test and Evaluation 2

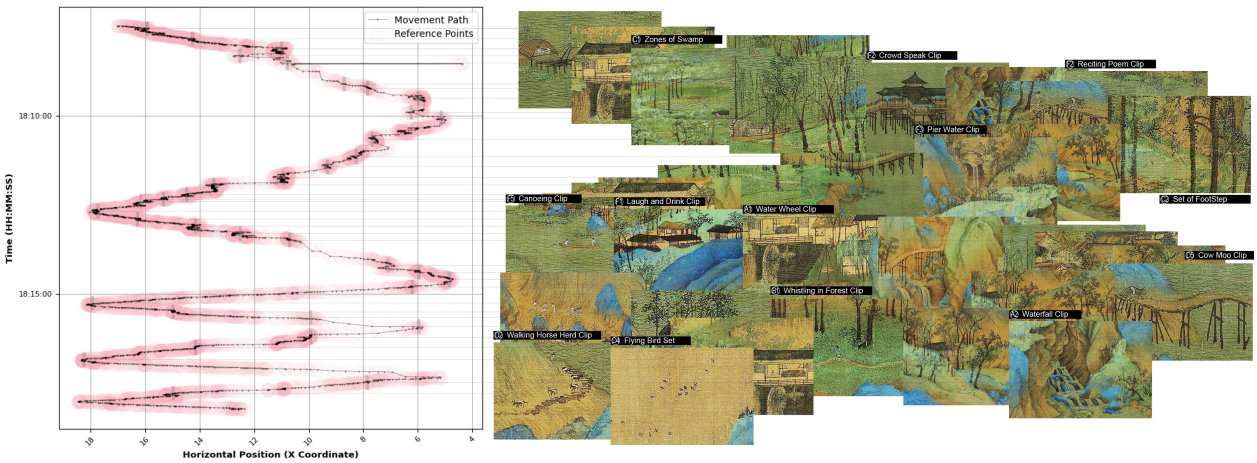


Figure 38: User Movement Over Time: Spatial-Temporal Visualization User 2, drawn by the author

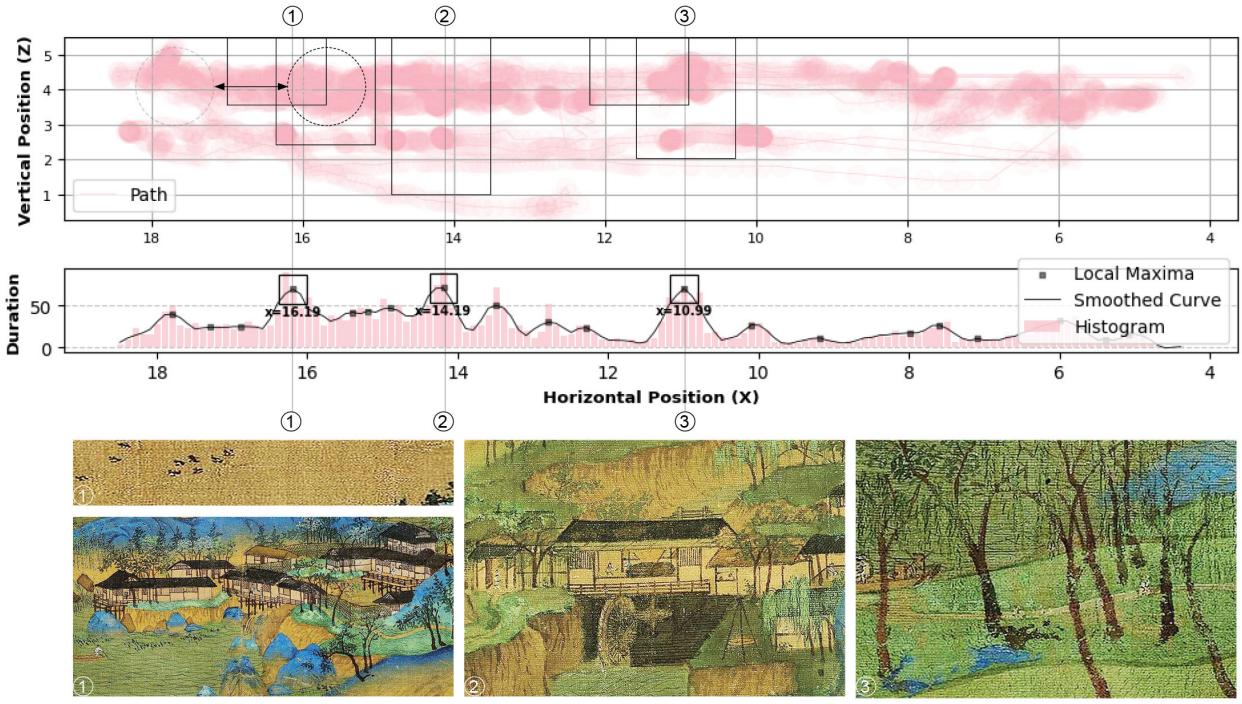


Figure 39: User Movement in Soundscape: X-Z Plane User 2, drawn by the author

The participant stands close to the scroll, almost pressing against it, his eyes scanning every stroke. The fine lines of figures and buildings emerge — villagers walking, boats gliding, rooftops stacked against steep cliffs. His head tilts slightly, listening, as if the painted world might speak.

The sounds rise around him. Nearby, quiet conversations drift from doorways, the soft clink of teacups punctuating the air. Women's laughter spills into the street, blending with the murmur of the river and the rhythmic clink of the oars. In the background, footsteps echo on stone and carts rattle over wooden bridges.

As the participant moves through his second and third passes, he steps away, no longer seeking detail but the larger flow. The mountains stretch further now, fading into mist. Villages become distant clusters. A bell tolls somewhere beyond the peaks, its sound barely reaching through the haze. He pauses, taking in the shifting soundscape — each step revealing a different layer of the world behind the ink.

"I listen, and something changes. Some areas highlight the environment — rushing water, rustling leaves—while others focus on human presence, the chatter of voices, the rhythm of daily life. Without the sound, I think I would have kept my distance, observing the entire scroll from afar. But the audio draws me in, shifting as I move. From a distance, it's ambient, abstract. As I step closer, details emerge — children's laughter near a school, the murmurs of conversation in what seems like a gathering place for adults."

"The painting is smaller than I expected. So many houses — I might have just walked past them without a second thought. The voices place me in the scene, letting me imagine who they belong to and what they are saying. I follow them, filling in the gaps with my own reasoning. This must be a courtyard. This, a school."

"Even when I don't understand the language, I recognize the emotions: excitement, routine, quiet contemplation."

4.5.4 User Test and Evaluation 3

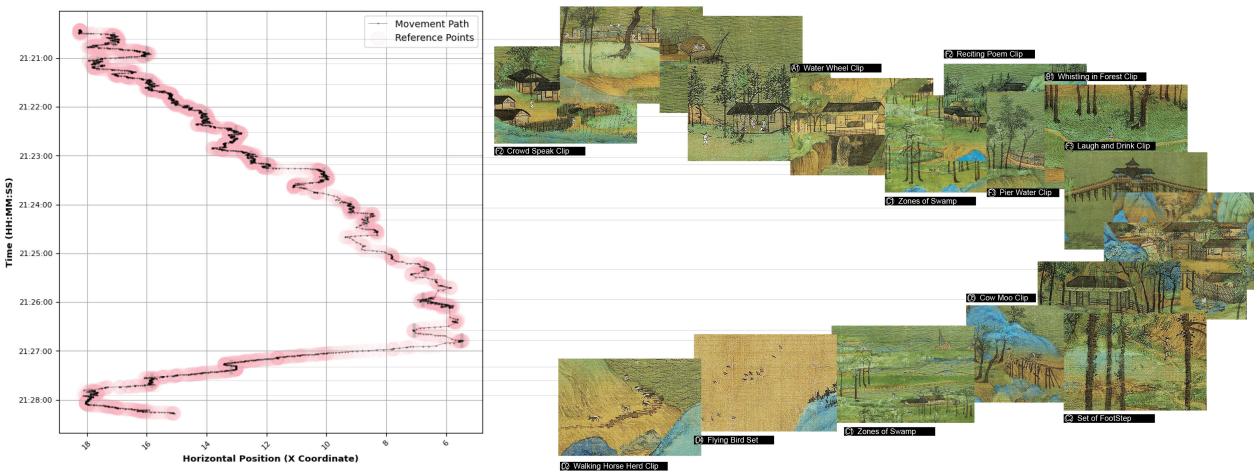


Figure 40: User Movement Over Time: Spatial-Temporal Visualization User 3, drawn by the author

