From Temporal to Spatial: Designing Spatialized Interactions with Segmented-audios in Immersive Environments for Active Engagement with Performing Arts Intangible Cultural Heritage

Yuqi Wang*
yuqiw.wang@mail.utoronto.ca
Ontario Institute for Studies in Education University of Toronto
Toronto, Canada

Sirui Wang*
sw5546@nyu.edu
New York University
New York, United States

Shiman Zhang
s2658776@ed.ac.uk
University of Edinburgh
Edinburgh, United Kingdom

Kexue Fu
kexuefu2-c@my.cityu.edu.hk
Studio for Narrative Spaces
City University of Hong Kong
Hong Kong, China

Michelle Lui
michelle.lui@utoronto.ca
Ontario Institute for Studies in Education University of Toronto
Toronto, Canada

RAY LC†
LC@raylc.org
Studio for Narrative Spaces
City University of Hong Kong
Hong Kong, China

Figure 1: Traditional Performing arts theatre experience versus screen-based video watching experience versus Spatial Interaction-based Segmented-Audio (SISA) prototype experience. (Left) Traditional performance, though sensory engaging and immersive, requires physical attendance at theatre venues, extended passive listening, and relies on conventional practitioner enactment, highlighting fundamental constraints in ICH engagement and preservation. (Middle) Screen-based videos provides temporal flexibility but still lacks immersive engagement essential for authentic cultural appreciation. (Right) Our SISA prototype enables active audience participation through spatial interactions with segmented audio elements in an immersive environment.

ABSTRACT

Performance artforms like Peking opera face transmission challenges due to the extensive passive listening required to understand their nuance. To create engaging forms of experiencing auditory Intangible Cultural Heritage (ICH), we designed a spatial interaction-based segmented-audio (SISA) Virtual Reality system that transforms passive ICH experiences into active ones. We undertook: (1) a co-design workshop with seven stakeholders to establish design requirements, (2) prototyping with five participants to validate design elements, and (3) user testing with 16 participants exploring Peking Opera. We designed transformations of temporal music into spatial interactions by cutting sounds into short audio segments, applying t-SNE algorithm to cluster audio segments spatially. Users navigate through these sounds by their similarity in audio property. Analysis revealed two distinct interaction patterns (Progressive and Adaptive), and demonstrated SISA's efficacy in facilitating active auditory ICH engagement. Our work illuminates the design process for enriching traditional performance artform using spatially-tuned forms of listening.

CCS CONCEPTS

• **Applied computing** → **Arts and humanities**.

KEYWORDS

Intangible Cultural Heritage (ICH), Virtual Reality (VR), Spatial Audio, t-SNE, Active Engagement

*Equal contribution.
†Correspondences should be addressed to LC@raylc.org.

DIS '25, 5th and 9th of July 2025, Funchal, Madeira

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Yuqi Wang, Sirui Wang, Shiman Zhang, Kexue Fu, Michelle Lui, and RAY LC. 2025. From Temporal to Spatial: Designing Spatialized Interactions with Segmented-audios in Immersive Environments for Active Engagement with Performing Arts Intangible Cultural Heritage. In *ACM SIGCHI Conference on Designing Interactive Systems (DIS '25), 5-9 July 2025, 2025, Funchal, Madeira.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Much of the richness of our culture comes from sources that cannot be easily documented or quantified, because they are not buildings or objects, but rather cultural practices that define us as a community. Traditional performing arts as a form of Intangible Cultural Heritage (ICH) [60], including music, dance, theatre, pantomime and beyond, is an example. UNESCO recognizes these cultural practices as "invaluable and irreplaceable sources of life and inspiration" [60]. However, ICH performing arts face significant threats due to weakened practice and transmission [5, 9, 64, 70]. These threats are intensified by challenges in sustainable generational transmission, as traditional ICH oral instruction methods [62] face increasing difficulty in maintaining continuous practice and knowledge transfer across generations in rapidly modernizing societies. As a consequence, some ICH performing arts events, such as Kun Qu Opera [61] and Peking Opera[62], that were once frequent and widely attended now occur rarely[63].

While museums strive to safeguard cultural heritage, they primarily focus on tangible artifacts, such as historical buildings, artworks, photos and models [27], often struggling to capture ICH's "living and ever-changing" nature [65] and integrate it effectively into their activities [49]. With ICH performing arts facing a steep decline, the sustainability challenges in preserving their auditory component are particularly significant, as traditional preservation methods require extensive resources and location-dependent engagement. Traditional methods of preserving and presenting audio ICH often lead to passive listening experiences for audiences. These methods typically involve extended periods of listening time of over two hours, demanding full attention of the audience without providing engaging interaction in return. The inherently passive, temporal, and linear nature of these artforms often fails to engage modern audiences and younger generations, who have shorter attention spans and prefer interactive, self-expressive content [36]. As a result, they often struggle to meaningfully engage with, understand, and contextualize the heritage value of these artforms [8, 11].

Current research have popularized the use of Virtual Reality (VR) to enhance engagement with audio-related ICH artforms [3, 10, 14, 32, 34, 40, 50, 53, 59, 75, 76]. However, those approaches still adhere to the temporal and linear nature of auditory ICH, often requiring long periods of passive listening. Studies have shown that compared to passive listening, active listening significantly improves engagement[33, 54]. Some recent studies [15, 16, 20] provide an inspiring approach to interactive listening experiences. These studies utilized machine learning algorithms, such as t-SNE, to transform temporal audio experiences into active interactions in VR 3D spaces [20] and interactive spatial online audio interfaces[15, 16]. By clustering audio segments based on their acoustic features, these approaches attempt to make audio exploration more tangible, visually engaging, and spatially interactive. However, novel approaches to interactive listening experiences remain unexplored in the ICH domain.

Inspired by these technological advances and the challenges in ICH engagement, we propose a new design methodology that aims to transform passive temporal auditory ICH experiences into active ones, through our Spatial Interaction-based Segmented-Audio (SISA) prototype. Our approach segments ICH audio recordings into 5-second segments and employs t-SNE algorithm [66] to cluster these segments spatially by their acoustic features, creating navigable 360-degree VR environments that may enable new forms of auditory ICH engagement. Our research examines both the theoretical considerations of spatial-temporal transformation in interaction design and its practical applications for ICH engagement, leading to two research questions:

RQ1: *How do we design engaging auditory ICH interactions utilizing the spatial properties of immersive experiences?*

RQ2: *What patterns of interaction emerge when users listen to auditory ICH audio segments that are mapped into spatial experiences?*

To address our research questions, we structured our investigation into three sequential phases, each building upon insights from the previous phase to systematically design, develop and refine the SISA system. Phase 1 and 2 uncovered critical design considerations for developing spatial transformations of temporal audio (RQ1), while Phases 3 examined how users experience and interact with spatially transformed auditory ICH content in SISA prototypes (RQ2). Refer to Methods section and Figure 2 for the detailed breakdown of the three phases. Phase 1 employed a participatory co-design approach with seven stakeholders representing diverse perspectives: one designer, one developer, one ICH practitioner, two researchers, and two user representatives. Through collaborative co-design sessions, we identified design rationale and critical design elements for transforming linear audio experiences into interactive spatial arrangements. Phase 2 consisted of testing prototyping with five participants to validate and refine the design elements identified in Phase 1. This exploratory investigation examined the technical feasibility of Prototype 1 and informed subsequent system refinements. Phase 3 comprised comprehensive user testing with 16 participants to investigate interaction patterns and user experiences of the final prototype. Analysis revealed two distinct interaction patterns: Progressive and Adaptive, demonstrating SISA's efficacy in promoting active ICH engagement.