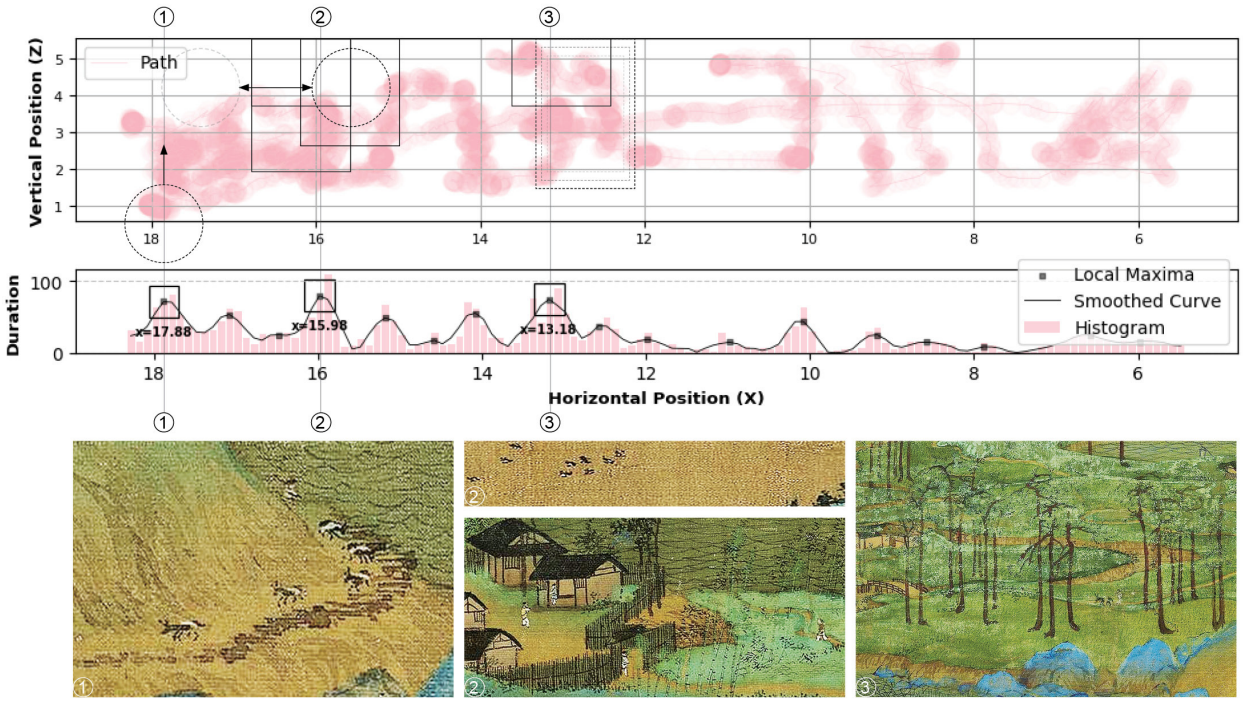


Figure 41: User Movement in Soundscape: X-Z Plane User 3, drawn by the author

The participant lingers by the water's edge, drawn in by the deep, rhythmic murmur of the current. In a distant corner, muffled conversations ripple through the air: blurred, indecipherable, yet irresistibly intriguing, perhaps villagers exchanging daily stories. Moving through a marshy stretch, he catches the rise and fall of frogs croaking, mingling with the sound of shifting water — a breath of summer

itself. The scholar's voice rings clear in a nearby school, yet as he steps away, it unravels into the wind. Near the pavilion, he pauses, expecting laughter and the soft clink of cups. From the forest, a whistle seamlessly blends with birdsong, drawing him forward. The children's recitation drifts through the valley, carried by the wind. Crossing a narrow wooden bridge, he suddenly encounters a herder, the call of his cattle rising — then vanishing before he can turn.

He returns to the beginning of the scroll, seeking to relive the experience as a whole. Stepping back to take in the grand scene, he stands still, only to trigger what had always been there — the soft chime of horse bells, emerging from behind and moving forward. He instinctively takes two steps ahead, as if drawn into the quiet procession of sound and movement.

"The sound of footsteps is the most immersive part. The moment I step onto the grass, it truly feels like I'm walking on something soft — I can even hear the faint snap of grass bending under my weight. As I keep moving, my own steps blend seamlessly with the ones already present in the painting, making me pause for a moment. Am I following someone or is someone following me?

"Above me, birdsong drifts and shifts, as if real birds are weaving through the forest canopy. Then, I notice—there are birds in the painting itself. It's such an unexpected moment, as if the painting and its sounds are playing a quiet trick on me, making me feel like I've stepped inside the scene.

4.5.5 User Test and Evaluation 4

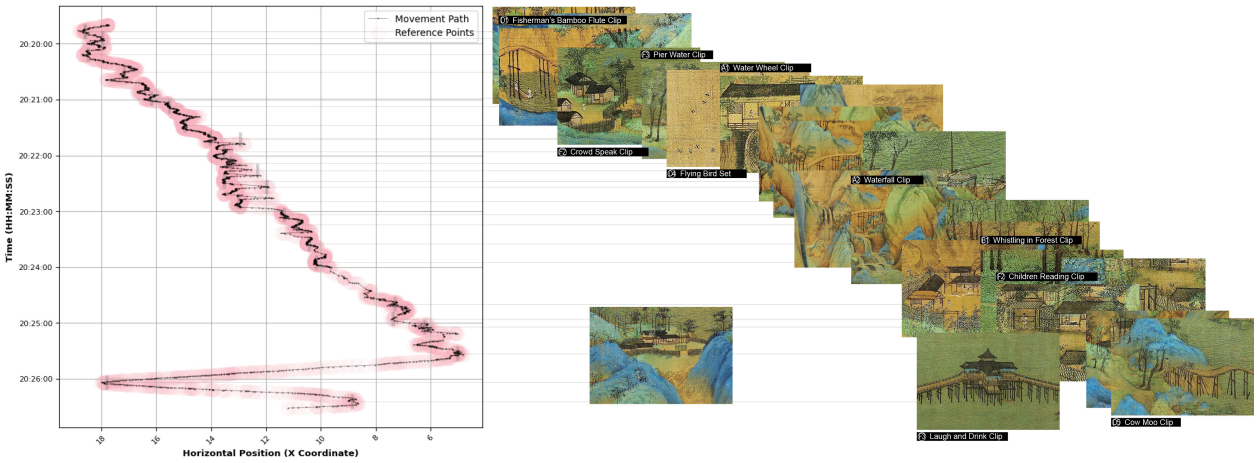


Figure 42: User Movement Over Time: Spatial-Temporal Visualization User 4, drawn by the author

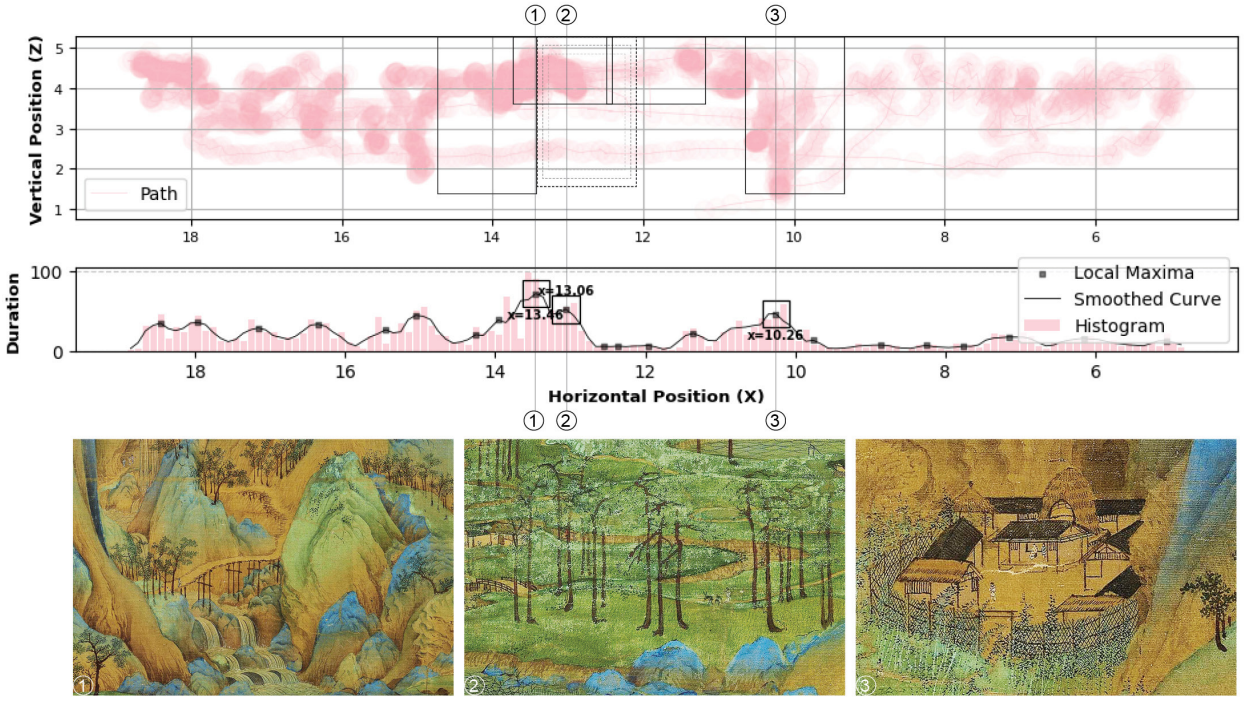


Figure 43: User Movement in Soundscape: X-Z Plane User 4, drawn by the author

The participant steps into the painting, greeted by the distant echo of a bamboo flute in the valley. He follows the murmur of flowing water, the whispers of a village, and the soft humming of a man carrying a hoe through the forest. As he ventures further, he stumbles upon a school, where children's recitations pause as he steps back — waiting. When he returns, the lesson has moved forward, the text advancing beyond what he last heard. A distant bell lures him toward a grand pavilion, an adult gathering space, before he wanders into the mountains. His footsteps on the floor sound like rustling grass, blurring the line between the indoors and the wilderness.

Suddenly, a short and abrupt cow call jolts him from his immersion. He stops, searching the painting for its source. Yet, as he moves, he realizes that the sounds do not simply vanish and reappear; rather, they ebb and flow, intertwining like threads in a tapestry. The receding whispers of the village blend seamlessly into the rising murmurs of the forest, the resonance of the schoolhouse lingers faintly beneath the chime of the distant bell. Each sound does not merely give way to another—it dissolves, merges, and reshapes, creating an ever-evolving symphony where past and present harmonize in a way that feels both natural and uncanny.

"That's when I notice the shift — when I'm near the painting, the sounds are rich with the human activities, conversations, movements, the texture of life. But as I step away, nature takes over. The

rustling of leaves, the distant call of birds, the wind over water through open spaces. It's as if the world within the painting dissolves into the vastness beyond it."

"I stop for a moment, listening. Somewhere nearby, people are chatting and laughing. The closer I get, the softer their sounds become. But as I step back, the noise swells, filling the space behind me. Why is it louder here? I move again, testing my theory, feeling the way the sound shifts, its presence shaping my experience in ways I hadn't expected."