While traditional ICH transmission faces constraints of resource intensivity and location dependency, our temporal-to-spatial SISA methodological framework offers a new accessible, interactive and sustainable way to experience ICH performing arts experiences. The SISA approach establishes a design methodology that reconceptualizes interactive auditory ICH experiences beyond conventional approaches. Our work makes significant societal contributions by contributing both theoretical frameworks for spatial audio interaction design and guidelines for creating sustainable, accessible and engaging auditory ICH experiences to preserve and present "the past"[60]. This work illuminates methodologies for enriching traditional performance artforms through spatially-tuned forms of listening.

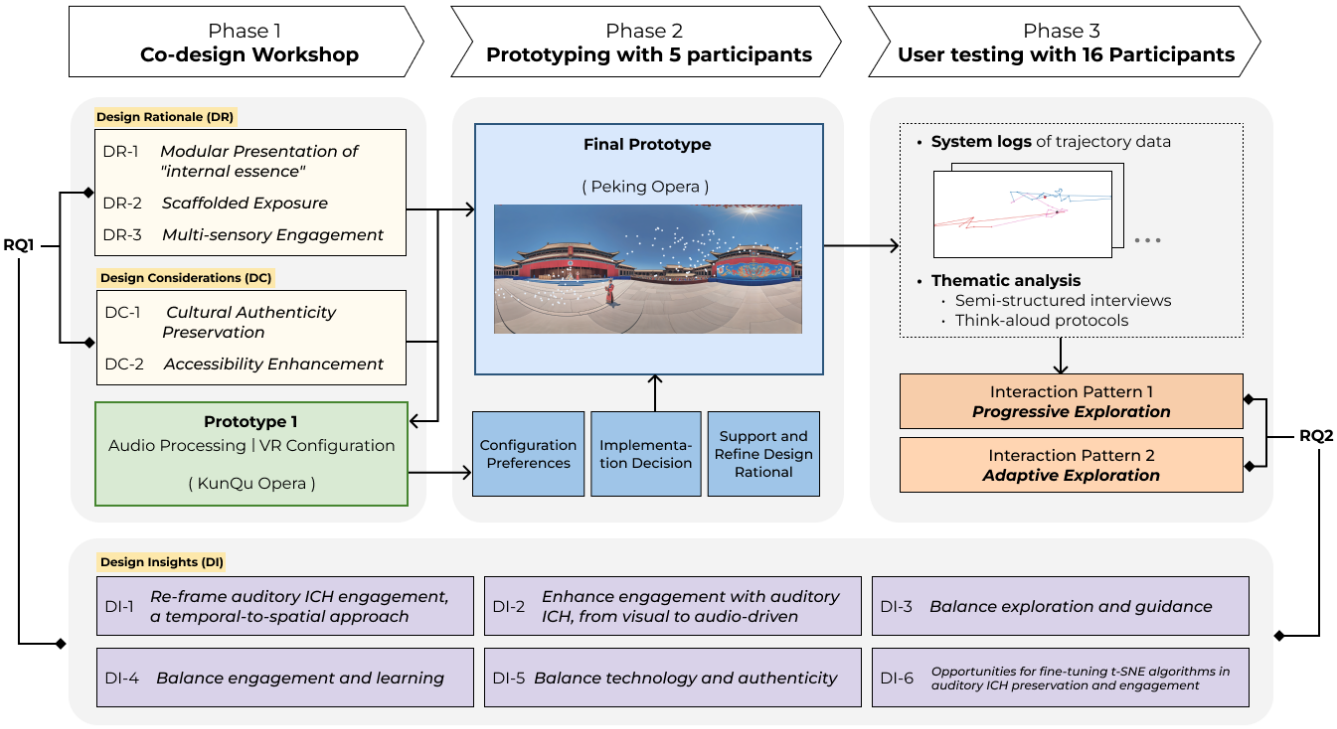

Figure 2: A detailed break down of our project phases.

2 BACKGROUND

2.1 Intangible Cultural Heritage and Performing Arts

The performing arts is one of the five domains of ICH. They include both visual and audio components, encompassing traditional theater performances that combine acting, singing, dance, music, dialogue, narration, recitation, puppetry, and pantomime [60]. These artforms shape cultural identities and enhance social cohesion [17]. Despite their significance, these cultural expressions face numerous modern challenges, such as aging practitioners, diminishing youth interest, hampered transmission, and reduced practice [64]. Traditional performing arts particularly illustrate these threats. For example, Kun Qu Opera and Peking Opera's vitality are eroding as they rely heavily on master-student relationships for transmission through oral instruction, observation, and imitation [62]. The preservation challenge is further complicated by evolving audience preferences and consumption patterns. Modern audiences, accustomed to interactive and fast-paced digital entertainment, often find traditional performances less engaging [36].

Recognizing the urgency of these challenges, international organizations have developed strategies to safeguard ICH. UNESCO's approach, in particular, advocates for transforming ICH from "fragile" to "truly alive" [60], encouraging the development of global appreciation and active engagement in these cultural practices. UNESCO suggests media, institutions, and cultural industries playing critical roles in developing audiences, raising public awareness, and promoting active participation in cultural expressions [60]. However, despite these efforts, auditory ICH remains challenging to engage with actively due to its traditionally passive, linear, and time-bound nature.

2.2 The Limitations of Passive Reception of Traditional, Linear, Time-bound Auditory ICH

Auditory ICH traditionally limits listeners to passive reception due to its linear, time-bound nature, constraining active exploration and deep engagement. This passive approach, akin to listening to an unstructured lecture without cognitive engagement, differs significantly from interactive experiences like making music or engaging in dialogue. This distinction between passive and active engagement with auditory ICH can be further understood through Schaeffer's four modes of listening: écouter (listening to identify sources), ouïr (passive perception), entendre (selective hearing), and comprendre (understanding meaning) [55]. Traditional ICH experiences often confine audiences to ouïr mode—passive perception without active interpretation. The contrast between passive and active listening lies in the level of engagement and the listener's agency to interact with manageable information within their attention span. Neuroscientific research underscores the importance of active engagement in auditory processing [33, 52, 54]. Research has

shown that active-response listening significantly enhances cognitive engagement and neural activity compared to passive listening [54]. The advantages of active engagement extend beyond auditory stimuli. A meta-analysis of 128 effect sizes from 33 experiments revealed a small to moderate advantage of active exploration over passive observation in spatial knowledge acquisition [52]. This research suggests that hands-on, experiential, interactive approaches to experiencing ICH, rather than passive reception, could be beneficial for understanding and remembering nuances of culture. These results align with the ICAP (Interactive, Constructive, Active, Passive) framework[13], which posits that learning outcomes improve as engagement progresses from passive to interactive modes. If applied to ICH, this framework could suggest the development of more effective content delivery methods, particularly for modern audiences. Collectively, these studies emphasize the need to move beyond mere documentation and archiving of auditory ICH. By promoting carefully designed interactive experiences, we can potentially enhance engagement levels and facilitate the transmission of cultural nuances, particularly among modern audiences. Our aim is to develop a new approach that addresses the limitations of traditional, passive methods of engaging with auditory ICH, converting the traditional passive reception of auditory ICH into active interaction experiences.

2.3 Current Practices of ICH engagement

We examined current technological practices in ICH engagement and preservation, focusing on audio-related aspects, given the potential of interactive experiences to enhance ICH engagement and preservation. Contemporary ICH engagement and preservation practices increasingly adopts advanced digital technologies, including Virtual Reality (VR) [3, 10, 14, 22, 32, 34, 50, 53, 59, 75, 76], Augmented Reality (AR) [39, 57, 58], Generative Artificial Intelligence (GenAI) [23, 26, 38, 71], interactive physical devices [48], live streaming technologies [43], and parametric expression techniques [53] to enhance public understanding and engagement with ICH by improving accessibility and interactivity. VR has demonstrated significant advantages in ICH interaction and engagement. Its strength lies in creating immersive, spatially interactive experiences that enhance learning and meaning-making in cultural heritage contexts [10, 14, 59, 75].

Audio-related ICH engagement and preservation practices are evolving but still limited. While basic preservation methods [47] and digitization efforts [4] are common, researchers are exploring more advanced technologies. These include multimodal VR presentations for aural heritage [32], and interactive systems like Virtual Guqin and Mixed Reality Guqin for granting greater accessibility and interest building for traditional musical instruments [74].

In ICH performing arts, various technological approaches are being applied. For dance movements, a VR application designed to enhance the demonstration of dance formations through the FormationCreator system enables real-time modifications via voice commands in a virtual environment [53]. VR models are also being used for capturing dance moves and rhythms [50]. In the realm of songs and music, an interactive Virtual Reality (VR) storytelling project explores the Chinese Hua'er "Baxiguliuliu" song [40]. For puppetry, the ShadowStory project uses digital narratives to promote creativity in Chinese shadow puppetry, incorporating both visual and audio elements [42]. However, for the auditory components of these technology-enhanced ICH performing arts practices, these approaches still adhere to the temporal, linear, and time-bound nature of auditory ICH and passive reception approach, indicating a need for more interactive and engaging auditory experiences.

Recent work around interactive listening experiences [15, 16, 20] offers an inspiring approach and can be applied for auditory ICH transmission and preservation purposes. These studies have utilized machine learning algorithms, such as t-SNE, to transform temporal audio experiences into active interaction experiences. For example, Cowen et al. created interactive online audio interfaces using t-SNE to visualize the complexities of emotions conveyed by brief human vocalizations [15] and music-associated subjective experiences [16]. Similarly, the Sound Of(f) project employed t-SNE to cluster and map sounds into a VR 3D space, allowing audiences to interact freely with audio segments in a virtual reality setting [20]. These works are inspiring and lead us to investigate the current audio processing practices and techniques to transform temporal audio experiences into active interaction experiences for auditory ICH engagement and preservation.

2.4 Audio Processing Practices

Dimensionality reduction techniques are essential in audio signal processing to simplify complex audio data and enable more accessible interpretation and visualization. A number of research [19, 21, 41] has focused on mapping high-dimensional audio data onto lower-dimensional spaces to facilitate both visual and statistical evaluation. The typical implementation pipeline involves feature extraction, reshaping, and dimensionality reduction of audio samples. Mel-Frequency Cepstral Coefficients (MFCCs) are widely used for feature extraction, while various dimensionality reduction techniques, including PCA, t-SNE, UMAP, and Isomap, have been evaluated for performance. Among these approaches, the combination of MFCCs for feature extraction and t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction has proven particularly effective for analyzing audio signals [51]. Studies adopting t-SNE technical approaches for interactive listening experiences [15, 16, 20] inspired us to design and develop an interactive system for transforming passive, linear auditory ICH experiences into active and interactive ones.

3 METHODS

To systematically investigate the design and development of interactive systems for auditory ICH engagement, we employed an iterative research methodology grounded in participatory design and prototype-based inquiry. Our work resulted in the Spatial Interaction-based Segmented-Audio (SISA) system, which represents the first application of machine learning algorithms to enhance ICH engagement within immersive VR environments.

Drawing from participatory design approaches and iterative prototyping methodologies, we structured our investigation into three interconnected phases to systematically develop and evaluate the SISA system. Each phase built upon insights from the previous one, enabling us to refine both interaction design elements and the

underlying technical implementation. We specifically focused on (1) understanding the design considerations for transforming temporal ICH audio into spatial interactions; (2) examining how users navigate and make sense of spatially-structured ICH audio content in the immersive environment; and (3) exploring the broader implications of SISA approach for ICH engagement, transmission and preservation in digital age. These objectives led to two primary research questions:

RQ1: *How do we design engaging auditory ICH interactions utilizing the spatial properties of immersive experiences?*

RQ2: *What patterns of interaction emerge when users listen to auditory ICH audio segments that are mapped into spatial experiences?*

We initiated the design process with a participatory co-design workshop in Phase 1, bringing together seven stakeholders representing diverse perspectives: one Human-Computer Interaction (HCI) designer, one VR developer, one ICH practitioner, two HCI researchers, and two user representatives. Throughout Phase 1 co-design workshop, we identified critical design rationale and design elements and created Prototype 1 using it as "an experimental component" [69]. This prototype served as a vehicle for inquiry [69], enabling stakeholders to explore and articulate design considerations for transforming linear audio experiences into interactive spatial arrangements. Phase 2 study utilized prototype 1 as both a "technology probe" [28] and a "provotype" [46] to validate and refine the design elements identified in Phase 1 with 5 participants. This exploratory investigation helped verify our design rationale and examine the technical feasibility of implementing spatial audio interaction mechanisms. The findings from this study directly informed system refinements and interaction design decisions for the final prototype. In Phase 3, we presented the final prototype as a research archetype [69] to conduct comprehensive user testing with 16 participants. Participants explored Peking Opera VR scene while researchers collected data through think-aloud protocols[12], semi-structured interviews, and system interaction logs. This phase focused on understanding how users navigate and make sense of spatially arranged ICH audio content in practice. Throughout this process, which served as a vehicle for inquiry [69], we documented, analyzed, and critically assessed our findings, focusing on research contributions tied to the design and development process rather than just the artifact itself [69]. The resulting design insights are summarized in the discussion section of this paper. Figure 2 illustrates the systematic progression of our three-phase design methodology, which is elaborated in subsequent sections. This study was approved by and conducted according to the guidelines of the University Institutional Review Board.

4 PHASE 1 STUDY: CO-DESIGN OF SISA SYSTEM

In Phase 1, we initiated our investigation with a participatory co-design workshop, bringing together seven diverse stakeholders to explore the transformation of traditional ICH audio experiences into interactive spatial arrangements. This section details our co-design process, including the participants, workshop formats, and the resultant design framework that informed SISA system development. Through stakeholder engagement in the ideation process, we established the SISA approach that balances technical innovation with cultural preservation while prioritizing user engagement.