4.5.6 User Test and Evaluation 5

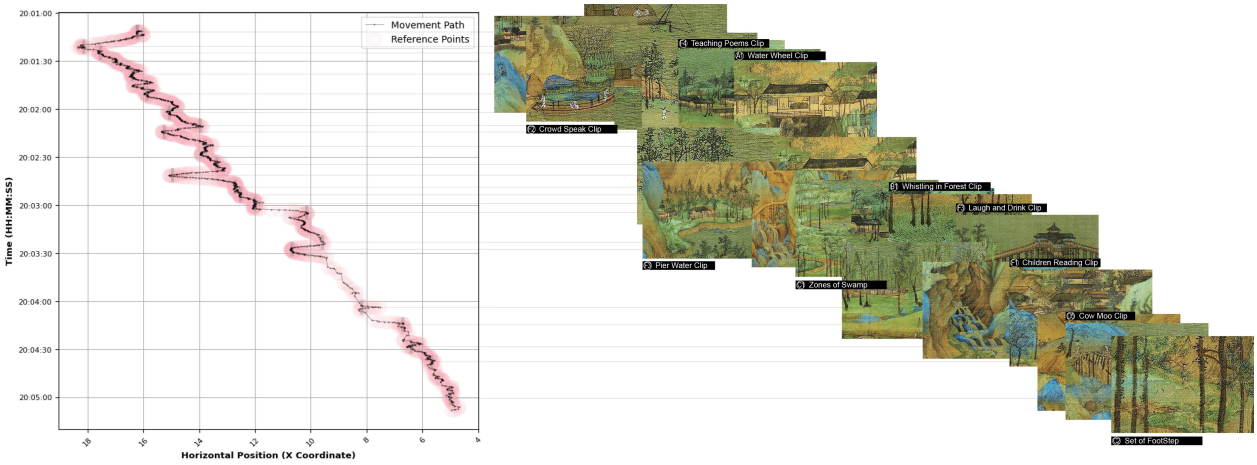


Figure 44: User Movement Over Time: Spatial-Temporal Visualization User 5, drawn by the author

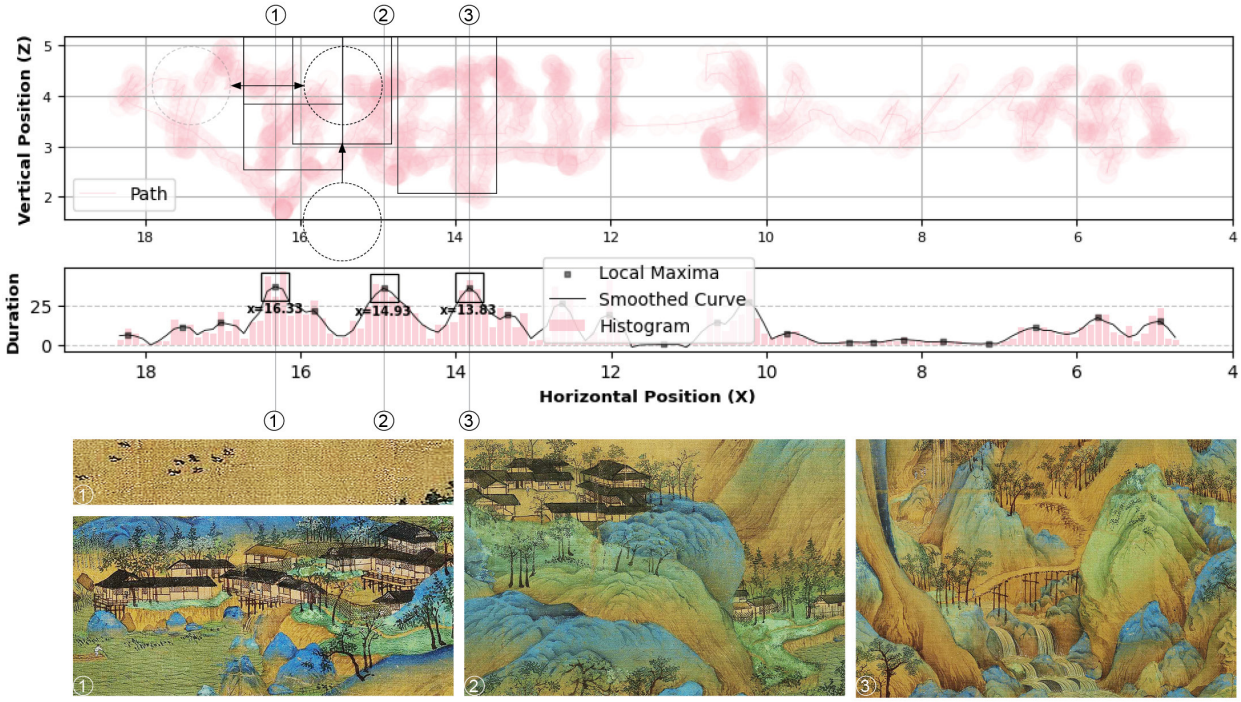


Figure 45: User Movement in Soundscape: X-Z Plane User 5, drawn by the author

The participant walks slowly and the path winds gently forward, with the scenery beside him unfurling like a scroll being revealed moment by moment. Just after passing a bamboo grove beside the rice fields, the faint voices of children reciting lessons drift in from afar, like pages rustling in the wind. It's as though the viewer becomes entranced by a particular thread of sound - captivated, not merely hearing, but actively seeking. Ahead, a water wheel creaks slowly in motion, its rhythmic groan interwoven with the gentle murmur of the stream. The two sounds blend like quiet companions, moving with unhurried grace. After winding through a mountain path, the occasional croak of a frog breaks the hush, joined by the faint flicker of fish brushing the water's surface.

The viewer sways gently — left, right, forward, back — as though unsure whether to move or to listen. With each shift of the body, the soundscape reshapes itself. Voices slip into the left ear and drift out through the right. The foreground swells and then recedes. The background looms close, then falls away, like chasing the echo of something just out of reach.

"I really liked the whole idea of how sound follows what I see in the painting. It's not just linear; it unfolds, swells, and recedes in response to where I am and how I move. That was incredibly successful."

“ I was especially drawn to how the sound of water varied—some places were soft and stream-like, others louder and more forceful, like a waterfall. It made me realize how finely tuned the sound palette was, even if most people might not notice those distinctions consciously.”

4.5.7 User Test and Evaluation 6

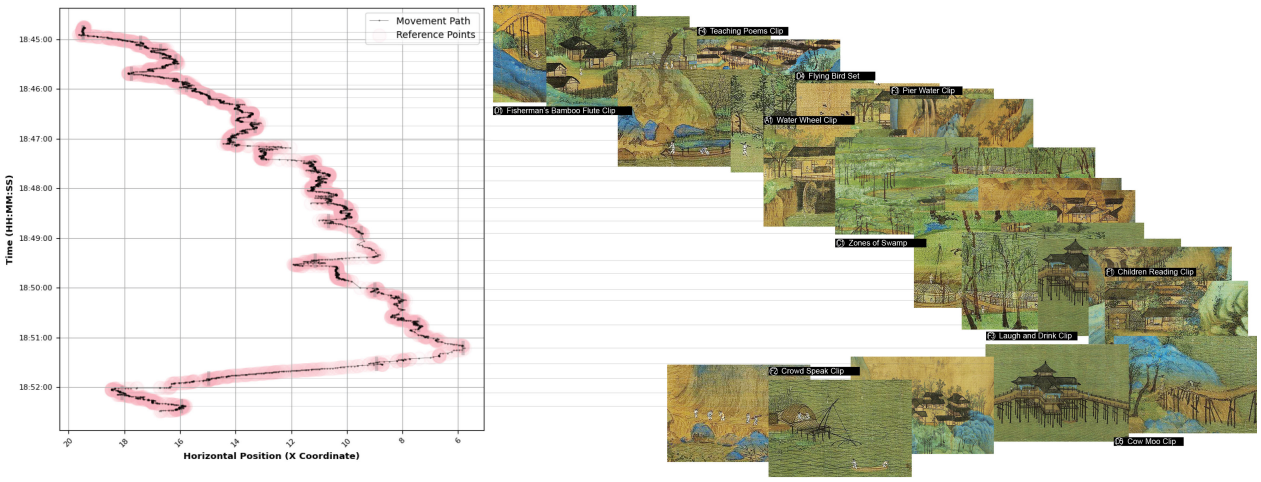


Figure 46: User Movement Over Time: Spatial-Temporal Visualization User 6, drawn by the author

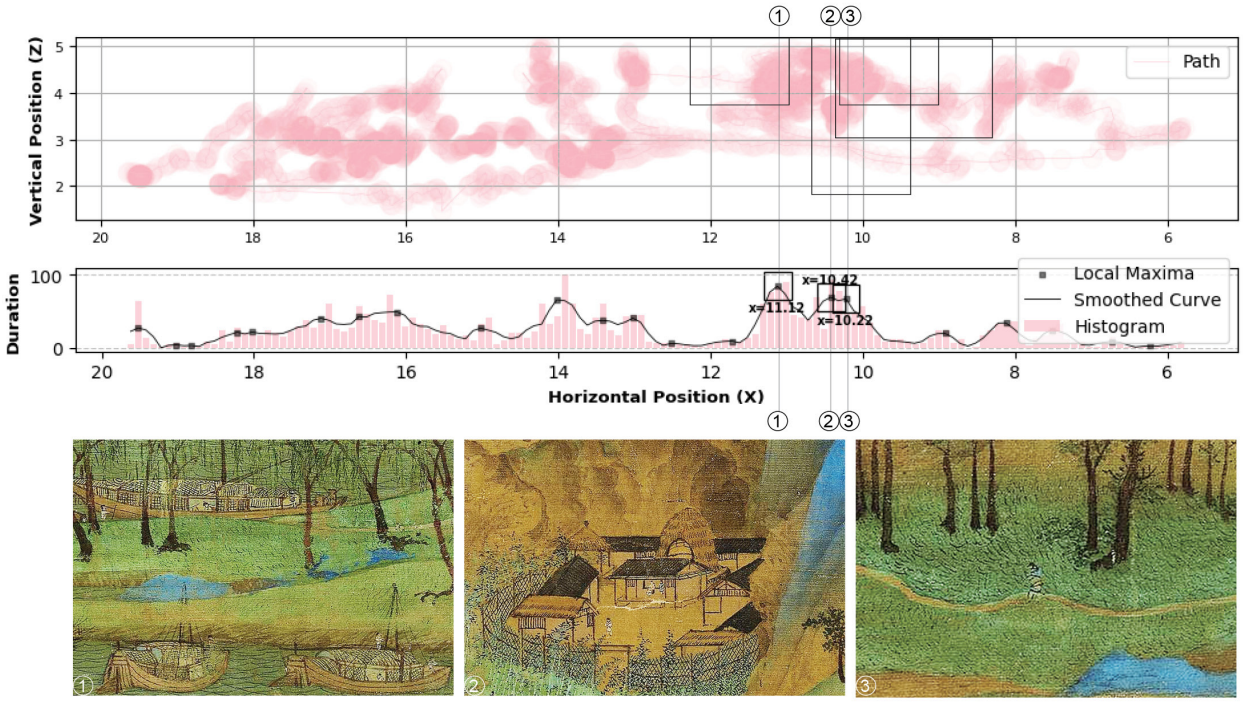


Figure 47: User Movement in Soundscape: X-Z Plane User 6, drawn by the author

The participant enters the scroll, guided first by the soft melody of a bamboo flute drifting over a tranquil river bend. The sound is gentle, almost hesitant, as if waiting for someone to listen. Then, laughter. The murmur of voices, the clink of ceramic cups, and the warmth of companionship ripple through a grand pavilion. Drawn by the faint creak of wood and the steady rhythm of a spinning waterwheel, he follows the meandering path. A cluster of birds rises into flight overhead, their cries scattered across branches and rooftops.

From within a courtyard, voices emerge—faint verses, a poem taught to unseen students. He leans closer, the cadence of the children's recitation folding into the rustle of reeds. It is a world set apart from the village chatter, yet echoing with the same rhythm of community. He lingers there, suspended between festivity and reverie. As he continues forward, a surprising discovery awaits: Within the painted landscape, a farmer walks home from the fields, whistling as he goes. Or maybe it's the birds that chirp on the trees, it's hard to tell. The two sounds intermingle so seamlessly that it becomes a game of guessing which is which. Maybe it's both. Either way, it's delightfully uncanny.

"I heard some words and that was quite surprising. It reminded me of those moments in games when you approach an NPC and overhear their conversation. The design made it feel like I was eavesdropping on something."

"I also noticed that when I walked back and forth in the same spot, the sounds would change slightly and I realized that I might be hearing different layers of sound depending on my position. It's fascinating how the sound changes based on where you stand. I tried walking back to see if it actually made a difference."

4.5.8 User Test and Evaluation 7

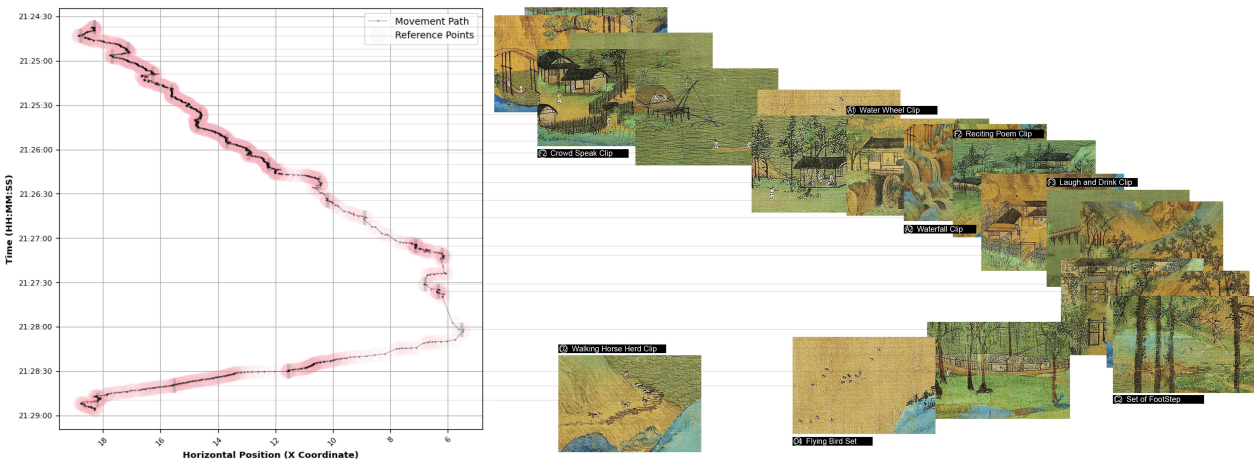


Figure 48: User Movement Over Time: Spatial-Temporal Visualization User 7, drawn by the author

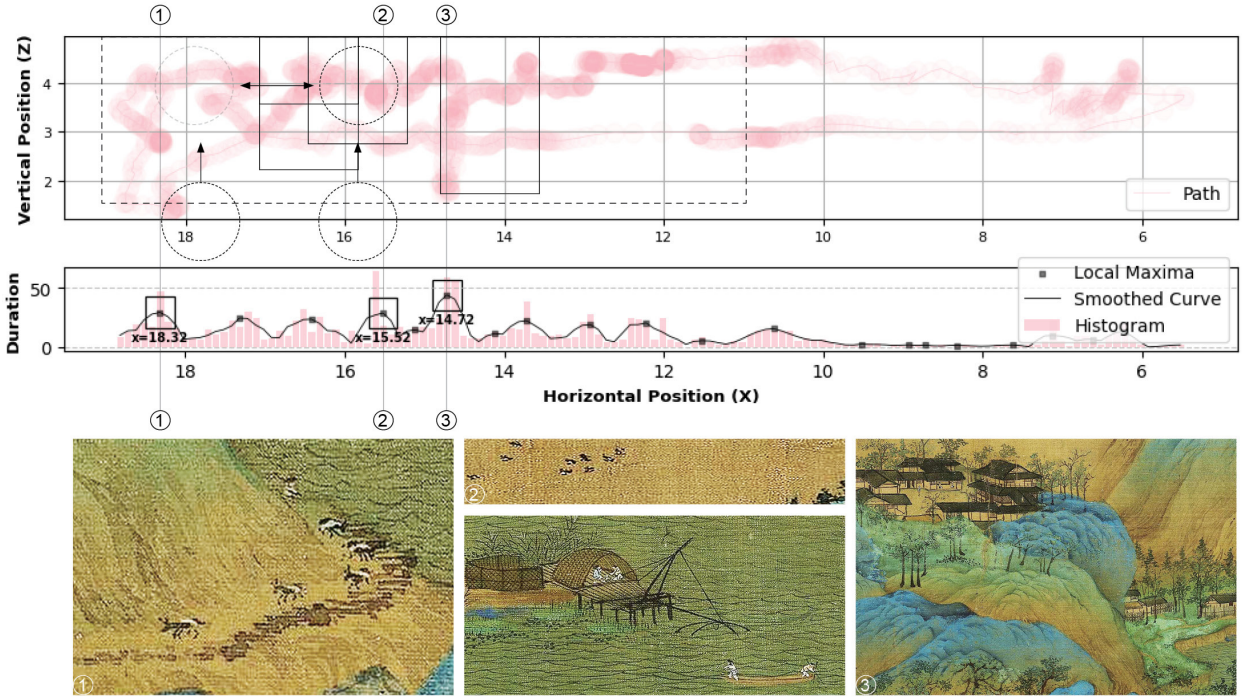


Figure 49: User Movement in Soundscape: X-Z Plane User 7, drawn by the author

A small boat glides across the water, its passengers moving in quiet harmony, as if performing a melody. The tune is soft, lilting, echoing through the valley. Just a few steps further, the participant enters the village — laughter rings out, cups clink, and conversations drift through the air. Eager to see more, he leans in close to the canvas, trying to catch the words of the tiny painted figures. But

stepping back a little, the voices fade, once again replaced by the calls of birds and the sounds of flowing water.

A man carrying a hoe passes through a shaded grove, humming a tune under his breath. His song lingers, curling around the viewer like a memory that refuses to fade. Suddenly, the clear, high voices of the children reciting lessons draw the viewer's attention deeper into the forest, to a hidden schoolhouse. He pauses, captivated — but when he leaves and returns, the lesson has moved on. In the distance, a faint bell chimes, leading him to a grand pavilion where laughter and a low murmur of conversation fill the air. He climbs mountain paths, crosses empty reed marshes, and gazes upon a vast lake, dotted with sails like stars scattered across the water. He reenters the village. Then a cow's cry shatters the delicate weave of sound, pulling him out of his dreamlike reverie. He freezes, scanning the scroll for its source. But instead of answers, his foot cracks a twig beneath the painted grass.

He begins to hurry back, hoping to retrace the story. But just as he slows down, planning to take it all in from a distance, a mounted caravan rushes past from behind. The kneading of the bridles surrounds him. He accelerates his pace and takes a few steps forward, chasing the lead horse's bell.

"I've never experienced a painting like this—where sound is embedded in the act of looking. It made me slow down, look closer and linger longer. It's as if the sound made the painting breathe."

"One of my favorite parts was when I heard this waterwheel sound and thought 'Wait, what is that?' It didn't fit at first. Then I scanned the painting, searching for it, and finally found it. That moment of discovery felt so rewarding, like a little hidden treasure."

"The long scroll and the way the painting slowly unfolds — especially with the sound — gives it this deep sense of time passing. Like I am walking through a moment stretched out. It made sense that I couldn't take it all in at once. I had to move slowly, bit by bit. It felt like reading a story in motion."