4.1 Participants

We recruited seven participants representing diverse stakeholder perspectives in the ICH and technology domains through professional networks. The participants included: one designer with experience in cultural heritage projects, one developer specializing in VR applications, one ICH practitioner with expertise in Peking Opera and Kunqu Opera, two researchers in HCI and ICH preservation, and two user representatives with varying levels of familiarity with traditional Chinese performing arts. This diverse group was intentionally selected to bring multiple perspectives to the design process, particularly balancing technical feasibility with cultural authenticity. Details see Table 1.

4.2 Workshop Format

The co-design workshop was structured as a three-hour intensive session employing multiple participatory design techniques to explore the transformation of linear ICH audio experiences into interactive spatial arrangements. The workshop includes a 15-min introduction, a 25-min open discussion, a 60-min design exploration discussion, a 30-min synthesis and refinement session, and a 30-min closing discussion. During introduction, participants were briefed on the project goals and signed consent forms. The context about current challenges in ICH engagement and preservation were provided. Then, an open discussion, employing contextual inquiry[31], was led by two researchers who ensured balanced participation and comprehensive documentation of emerging themes. This session explored fundamental questions about ICH engagement, cultural preservation, and technological intervention. This semi-structured dialogue allowed stakeholders to share their diverse experiences and perspectives on current ICH engagement methods, establishing a rich foundation for subsequent design activities. Next in the 60-minute design exploration session, participants engaged in collaborative ideation focusing on defining design goal, rationale, considerations, and potential approaches for enhancing auditory ICH engagement. To supplement the contextual inquiry discussions, we conducted case study analysis, where participants examined existing ICH preservation initiatives and interactive systems to identify effective design elements and potential implementation challenges. The case studies included traditional performance documentation methods[62], existing ICH applications[3, 40, 42, 71, 73, 74], and emerging spatial audio technologies[15, 16, 20], providing a comprehensive contextual foundation for ideation. Furthermore, the group worked together to consolidate ideas and establish key design elements for the SISA system and Prototype 1.

4.3 Data Collection and Analysis

The entire workshop was audio-recorded and later transcribed. Additional data sources included researchers' field notes documenting key discussion points and emerging themes, photographs of sketches and diagrams created during ideation sessions, and written feedback from participants collected at the workshop's conclusion. Our analysis followed a rigorous qualitative approach based on

Co-Design Participant ID	Occupation	Relevance to the Study
CoD1	Designer with 1 year experience in cultural heritage projects	Focused on digital preservation of Chinese artifacts and contributed insights on integrating cultural elements into the design specifically regarding visual aesthetics, symbolic movements, and performance conventions unique to Chinese opera traditions.
CoD2	Developer specializing in VR applications for cultural exhibition contexts	Provided technical feasibility perspectives for immersive technology integration.
CoD3	ICH practitioner with expertise in Peking Opera and Kunqu Opera	Offered detailed knowledge on traditional performing arts for cultural authenticity.
CoD4	Researcher in HCI	Focused on user-centered design principles and interaction patterns.
CoD5	Researcher in HCI and education	Provided insights on instructional design, emphasizing effective content delivery.
CoD6	User representative with familiarity with traditional Chinese performing arts	Contributed as a knowledgeable end-user, highlighting engagement factors.
CoD7	User representative with limited exposure to traditional Chinese performing arts	Provided a perspective on accessibility and initial impressions.

Table 1: Demographic Information of Phase 1 Co-design Workshop Participants (N=7)

Braun and Clarke's thematic analysis methodology[7]. The emerging themes were developed and refined through two rounds of review.

4.4 Phase 1 Study Results: Design Rationale, Considerations and Approach

Through systematic analysis of co-design workshop data, we developed a comprehensive design framework encompassing high-level design goals, underlying rationale, specific considerations, design approach and their implementation in the SISA system design and Prototype 1.

4.4.1 Design Goal. Traditional ICH audio content is typically experienced through passive, linear listening, limiting engagement and potentially hindering cultural transmission. Our core focus is on creating more active and engaging ways for users to experience and listen to ICH audio content. As CoD3 articulated:

> *"The goal is to inspire more people to listen to these ICH genres...not just within the Chinese community, who already have a deeper understanding of our art...but also to communicate these artistic values to non-Chinese speakers." (CoD3)*

This insight underscores the need for more accessible and engaging approaches to ICH audio content presentation. Our primary design goal is to enhance auditory ICH engagement by transforming passive linear temporal experiences into interactive spatial arrangements within immersive environments. This transformation aims to (1) enable active exploration of ICH audio content; and (2) foster deeper engagement through spatial-temporal interactions. This approach should create a more engaging pathway for users to listen and explore auditory ICH content while supporting ICH preservation and transmission purposes.

4.4.2 Design Rationales. Three fundamental rationales emerged from our co-design discussions and practitioner insights: (1) Modular Presentation; (2) Scaffolded Exposure; (3) Multi-sensory Engagement.

(DR-1) Modular Presentation of "internal essence". ICH practitioner CoD 3 emphasized the modular nature of performing arts, noting that *"each performance genre has its own internal essence, modules and grains...own formula and convention"*. This revealed the inherent modularity in traditional performances. CoD 5 suggested a pedagogical approach through *"breaking the genre into modules and components, and showing what each module means"*. This was echoed by CoD 3, *"when people accumulate this knowledge and listening experience, it becomes easier for them to understand and appreciate the performance genre"*. CoD 3 further suggested enabling SISA system to create what practitioners describe as a *"symbol system that allows better communication"*, and to maintain cultural authenticity while facilitating engagement. Hence, we aim to design and develop SISA system to decompose complex cultural elements into discrete, learnable modules while maintaining their interconnections. This can be achieved by applying t-SNE algorithm to cluster audio segments spatially by their similarity in audio property. This approach should facilitate non-linear exploration of ICH listening experience while preserving the authentic features of the artform.

(DR-2) Scaffolded Exposure. Stakeholders suggested considering the cognitive load management of users when them experiencing the cultural contexts and content in the system. Both researchers and practitioners emphasized the need for carefully calibrated complexity levels in cultural content presentation, emphasizing the importance of structured initial exposure and gradual deepening of engagement. As researcher CoD4 emphasized, *"We need to start from the surface level, making the system more accessible, interactive, and fun first"*, while researcher CoD5 advocated for gradual information disclosure: *"Don't overwhelm them with too much information. Let the exploration journey progress bit by bit."* This perspective was strongly reinforced by designer insights, with CoD1 noting:

> *"You can't start by explaining deep theories, everyone would find that boring, especially young people. There are too many things to see in the world today - why would people dedicate time to this?" (CoD1)*

Drawing from personal experience, CoD3 illustrated this scaffolded exposure principle: *"I started learning Kun Qu and Peking Opera because I loved the visual elements - the beautiful costumes, the makeup,*

the headdresses. The music came later." These converging perspectives from multiple stakeholders underscore the importance of implementing a scaffolded exposure architecture in the SISA system that strategically sequences engagement - beginning with visually appealing and immersive auditory elements before gradually introducing deeper cultural contexts and modules.

(DR-3) Multi-sensory Engagement. Stakeholders stressed the need of incorporating various sensory dimensions to create comprehensive cultural experiences. As articulated by ICH practitioner CoD 3: *"Performance arts provide many sensory impacts from hearing, vision, and the whole atmosphere. Audience will build emotional connection within this immersive environment."* User presentative CoD6 further reinforced this approach:

> *"Theater art emphasizes immersion and sensory impact. The experience is completely different from watching on a computer screen... Even if you say 'I don't understand what they're singing or what they're saying,' for complete beginners, there's still a lot of sensory information...Music itself carries emotions... lyrics are just one aspect." (CoD6)*

Hence, we design our SISA system to combine visual, auditory, and interactive elements to provide the rich sensory and immersive experience of traditional performances in the VR environment.