4.5.9 User Test and Evaluation 8

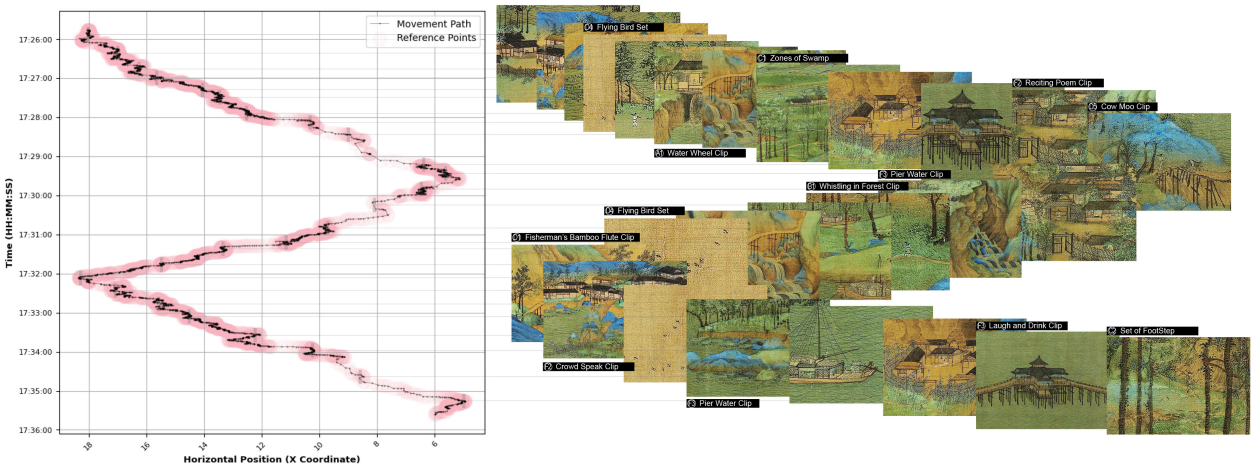


Figure 50: User Movement Over Time: Spatial-Temporal Visualization User 8, drawn by the author

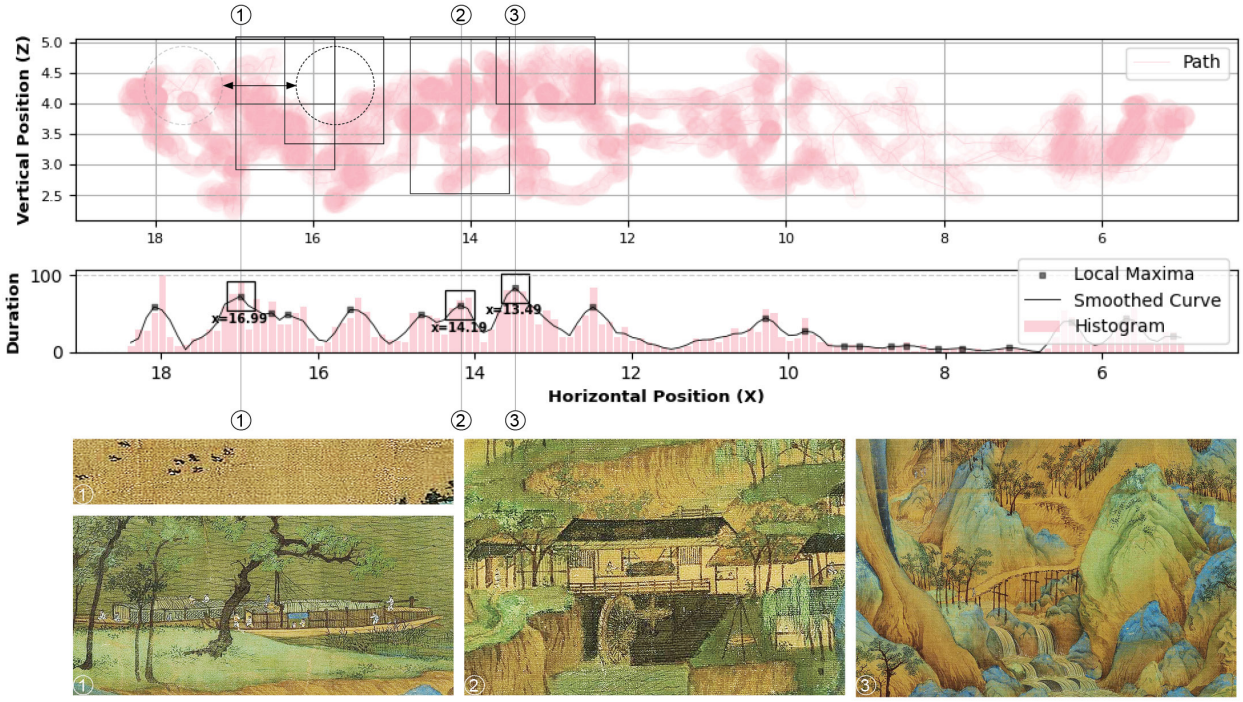


Figure 51: User Movement in Soundscape: X-Z Plane User 8, drawn by the author

Along the path, a soft melody from a fisherman's flute drifts from the riverbank. Ahead, laughter echoes from a pavilion shaded by trees, where adults gather and converse. Near the pier, the participant pauses — the gentle water sounds and the creaking of boats suggest a village that feels both historic and alive.

As a flock of geese glides overhead, the participant is drawn — almost unknowingly — into the mountains. In a hidden valley, the sound of an old waterwheel turning rings out, bold and rhythmic, its pulse accompanied by the thunder of a nearby waterfall. Then, as if the landscape is holding its breath, the noise gives way to a sudden hush. Nestled within this quiet is a secluded fishpond. A fishing net stands propped and ready, yet no soul is in sight. The silence is so complete that even the distant chirping of the frogs feels amplified. The participant instinctively slows, holds their breath, and walks on tiptoe, as if afraid to disturb something.

Returning to the edge of the water, they now face a vast lake dotted with sails. The mist blends the village with the sky and the flute sound returns - faint and distant. On the way back, he notices a school built on the hillside, where the children's voices reciting poetry become clearer with each step.

"There was a kind of playfulness in that. Sometimes I heard a sound and looked around to find its visual match, like with the waterwheel. Other times, I saw something, and just at that moment, its sound started playing."

"When I approached areas with buildings, I could hear human voices. One area really stood out—it felt so lively. There were people gathered, drinking, clinking glasses, chatting... I clearly remember those spots."

"Honestly, it felt like I had traveled through time. Suddenly, I was in ancient times. It was like I was physically walking inside the painting. Wherever my eyes went, especially when I got closer, the sound I heard matched what I imagined in my head. That was wild."

4.5.10 Second Observation and Conclusion

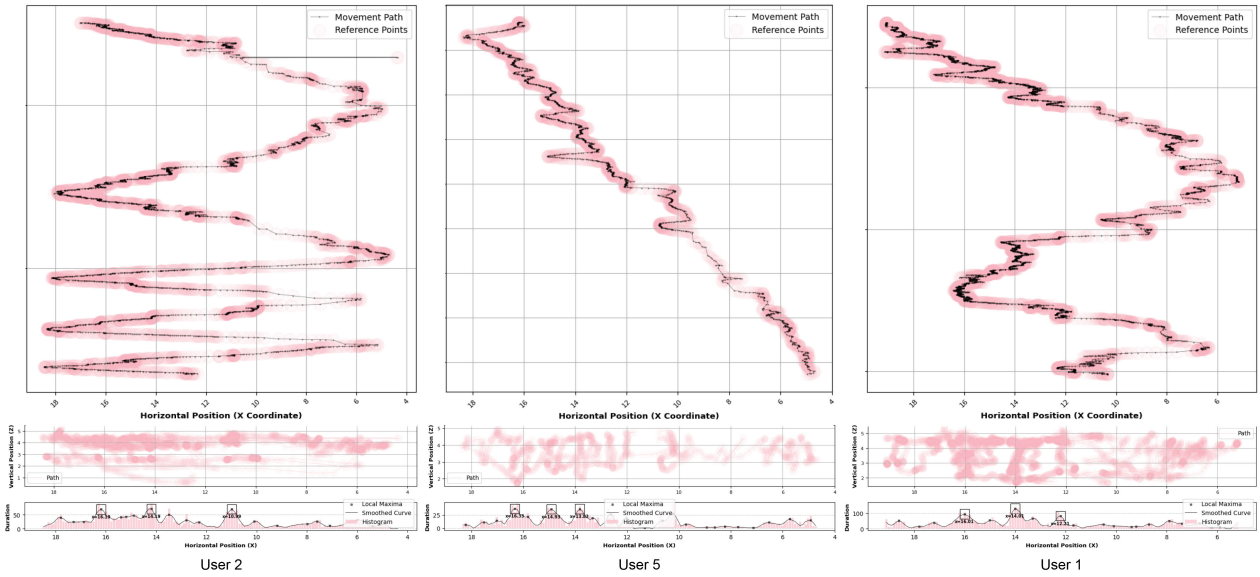


Figure 52: Three Path Patterns, drawn by the author

The Figure 52 illustrates three distinct user movement patterns, each shaped by individual behavioral preferences during their exploration of the soundscape.

- **Pattern A** (e.g., User 2) reflects a zigzagging movement strategy. These users tend to maintain a consistent viewing distance while exploring the entire artwork horizontally. After completing one full pass, they adjust their distance and begin a new pass, continuing this back-and-forth approach. This behavior is often driven by the desire to first examine the detailed visual elements up close and then step back to grasp the broader, more global view. These users are typically highly focused on the visual content itself. They actively search for sources of sounds, try to match audio cues with visual elements, and seek different interpretive perspectives within the visual field.

- **Pattern B** (e.g., User 5) is characterized by a more spatially embedded exploration strategy. These users progress steadily along the horizontal axis (X-axis), moving from one end of the artwork to the other, while frequently shifting their body position forward and backward along the depth axis (Z-axis). Their behavior is primarily driven by a fascination with the spatial qualities of sound. Rather than focusing on the visual content, they are captivated by how the soundscape transforms with each change in position. For them, every step offers a new auditory perspective and they aim to maximize the spatial richness of the sound experience.

It's not uncommon for these users to pause, retrace their steps, or move forward and backward around specific areas to re-engage with particular sound cues that resonate with them. The immersive nature of the audio sometimes guides their path more than the visual imagery itself.

- **Pattern C** (e.g., User 1), the most common behavior observed, falls between the two extremes. These users shift between focusing on the correlation between visual details and sound elements and being captivated by prominent foregrounds, backgrounds, or evolving soundscapes themselves. They sometimes dwell on the audiovisual interplay, while at other times they follow their curiosity toward specific acoustic or visual features.

In summary, the diversity of user trajectories in the space is a result of the interplay between personalized exploration styles, individual agency, and the immersive qualities of the spatial soundscape. This synergy fosters rich and open-ended experiences that vary widely from one visitor to another.

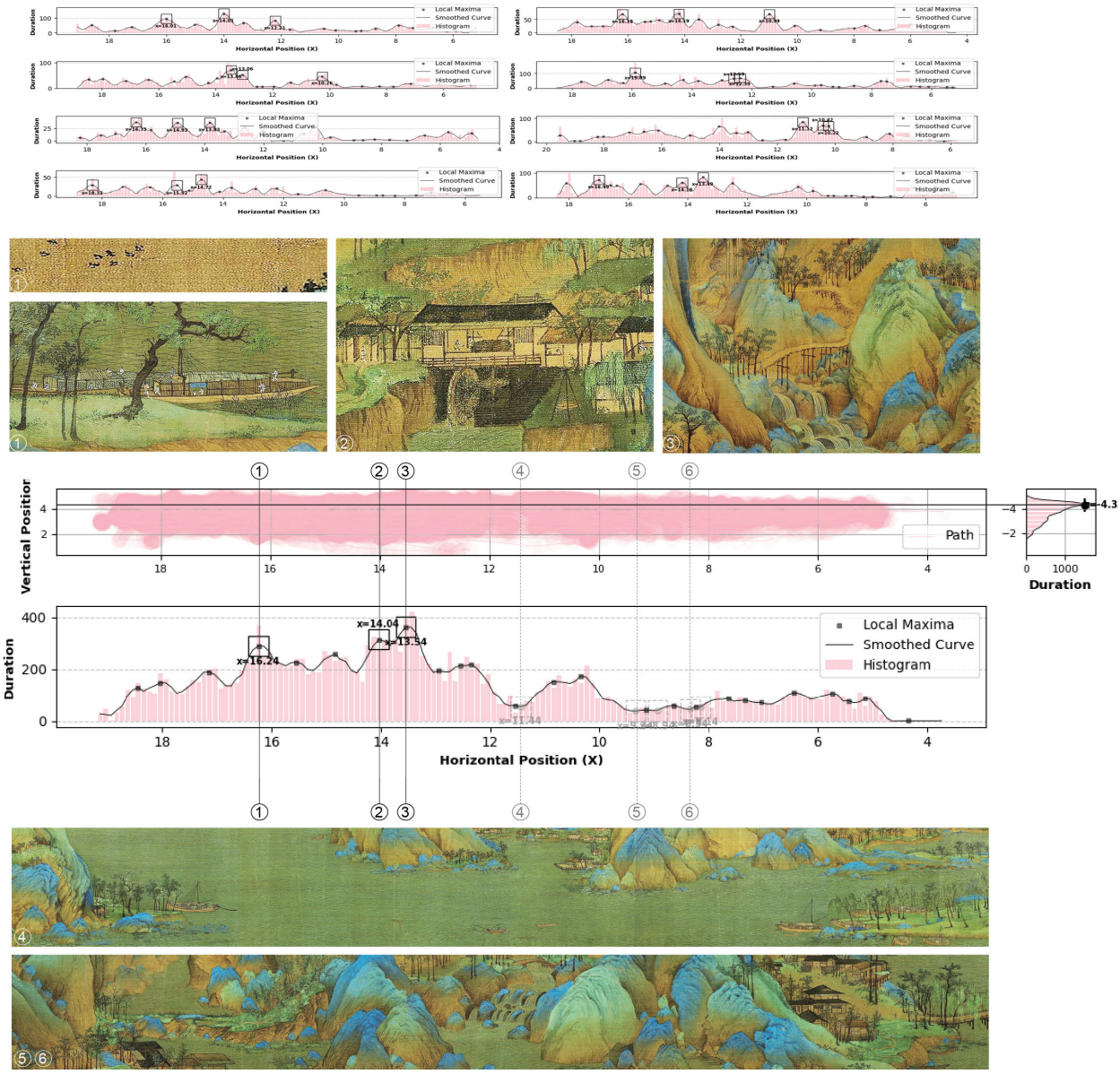


Figure 53: Cumulative Stay Distribution Statistics, drawn by the author

By overlaying the trajectories of the eight users in a single composite visualization, we identified a clear behavioral trend: The participants consistently spent the longest time in the **mid-left area** of the soundscape. Rather than the exact center, this region featured a higher density of dynamic sound events and richer audio layering, which contributed to its role as a key engagement hot spot. The spatial audio elements that overlapped in this zone created a complex and evolving sonic environment that naturally encouraged users to slow down, explore, and revisit. The most engaging regions correspond closely with the locations of specific spatial sound prompts, which can be categorized into the following types:

- **Ephemeral and Evolving Audio Prompts**: (Example: Bird Singing Clip Sets) These light, dynamic, and often playful audio cues shift subtly over time, evoking a sense of curiosity and liveliness. Their nomadic and moving quality draws users away from fixed positions, encouraging broader spatial exploration across the scene. Although they may not always result in prolonged pauses, they effectively promote fluid movement and repeated engagement, as users are drawn to follow their transformations and re-encounter them in new contexts.

- **Shifting Sonic Focus**: (Example: Crowd Speaking Clip) These clips create a rich auditory interplay between ambient background and localized foreground elements. Users tend to linger around conversational or socially charged scenes, intuitively seeking the origin or narrative behind the sounds. By simulating human presence and social interaction, these sound cues anchor attention and heighten immersion. The subtle reduction of ambient noise in surrounding areas further accentuates focal points, guiding the users toward detailed visual elements. This invites active interpretation, encouraging users to mentally reconstruct the scene, its stories, and imagined inhabitants through both visual and sonic cues.

- **Localized Audio Anchors Based on Dynamic Proximity**: (Examples: Water Wheel Clip, Waterfall Clip) These sound sources are tightly bound to specific locations, with their volume, clarity, or texture changing significantly depending on the listener's proximity. This spatial sensitivity triggers micro-movements — users often approach, retreat, or circle around these points to experience acoustic variation. The strong locational identity of these clips turns them into perceptual anchors, drawing repeated attention and encouraging users to return for deeper sonic interaction.

In contrast, the area where the participants spent the least amount of time corresponded to a deliberately sparse region of the soundscape. This low-engagement zone featured minimal auditory layering — typically only a single mid-ground audio source and the occasional drifting far-field sound. Visually, this section of the scroll depicted expansive and relatively uniform elements such as water surfaces or vegetation, with limited motion or fine-grained detail. The reduced sonic and visual complexity in this region appears to have offered fewer stimuli for exploration, reinforcing its role as a designed pause or transitional space within the overall experience.

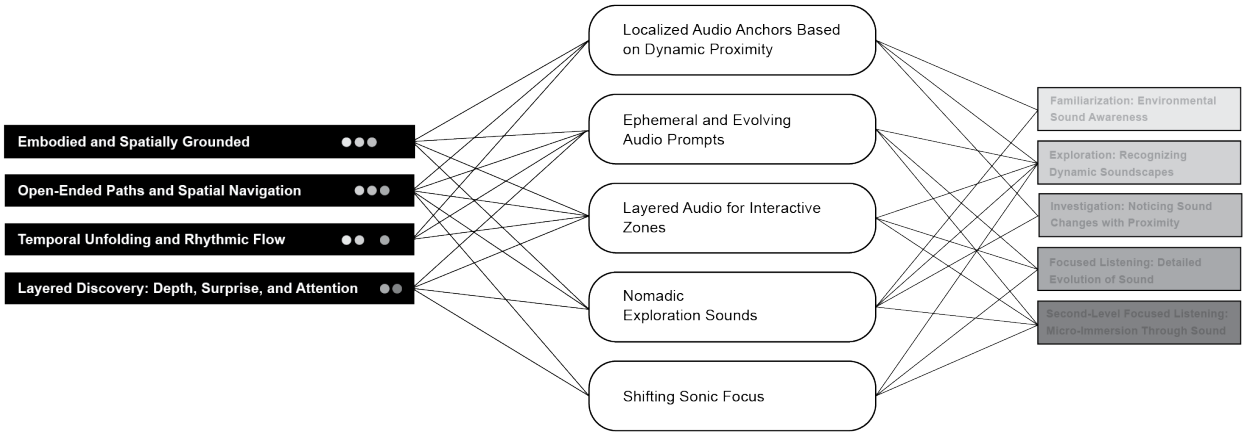


Figure 54: Criteria Insights, drawn by the author

Review the design criteria (Figure 54):

- **Embodied and Spatially Grounded** – This criterion received strong validation through user engagement patterns across all types of movement. Participants consistently interacted with spatial audio in a physically responsive manner — approaching, retreating, and repositioning themselves in relation to sound cues. Particularly with proximity — based anchors (e.g., the waterfall and waterwheel), users demonstrated high levels of embodied listening, confirming the effectiveness of spatially grounded design.

- **Open-Ended Paths and Spatial Navigation** – Observed behaviors revealed dynamic and varied movement strategies (Patterns A–C), reflecting a successful facilitation of non-linear exploratory navigation. Although many users carved their own paths through the soundscape, most of the trajectories still aligned with a few dominant zones, suggesting further potential for designing more evenly distributed points of interest to encourage greater path diversity and reduce clustering.

- **Temporal Unfolding and Rhythmic Flow** – Engagement with temporally dynamic elements - such as evolving bird song clips or shifting conversational scenes - indicated moderate success. These moments invited users to revisit areas and follow the sonic changes over time. Rhythmic flow was especially evident in Pattern C behaviors, where users alternated between lingering in audiovisual interplay and following emergent acoustic cues, creating a natural sense of pacing and re-engagement. However, occasional abrupt transitions or sparse zones disrupted the overall rhythm. This suggests opportunities for refining the pacing and crafting smoother temporal gradients that better guide user attention through unfolding audio layers.

- **Layered Discovery: Depth, Surprise and Attention** – Although the general engagement with this dimension was lower in areas lacking visual or auditory richness, the user's responses revealed a strong appreciation for moments of subtle surprise and sonic layering. Several participants, including Users 6 and 7, described the experience as "rewarding" and "playful," highlighting the enjoyment of discovering hidden sounds embedded in the environment. These reflections affirm the potential of carefully crafted audio layers to spark curiosity and sustain attention. However, such moments were relatively limited in scope. Increasing the density of layered cues, embedding more nuanced surprises, and improving the audio-visual interplay in quieter regions could amplify the effectiveness of this criterion and deepen the user's attention.