4.4.3 Design Considerations. Several key considerations emerged from our co-design process that shaped the SISA system's design: (1) Cultural Authenticity Preservation; (2) Accessibility Enhancement.

(DC-1) Cultural Authenticity Preservation: The ICH practitioner underscored the critical importance of preserving cultural authenticity, describing traditional opera as *"a living fossil of our culture, allowing us to glimpse historical perspectives and cultural evolution (CoD 3)."* At the same time, CoD 3 highlighted the balance between maintaining tradition and fostering innovation, stating, *"There are things that are strictly fixed, and things that are flexible. For instance, there are disciplines and manners you have to follow, but there is room for personal creation and innovation."* Furthermore, CoD3 advocated for innovative approaches to promote engagement, transmission, and preservation within the ICH domain, noting that *"when efforts are made to discover and present the internal essence using different methods, and when it's done with dedication, that's respecting the art form."* This perspective underscores the dual necessity of respecting cultural authenticity while exploring creative and innovative pathways in the design of interactive systems for ICH engagement.

(DC-2) Accessibility Enhancement. The ICH practitioner CoD 3 identified accessibility as a fundamental challenge in traditional performance contexts, explicitly noting that

> *"The main challenge is getting people to come to the theater in the first place. Even for the overseas Chinese community, opera feels distant - they know it's national heritage and it's held in high regard, but they don't understand its internal essence (CoD3)."*

This challenge highlights a critical barrier to cultural transmission and engagement in conventional performance settings. The practitioner demonstrated strong support for innovative technological approaches to address these accessibility constraints, particularly emphasizing the potential of digital preservation and virtual engagement platforms. The practitioner also emphasized that traditional performance settings, while valuable, can present significant barriers to new audiences. CoD3 highlighted this challenge: *"Another difficulty is how to help people who don't speak Chinese understand the deeper value of our art, including its cultural significance and abstract elements."* This insight reinforces the need for innovative approaches to cultural transmission that can overcome linguistic and cultural barriers.

4.4.4 Design Approach. Drawing upon our established design goals, rationales, and considerations, we developed a systematic approach leveraging MFCC-TSNE algorithms to analyze and cluster segments of traditional performance pieces. To transform temporal performance arts music into spatially interactive listening experiences, we developed a novel method to process traditional performance arts music recordings. This approach segments lengthy audio recordings into shorter, manageable units and distributes them across a 360-degree spatial environment. Using MFCC (Mel-frequency Cepstral Coefficients) combined with t-SNE (t-Distributed Stochastic Neighbor Embedding) dimensionality reduction, we created an explorable soundscape that preserves the essence of the original performances (DR-1 and DC-1) while enabling interactive engagement with both cultural visual and auditory elements in the immersive VR environment (DR-3).

This approach aligns with Schaeffer's four modes of listening [55]- Écouter (listening to identify sources), Ouïr (passive perception), Entendre (selective hearing), and Comprendre (comprehending meaning)- facilitating users' progression through these hierarchical stages of listening. The scaffolded exposure principle (DR-2) enables users to advance from initial source identification (Écouter) toward more sophisticated comprehension (Comprendre), while multi-sensory engagement (DR-3) creates optimal conditions for developing selective hearing (Entendre) capabilities. For the remainder of this paper, we refer to this method as the Spatial Interaction-based Segmented-Audio (SISA) approach. By mapping acoustic features into a 3D virtual space using t-SNE dimensionality reduction, we create a taxonomy of soundscapes that aims to facilitate immersive exploration. This spatial organization transforms complex temporal performances into navigable spatial audio segments. It aims to enable users, especially novices, to actively and gradually discover relationships between audio segments and then identify genre-specific characteristics such as vocal techniques, instrumentation patterns, and character roles (DR-2). By shifting from passive linear listening to embodied spatial exploration, the system aims to lower cognitive entry barriers and suggests possibilities for enhancing users' affective and analytical engagement with the genre — directly addressing challenges[60] reported by practitioners in cultural transmission and accessibility (DC-2).

4.5 SISA Prototype 1

Based on previous design rationale and considerations, the SISA approach emphasizes *modular presentation of "internal essence" (DR-1)*, *scaffolded exposure (DR-2)*, *multi-sensory engagement (DR-3)*, *cultural authenticity preservation (DC-1)*, and *accessibility enhancement (DC-2)* in its design. To implement this approach, the SISA system is designed with two primary components:

(1) Audio Processing: Implements culturally-authentic audio segmentation and modular organization, enabling scaffolded content exploration while preserving internal essence through audio processing methodology.
(2) VR Configuration: Delivers multi-sensory stimulation via scaffolded progression mechanisms, integrating cultural authenticity and enhanced accessibility through an immersive environment.

For our initial prototype development, we selected Kunqu Opera [61] as the representative ICH performing arts genre. Following consultation with Kunqu Opera ICH practitioner CoD3, we identified a specific temporal audio segment from the performance titled "The Peach Blossom Fan, 1699"[72], specifically utilizing the sequence from 0:30 to 0:42 minutes. The Peach Blossom Fan[72] by Kong Shangren is a historical play about the doomed love between scholar Hou Fangyu and courtesan Li Xiangjun during the Ming dynasty's fall. Their romance, symbolized by a bloodstained fan, is shattered by political turmoil, reflecting the era's personal and national tragedies. Details of Prototype 1 are provided in the next section.

4.5.1 *SISA Prototype 1: Audio Processing.*

Audio Separation and Segmentation. To achieve *modular presentation of internal essence (DR-1)* while preserving *cultural authenticity (DC-1)*, our audio processing workflow begins with careful separation and segmentation of the source material. To prepare the audio data for extraction of MFCC characteristics, we performed a process that involved isolating the different components and segmenting the audio. We first used an open source online AI tool [29] specifically designed for audio separation. This tool utilizes machine learning techniques to identify and isolate vocals from audio tracks. By separating these components, we can independently analyze the vocal and instrumental content which can lead to more accurate results when applying t-SNE. With the separated tracks, we used a command-line tool, audio-splitter [67] to divide each audio file into equally-sized segments. This tool allows for precise control over the length of each segment, which maintains consistency across the dataset. By specifying the desired chunk length, audio-splitter processes the entire audio file and outputs a folder of segments of uniform duration. This segmentation is crucial to ensure that each audio segment is of manageable size for further processing and analysis. Research has shown that a 5-second window size achieves the highest accuracy for pattern recognition using MFCC [6]. With this technical consideration, we opted to proceed with this 5-second segment length configuration for our Prototype 1.

Feature Extraction. In support of *scaffolded exposure (DR-2)* and *accessibility enhancement (DC-2)*, we implemented a comprehensive feature extraction process. MFCC compresses the entire spectrum into a smaller set of coefficients that together represent the overall shape of the spectrum. It converts raw audio signals into computational data. To balance feature richness and performance, the first 13 MFCCs are typically used, as they align with human auditory sensitivity to different frequencies. These are supplemented by Delta MFCCs, which capture changes in cepstral features over time, and delta-delta MFCCs, or acceleration coefficients, which add a longer temporal context [56]. Together, these form a 39-element vector that is assumed to retain sufficient features about the audio sample.