5 Findings and Discussion

By analyzing prototypes, it examines how Augmented Audio Reality (AAR) enhances spatial perception, guides movement, and deepens narrative engagement. The discussion highlights key successes, challenges, and design considerations.

5.1 Insights on "Flowing Space"

The concept of 'flowing space', which originates in architectural theory, offers a lens to understand how users interact dynamically with spatial environments. Applied to traditional Chinese scroll paintings, this concept transforms the experience from static visual observation to an interactive, multi-sensory journey. Through Augmented Audio Reality (AAR), users don't simply view the artwork—they move through its spatial rhythms, allowing the scroll to unfold gradually as a living and responsive environment shaped by both motion and sound.

Spatial audio becomes a guiding force in this unfolding, reinforcing depth, directionality, and a sense of immersive movement across the painted landscape. Precisely placed sounds construct an acoustic perspective that mirrors the visual composition, where directional cues and binaural audio simulate proximity and orientation. For instance, a waterwheel on the left emits a subtle, left-ear dominant sound that swells upon approach and fades when passed, while voices from a pavilion become clear and articulate when approached at close range, then gradually dissolve into ambient texture as the listener moves away. Foreground sounds, such as footsteps on a bridge or rustling leaves, are rendered crisply and immediately, while ambient elements, such as templating bells or distant chatter, gently ebb and flow, creating a layered sonic atmosphere that sustains continuity without abrupt transitions. Thematic audio cues further anchor narrative moments: the steady creak of a waterwheel signifies labor, while bird calls suggest a living, breathing natural world. These sounds not only guide perception, but also reinforce the temporal and emotional tone of each scene. The system invites slower, more contemplative movement, encouraging users to dwell, listen, and follow the internal rhythm of the scroll as it unfolds over time.

By weaving together spatial direction, intensity, and narrative theming, the scroll becomes a navigable soundscape — one that blurs the boundaries between physical space and virtual presence. Users are gently drawn into a sensory dialogue, where image and sound unfold in tandem, fostering a deeper connection with the painting as a continuously evolving immersive environment.

5.2 Active Roaming vs. Passive Viewing

Traditional Chinese scrolls are typically experienced as static visual narratives, where the audience observes the artwork in a linear and predetermined manner. The integration of AAR allows for a departure from this passive mode of interaction. Audiences are empowered to dictate the rhythm of their exploration, controlling their journey through the landscape, and discovering elements at their own pace. This personalized engagement fosters a deeper connection between the viewer and the artwork, enabling new interpretations and a richer appreciation of its narrative structures.

Sound plays a crucial role in shaping this experience. Unlike static visuals, spatial audio elements create an evolving soundscape that encourages spontaneous navigation. Some auditory cues act as focal points, drawing users to linger and examine intricate visual details, while moving sound sources stimulate curiosity, nudging users to follow the direction of the sound.

Beyond guiding movement, sound also interacts with personal perception and imagination. While the visual composition provides a foundation for interpretation, sound activates new layers of engagement, encouraging each user to construct their own version of the scene. The same landscape may evoke entirely different narratives depending on which sounds capture an individual's attention: One user might be drawn to the bustling energy of a market, while another may focus on the tranquility of a riverside moment. This subjectivity extends to the pace and rhythm of movement; some may drift leisurely between points of interest, while others navigate in quick, deliberate steps, their journey shaped by a unique interplay of visual curiosity and sonic intrigue. Ultimately, the fusion of sound and image transforms the experience of the scroll from a fixed linear sequence into a deeply personal exploration. The scroll, in turn, ceases to be a singular artwork and instead becomes a dynamic, ever-changing environment—one that is rewritten with every journey.

However, the soundscape design is suggested to be grounded in a respect for the artistic integrity of the original scroll. Designing an intuitive auditory experience requires a careful balance to ensure that spatial sound complements rather than overpowers the visual and narrative coherence of the artwork. If sound placement is too rigid, it may unintentionally impose a structured path, limiting the sense of spontaneous discovery. In contrast, excessive randomness in the sound distribution could create a fragmented experience, making navigation feel disjointed rather than fluid. By striking a balance between structure and flexibility, the auditory experience encourages natural exploration, allowing each individual to shape their own journey through the scene while preserving the essence of the artwork's storytelling.

6 Look Forward

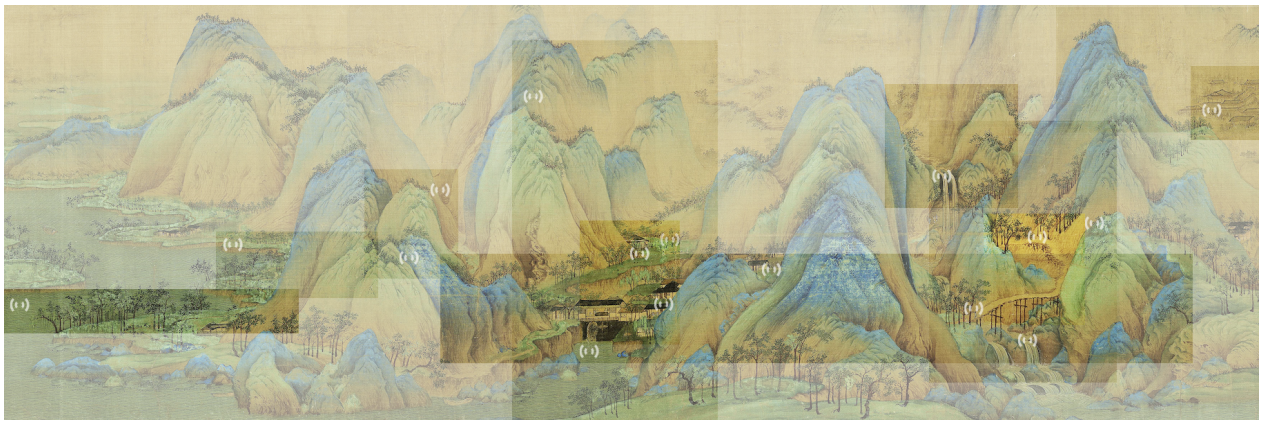


Figure 55: roaming in the soundscape, drawn by the author

This part summarizes the research contributions and limitations, outlines recommendations for integrating AAR into cultural heritage projects, and suggests potential extensions to other art forms.

6.1 Limitations

- **Technological Constraints** – The sensor data for user orientation had latency and accuracy issues, limiting its effectiveness in designing interactive soundscapes. Additionally, dynamically adjusting the soundscape in Unity based on real-time sensor input proved complex. The current implementation supports only a single-user experience, without multi-user interaction capabilities.

- **Fidelity Constraints** – The artwork was presented as a printed reproduction, resulting in a loss of clarity compared to the original. Due to printing constraints, the scroll was divided into four segments for display, which may affect the continuity of the experience. Moreover, the exhibition space had fixed lighting conditions that could not be adjusted to enhance immersion.

- **Time Limitations** – The research, development, and writing process took place within six months, during which sensor integration, data transmission, soundscape design, iterative refinement, and user testing had to be completed. This limited the depth of testing and optimization.

- **User Testing and Feedback** – The diversity of user testing was constrained by time and availability. Factors such as participants' linguistic and cultural backgrounds, prior familiarity with the artwork, and the limited duration of the exhibition (due to space rental constraints) impacted the breadth of feedback collected.

Addressing these challenges in future iterations will enhance the fluidity, accessibility, and depth of engagement with the AAR experience.

6.2 Future Work

Looking ahead, this research lays the groundwork for integrating AAR into cultural heritage initiatives. The following directions highlight potential areas for growth.

6.2.1 Multi-User Interaction and Emergent Soundscapes

An direction for future work involves introducing multi-user interaction, with the hope of fostering shared engagement with the scroll. By allowing multiple participants to explore the environment simultaneously, it becomes possible to imagine a system in which each person's movement and interaction gently contribute to an evolving soundscape. These auditory layers, shaped by user locations, interaction histories, and proximities, could give rise to patterns that are not fully predictable but still grounded in the underlying logic of the artwork.

This approach remains exploratory. While it may encourage moments of social legibility and collaborative interpretation, it also raises open questions: How could one user interaction shape the experience of another? Could overlapping sonic traces promote dialogue, or introduce distraction? These questions are not easily answered and will require careful testing. Future iterations might experiment with proximity-based audio blending, or lightweight auditory cues that respond to density and movement, in order to explore how collective behavior subtly shifts the listening field. Ultimately, the aim is to invite co-authorship without losing the reflective, individualized quality central to the original scroll-viewing experience.

6.2.2 Cultural Specificity vs. Generalizability

This project draws heavily from the logic and aesthetics of Chinese hand-scrolls, whose viewing conventions are deeply rooted in specific cultural, historical, and artistic traditions. Although the system was developed with deep respect for this lineage, more work is needed to better understand its limitations, particularly in terms of generalizability. It remains to be seen whether the interactive

techniques used here can meaningfully support other forms of narrative art without flattening their cultural nuances.

This work will proceed in close collaboration with curators, heritage professionals, and cultural historians to avoid flattening distinct traditions. Issues of translation, both literal and conceptual, must be addressed: How might localized soundscapes be meaningfully adapted to different audiences or languages? What curatorial judgments are involved when augmenting sacred or historically loaded materials? The goal is not to build a one-size-fits-all framework, but to evolve a flexible, respectful methodology for cultural interpretation through sound and movement.

Future exploration will involve thoughtful comparisons with other traditions, such as narrative frescoes, friezes, or panoramic compositions. These experiments will be undertaken with caution and in close consultation with heritage scholars and cultural experts, with the goal of identifying what aspects of the current system might translate and what should be rethought entirely. Rather than assuming universality, this work acknowledges the risk of oversimplification and approaches each new application as a site for learning. Through this process, the project hopes to contribute not a fixed model, but a set of evolving practices for culturally sensitive interaction design.

6.2.3 Expert vs. Lay Engagement

This prototype was designed primarily with lay audiences in mind, emphasizing accessibility, sensory engagement, and spatial intuition. The goal was to offer a form of encounter that invites curiosity and affective connection, even for those with little prior knowledge of Chinese scrolls or art history. By translating visual narrative into spatialized sound and movement, the system supports an intuitive entry point into a complex tradition.