A positive cepstral coefficient suggests that most of the spectral energy is concentrated in the low-frequency regions, while a negative coefficient indicates that the energy is primarily in the high-frequency regions. Consequently, vocalization in elevated frequency ranges characteristic of Kunqu opera, such as the melismatic singing techniques utilized in the "water sleeves" passages of dan role performances, exhibits comparable MFCC patterns that contrast markedly with those in lower frequency ranges or the accompanying ensemble of dizi, pipa, and other traditional instruments [18, 61]. Thus, MFCCs are powerful in facilitating efficacious classification predicated on these spectral attributes. A corresponding relationship between MFCC coefficients and acoustic features can be observed in Figure 3. To convert audio signals into these coefficients, we used Librosa [45], a Python package for music and audio analysis. Librosa generates a coefficient for each frame, resulting in a $(39, F)$ MFCC feature vector for each audio segment, where F equals number of frames per segment. If we split the original audio sample into N segments, we will obtain N number of MFCC feature vectors of size $(39, F)$.

Since t-SNE takes a 2D array as input, we have to reshape the feature vectors through two steps: aggregation and flattening [51]. We calculated the mean, standard deviation, minimum, and maximum of each feature to aggregate features over frames. This resulted in each feature vector having a size of $(39, 4)$. We then flattened all feature vectors into one vector with a size of $(N, 39 * 4)$, allowing t-SNE to perform dimensionality reduction on all the audio segments.

Dimensionality Reduction. To facilitate *multi-sensory engagement (DR-3)* while maintaining *cultural authenticity (DC-1)*, we employed t-SNE dimensionality reduction. This technique allows us to create meaningful spatial relationships between audio segments that reflect their inherent similarities, supporting intuitive exploration of the cultural material. t-SNE is applied to this $(N, 39 * 4)$ vector where each row corresponds to a high-dimensional data point, and each column represents a feature. This process will result in a two-dimensional coordinate map where similar-sounding segments are positioned closer together [66]. In our case, t-SNE processes the $(N, 39 * 4)$ audio feature vector by calculating pairwise similarities between audio segments in high-dimensional space, creating a similar distribution in low-dimensional space, and iteratively minimizing the Kullback-Leibler (KL) divergence between these distributions [35]. We use the scikit-learn's tool [2] to perform t-SNE. The results vary between runs, so we tested various songs and genres to explore how to optimize results. We observed that (1) a minimum similarity across segments < 0.9 and (2) KL divergence < 1, ideally < 0.5, can meet our expectation. We therefore tried with different combinations of parameters ($perplexity$ and $learning\_rate$), compared multiple runs, and visually evaluated the results to determine which one to adopt. Please refer to Figure. 3 for a detailed visual representation of the complete audio processing workflow.

To better understand the audio features in our t-SNE result and ensure *cultural authenticity (DC-1)* is preserved throughout audio processing, we implemented a density-based clustering algorithm.

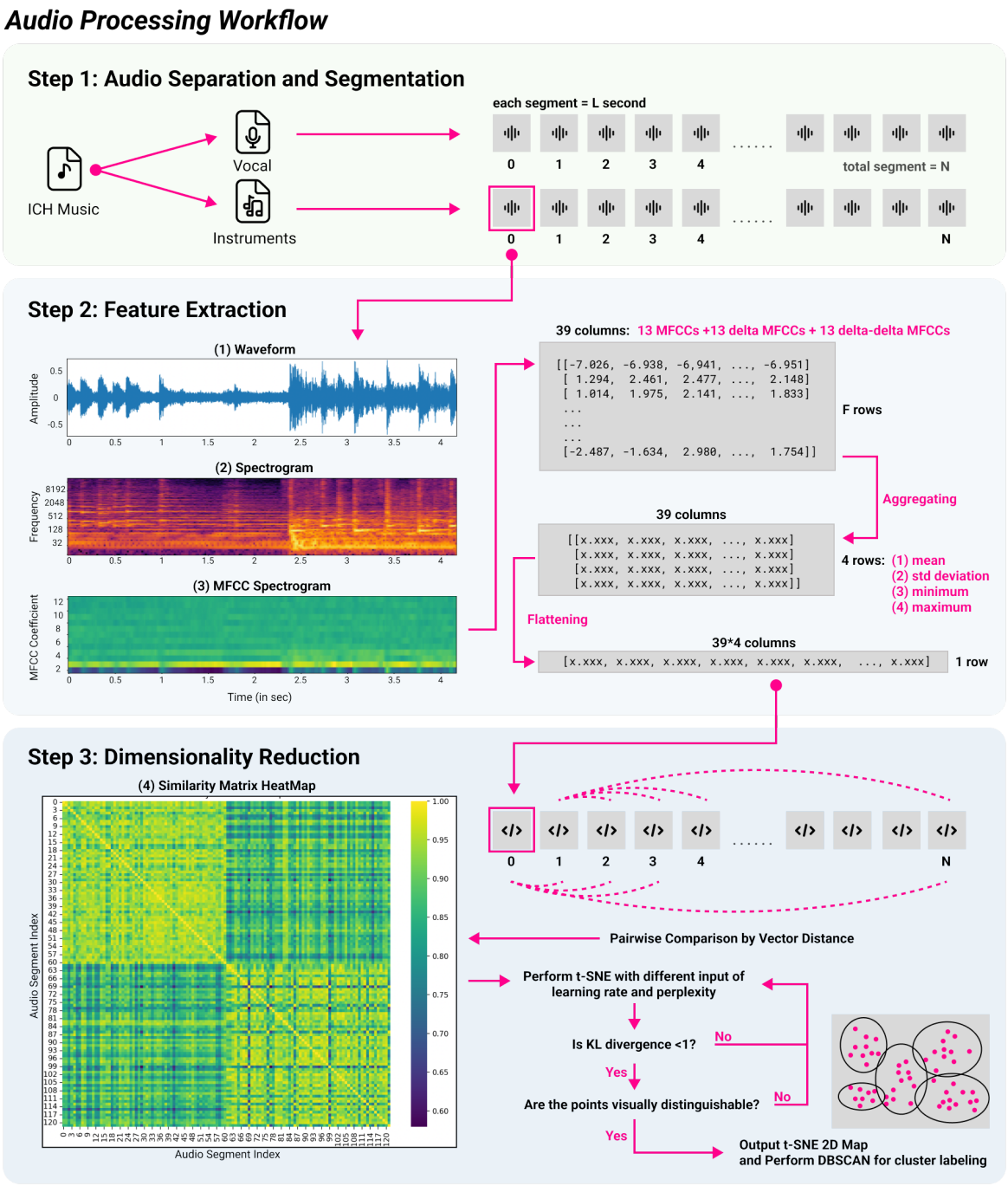

Figure 3: SISA Audio Processing Workflow consists of the following steps: *(1) Audio Separation and Segmentation, (2) Feature Extraction, (3) Dimensionality Reduction.* In this illustration, the waveform, spectrograms, and the heat map are derived from a short testing clip of Kunqu Opera.

We selected Density-Based Spatial Clustering of Applications with Noise (DBSCAN) among various clustering algorithms because it effectively identifies clusters of arbitrary shapes and it is powerful enough to handle small datasets like ours [30]. Given that DBSCAN requires the specification of key parameters—the minimum number of points per cluster (*min_samples*) and the maximum distance

between two points (*eps*). We fixed *min_samples* to the default value of 5 and ran the algorithm across a range of *eps* values to maximize the number of distinct clusters while minimizing noise points. Then, we meticulously examined the clusters generated by DBSCAN to characterize their distinct features and labeled according to UNESCO descriptions [61], with additional validation from CoD3. Our annotations revealed that the clusters naturally aligned with traditional Kunqu opera character roles: young male lead, female lead, and elderly male, which are quintessential to this art form. Notably, we also identified a cluster containing purely instrumental passages without vocals, another defining characteristic of Kunqu opera. Any noise points detected by DBSCAN were manually assigned to their most suitable clusters. Figure 4 illustrates the cluster categorization of the selected Kunqu Opera musical segments utilized in prototype 1.

This clustering approach was instrumental in preparing for our user study analysis. By mapping the spatial distribution of the audio features, we established a structured framework for analyzing user interactions with ICH content in our SISA prototype. The labeled cluster maps would later serve as a reference for overlaying user trajectories, enabling us to discern patterns in audio landscape navigation and reveal insights into exploration strategies and preferences.

4.5.2 *SISA Prototype 1: VR Configuration.*

Spatial Mapping. To implement *multi-sensory engagement (DR-3)* and support *scaffolded exposure (DR-2)*, we translated the t-SNE result into navigable 3D space. We extend the 2D t-SNE coordinates (x_{2D}, y_{2D}) into 3D space by introducing a z-coordinate. This 3D representation maintains the original t-SNE relationships. We set z_{3D} equal to y_{2D} and use a fixed radial distance r to calculate new x_{3D} and y_{3D} coordinates as follows [24].

$$\theta = 2\pi * \frac{(x_{2D} - x_{min})}{(x_{max} - x_{min})}$$
$$x_{3D} = r * cos(\theta),\ y_{3D} = r * sin(\theta),\ z_{3D} = y_{2D}$$

The spatial mapping workflow can be visualized as shown in Figure 4.

Visual Environment. To enhance *cultural authenticity (DC-1)*, and support *multi-sensory engagement (DR-3)* and visual immersion, we use a generative AI tool, Skybox AI [37], to create a culturally and contextually relevant 360-degree image for the VR environment. We used ChatGPT to extract key terms from UNESCO descriptions of the selected ICH genre sites and refine them into a concise prompt. This refined prompt was then input into Skybox AI to generate a panoramic image representing the topic associated with the selected genre. The use of panoramic visual environments aims to achieve contextual fidelity and cultural authenticity (DC-1), addressing the critical role that spatial surroundings play in framing audience interpretation of audio segments.This visual-auditory integration supports users in constructing meaningful relationships between modular audio elements (DR-1) and the holistic cultural experience, facilitating comprehension of the genre's distinctive characteristics through multi-sensory channels (DR-3). Refer to Figure.5 for the visual generation workflow.

Interaction Design. Users can interact with the VR environment with rotational movement, employing three degrees of freedom [68] (rotation, yaw, and roll). As users direct their controller toward these points, they trigger the playback of the corresponding audio segments. To enhance participants' understanding of their exploration progress, when a participant interacts with an audio segment represented by a specific shape and the color of the shape changes from white to red while the segment is playing. Once the segment finishes, the color changes to green, providing visual feedback to indicate that the point has been activated and explored.

Implementation. The technical implementation focuses on *accessibility enhancement (DC-2)* through an intuitive VR interface. A-Frame [1], a web-based VR framework, was used to create the SISA environment by placing 3D coordinates within a generated visual environment. Audio segments were assigned to spherical points and played from their positions when selected, utilizing A-Frame's built-in spatial audio functionality.

5 PHASE 2 STUDY: SUPPORT OF DESIGN RATIONALE

To validate our design approach and gather insights for future iterations, we recruited five participants to use Prototype 1 to understand user preferences on certain configuration settings of SISA system and collect feedback to support our design rationale. The study focused on investigating two primary aspects: Audio Processing Preferences and VR Configuration Preferences.

Audio Configuration Preferences. We used t-SNE to reduce the properties of audio from our selected ICH performing arts genre - Kunqu Opera- creating clusters based on these properties. To better understand the feasibility of presenting these clusters in the ICH context regarding the algorithm's specific configurations, we aimed to understand participants' system preferences regarding:

(1) *Mixed vs. Split Audio:* Whether participants preferred audio that was not separated into different tracks ("mixed") or audio that was separated ("split").
(2) *Audio Segment Duration:* Whether the current 5-second audio segment duration was sufficient for content consumption and exploration.

VR Configuration Preferences.

(1) *Visual environment:* We aimed to collect participants' preferences regarding the absence (A) or presence (B) of people in these images. We generated two 360-degree images based on our prompts using SkyboxAI. Please refer to Figure. 5 for details.
(2) *Interaction:* Whether participants preferred audio points that changed color after interaction or those that remained unchanged.

5.1 Participants and Procedure

We recruited five participants to explore our Prototype 1 featuring Kunqu Opera songs in 5-second audio segments, as suggested by Beritelli and Grasso[6]. Participants were asked to think-aloud[12] during navigation, and we conducted open-ended interviews afterwards for in-depth insights. The thematic analysis of the interview transcriptions are conducted by two researchers. The themes were