Yet this emphasis on accessibility also raises a critical question: how might domain experts—such as curators, art historians, or scholars of Chinese visual culture—interpret and evaluate such an experience? This is not simply a matter of assessing accuracy or pedagogical utility. Rather, it points to deeper questions about what forms of knowledge and heritage are surfaced through immersive media, and which are obfuscated in the process. It reveals a vital tension at the heart of heritage-oriented immersive design: the desire to innovate without compromising cultural specificity or scholarly depth.

Looking ahead, this suggests a broader opportunity—not only to adapt the system for expert audiences, but to design with them. Future iterations could emerge from co-creative processes

involving curators, conservators, and cultural historians, whose insight might shape how sound is mapped to visual or symbolic elements, how metadata is layered, or how historical context is spatialized. Such collaborations would move beyond merely validating the system's fidelity to heritage materials, and instead explore how immersive tools might expand, challenge, or nuance expert understandings of these works.

Acknowledgements

As spring grasses bloom across campus and this study chapter at Carnegie Mellon draws to a close, I find myself filled with deep gratitude for the many people who have accompanied me on this journey.

First, I express my sincere thanks to my thesis advisor, Professor Daragh Byrne. His generous mentorship has guided me not only in how to conduct research and write critically but also in how to read deeply, engage dialectically, and expand my understanding of what it means to design with sound. He has taught me to seek depth in soundscape design and to remain open to its many unexplored possibilities. His steady support has been a source of strength throughout this process.

I am also profoundly grateful to Professor Golan Levin, whose mentorship has brought extraordinary inspiration and color to my CMU experience. From the ingenuity of the Experimental Capture projects to the creative breakthroughs that emerged in every thesis discussion, his ideas have consistently lifted my spirits and sparked new directions. His guidance helped me consider not only the conceptual but also the expressive and experiential dimensions of my work. At times when I was lost in complexity, his insight always illuminated the path ahead.

I extend my thanks to all the MSCD faculty, including Daniel Cardoso Llach, Vernelle A. A. Noel, Joshua D. Bard, Jimmy Wei-Chun Cheng, and Omar Khan, etc. Your teaching and feedback, whether during everyday conversations, prethesis critiques, or presentations, have challenged me to think rigorously, design more thoughtfully, and grow with each step.

To my peers in the MSCD community, thank you for your companionship, encouragement, and collaboration — especially during long nights of writing and rounds of user testing. I will always remember the Wednesday afternoon research meetings, the delicious milestone meals, and the joyful Codelab events that made this program feel like home.

I am deeply grateful to my parents, whose unwavering support and belief in me at every stage of life have given me the courage and confidence to continue moving forward.

Thanks to all who supported my ideas, shared my questions, and celebrated the small breakthroughs along the way.

References

[1] Horizontal scroll: Thousand miles of mountains and rivers. Accessed: 2025-02-25.

[2] Wikipedia, 2024.

[3] Artlist. Natural ambiences sound effects pack, 2025. Accessed: 2025-03-03.

[4] F. Behrendt. The sound of locative media. *Convergence: The International Journal of Research into New Media Technologies*, 18(3):283–295, 2012.

[5] N. Blumberg. linear perspective, 2024.

[6] Michael Bull. *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Materializing Culture. Berg Publishers, Oxford, UK, illustrated, reprint edition, 2000.

[7] Janet Cardiff. Her long black hair, 2004.

[8] David Carrier. *The psychology of perspective and Renaissance art: Brunelleschi's Peepshow and The Invention of Perspective*. Columbia University Press, 1988.

[9] Ishan Chatterjee, Robert Xiao, and Chris Harrison. Gaze+gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, page 131–138, New York, NY, USA, 2015. Association for Computing Machinery.

[10] Lam Meng Chun, Haslina Arshad, Thammathip Piumsomboon, and Mark Billinghurst. A combination of static and stroke gesture with speech for multimodal interaction in a virtual environment. In *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*, pages 59–64, 2015.

[11] L. Cliffe. *Audio Augmented Objects and the Audio Augmented Reality Experience*. PhD thesis, University of Nottingham, 2022. December 31, 2022.

[12] Laurence Cliffe, James Mansell, Joanne Cormac, Chris Greenhalgh, and Adrian Hazzard. The audible artefact: Promoting cultural exploration and engagement with audio augmented reality. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, page 176–182, New York, NY, USA, 2019. Association for Computing Machinery.

[13] Policarpo del Canto Baquera. Interfacing the multiplexer room : Speculative spatio-mediated assemblages for the digitally-interfered home. Master's thesis, CMU, 10 2022.

[14] Earth.fm. Nature soundmap, 2025. Accessed: 2025-03-03.

[15] Valdemar Erlingsson. extOSC: Open Sound Control for Unity, 2024. Accessed: 2024-02-20.

[16] William W. Gaver. Auditory icons: Using sound in computer interfaces. *Human–Computer Interaction*, 2(2):167–177, 1986.

[17] J. T. Hansberger, C. Peng, V. Blakely, S. Meacham, L. Cao, and N. Diliberti. A multimodal interface for virtual information environments. In J. Y. Chen and G. Fragomeni, editors, *Virtual, augmented and mixed reality. Multimodal interaction*, pages 59–70. Springer international publishing, 2019.

[18] Adrian Hazzard, Jocelyn Spence, Chris Greenhalgh, and Sean McGrath. The rough mile: Reframing location through locative audio. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, AM '17, New York, NY, USA, 2017. Association for Computing Machinery.

[19] A. Herruzo and N. Pashenkov. 'what i see is what you get' explorations of live artwork generation, artificial intelligence, and human interaction in a pedagogical environment. In A. Brooks and E. I. Brooks, editors, *Interactivity, Game Creation, Design, Learning, and Innovation*, pages 343–359. Springer International Publishing, 2020.

[20] National Museums Northern Ireland. Unlocking our sound heritage, 2019. Accessed: 2025-02-09.

[21] A. Jaimes and N. Sebe. Multimodal human–computer interaction: a survey. *Computer Vision and Image Understanding*, 108(1-2):116–134, 2007.

[22] Jaeneen K. Janet cardiff & george bures miller, 2012.

[23] Mixkit. Free nature sound effects, 2025. Accessed: 2025-03-03.

[24] Pallavi Mohan, Wooi Boon Goh, Chi-Wing Fu, and Sai-Kit Yeung. Head-fingers-arms: Physically-coupled and decoupled multimodal interaction designs in mobile vr. In *Proceedings of the 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, New York, NY, USA, 2019. Association for Computing Machinery.

[25] Nils Montan. Aar an audio augmented reality system. Master's thesis, Royal Institute of Technology, 2003.

[26] The Palace Museum. Digital collection of the palace museum, 2024. Accessed: 2025-02-03.

[27] Chao Peng, Jeffrey T. Hansberger, Lizhou Cao, and Vaidyanath Areyur Shanthakumar. Hand gesture controls for image categorization in immersive virtual environments. In *2017 IEEE Virtual Reality (VR)*, pages 331–332, 2017.

[28] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, page 99–108, New York, NY, USA, 2017. Association for Computing Machinery.

[29] Roberto Pierdicca, Emanuele Frontoni, Primo Zingaretti, Mirco Sturari, Paolo Clini, and Ramona Quattrini. Advanced interaction with paintings by augmented reality and high resolution visualization: A real case exhibition. In *Proceedings of the Second International Conference on Augmented and Virtual Reality - Volume 9254*, page 38–50. Springer-Verlag, Berlin, Heidelberg, 2015.

[30] Andreas Schneider, Daniel Cernea, and Achim Ebert. Hmd-enabled virtual screens as alternatives to large physical displays. In *2016 20th International Conference Information Visualisation (IV)*, pages 390–394, 2016.

[31] Shouqian Shi. *Mountain Echoes: Chinese Landscape Painting and the History of the Viewer*. Shanghai Calligraphy and Painting Publishing House, 2019.

[32] Marjan Sikora, Mladen Russo, Jurica undefinederek, and Ante Jurčević. Soundscape of an archaeological site recreated with audio augmented reality. *ACM Trans. Multimedia Comput. Commun. Appl.*, 14(3), July 2018.

[33] T. Thiel. Critical interventions into canonical spaces: Augmented reality at the 2011 venice and istanbul biennials. In V. Geroimenko, editor, *Augmented Reality Art: From an Emerging Technology to a Novel Creative Medium*. Springer, 2014.

[34] Aneta Trajkoski. The shifting spaces of janet cardiff and george bures miller's 'berlin files'. *Burlington Contemporary*, 9, November 2023.

[35] Unity Technologies. *AudioSource*, 2024. Accessed: 2025-02-25.

[36] Unknown. A thousand li of rivers and mountains, n.d. Accessed: 2025-02-25.

[37] Daniel Vogel and Ravin Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, page 137–146, New York, NY, USA, 2004. Association for Computing Machinery.

[38] Theo Watson. Audio space, n.d. Accessed: 2025-02-09.

[39] We AR in MoMA. We ar in moma, 2024. Accessed: 2025-02-03.

[40] Jing Yang, Yves Frank, and Gábor Sörös. Hearing is believing: Synthesizing spatial audio from everyday objects to users. In *Proceedings of the 10th Augmented Human International Conference 2019*, AH2019, New York, NY, USA, 2019. Association for Computing Machinery.

[41] Jing Yang and Friedemann Mattern. Audio augmented reality for human-object interactions. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, page 408–412, New York, NY, USA, 2019. Association for Computing Machinery.

[42] Nan Zhang and Yuhang Kong. Garden patterns and spatial transformation in landscape painting—three hanging scrolls from the northern song dynasty as examples. *Chinese Landscape Architecture*, 38(8):66–71, August 2022.

[43] Yanyuan Zhang. *Notes on Famous Paintings of the Past Dynasties*. 1986.

[44] Andreas Zimmermann and Andreas Lorenz. Listen: a user-adaptive audio-augmented museum guide. *User Modeling and User-Adapted Interaction*, 18:389–416, 2008.