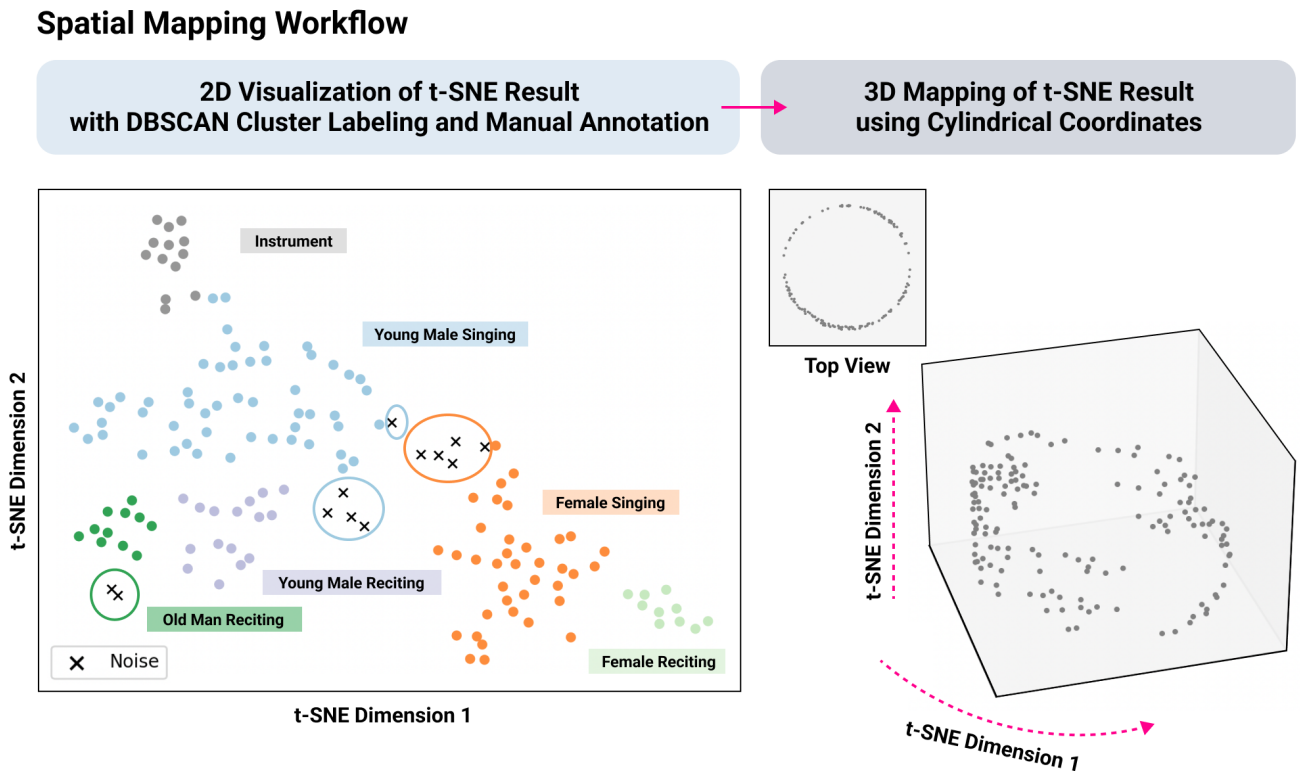

Figure 4: SISA Spatial Mapping Workflow. (Left) 2D Visualization of t-SNE Result with DBSCAN Labeling and Manual Annotation. For the selected Kunqu Opera music, t-SNE effectively grouped similar features together. Using DBSCAN, six distinct clusters were identified, differentiating between instruments and character roles, and whether the performance involved singing or reciting. (Right) 3D Mapping of t-SNE Result and its top view. Using a cylindrical coordinate system, the 2D t-SNE results were transformed into a 3D space, ensuring that when the user is positioned at the center, they are equidistant from all points.

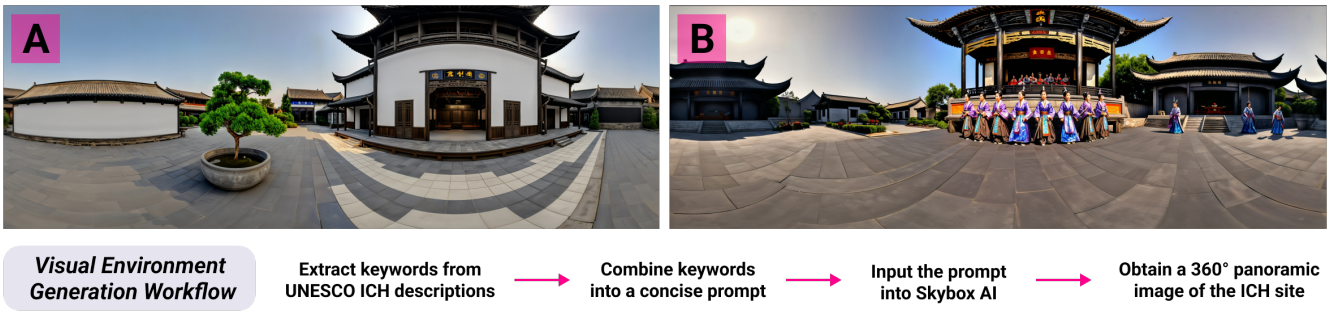

Figure 5: SISA Visual Environment Generation Workflow. We used Skybox AI to generate a visual environment of the ICH scene. The process begins with extracting keywords from UNESCO genre descriptions, followed by combining these keywords into a Skybox-specific prompt. Skybox AI then uses this prompt to create the 360° panoramic image, resulting in a detailed visual representation of the ICH site. (A) Panoramic image of Kun Qu Opera with the absence of people. (B) Panoramic image of Kun Qu Opera with the presence of people.

revised and discussed with the whole research team until final consensus was achieved[7].

5.2 Data Analysis and Results

Thematic analysis of interview transcriptions showed that participants showed a strong preference (four out of five) for the mixed audio configuration, describing it as richer and more immersive. While the distinction between close and far sound was less significant, they appreciated the ease of interaction with larger, closer audio points. Additionally, the interaction mode, where audio points changed color after interaction, was overwhelmingly favored (five out of five) as it provided clear feedback on their exploration progress. This finding strongly supports our design rationale for *multi-sensory engagement (DR-3)* through integrated auditory experiences, while the color-changing feedback mechanism validates our approach to *scaffolded exposure (DR-2)*. However, participants (three out of five) expressed that the 5-second audio segments were too short, as they found it difficult to grasp the content within such a brief duration. They suggested extending the duration of audio segments, such as to 10-second, for better content consumption and more thorough exploration. This insight particularly informs our *modular presentation approach (DR-1)*, suggesting the need for appropriate temporal granularity in content modules to ensure effective cultural transmission.

Thematic analysis of interview transcriptions showed that most participants (four out of five) preferred images that included people, as they felt this added a sense of culture and authenticity to the environment. This preference aligns with our design constraint of *cultural authenticity preservation (DC-1)*, emphasizing the importance of human elements in cultural representation. However, once they realized the images were AI-generated, concerns about authenticity emerged, with some participants expressing doubts about the accuracy of the representations. This finding highlights the tension between technological innovation and cultural authenticity, requiring careful consideration in our implementation of *accessibility enhancement (DC-2)*.

These empirical findings directly supported and refined our initial design rationales while informing specific implementation decisions and configuration preferences for subsequent prototype development. The study results led to the implementation of mixed audio as the default configuration, emphasizing immersive experience and cultural authenticity. Additionally, we integrated interactive feedback to enhance user engagement and exploration awareness. In selecting AI-generated visual elements, we took cultural authenticity into account as part of the design process.

6 PHASE 3 STUDY: USER INTERACTION PATTERNS

Building on insights from phase 1 and 2 studies, we refined our prototype design for final prototype and conducted comprehensive user testing with 16 participants to investigate spatial interaction patterns with this prototype presenting Peking Opera[62]. This phase 3 study aimed to uncover how users navigate and make sense of spatially transformed ICH content. Please refer to Figure 6 for the final prototype design workflow chart.

6.1 SISA Final Prototype

For the genre analysis, we selected Peking Opera. Reasons are presented as follows. First, this genre represents distinct linguistic traditions within China: Peking Opera is performed in Mandarin, China's official language. Second, it exhibits notably different levels of exposure among the general Chinese population. Peking Opera, while not actively sought out by younger generations, maintains broader visibility through national media platforms such as the Spring Festival Gala. Third, Peking Opera was inscribed on the Representative List of the ICH in 2010 [62]. Finally, this genre originates from China, making it particularly relevant for our user study focused on participants with Chinese cultural backgrounds.

Peking Opera represents a highly stylized form of traditional Chinese theater that combines music, singing, acrobatics, and elaborate costumes. Its distinctive performance style features unique vocal techniques and orchestral compositions, utilizing traditional instruments such as the thin high-pitched jinghu, the flute dizi, percussion instruments like the bangu and daluo. The genre is renowned for its sophisticated integration of various artistic elements and its rich historical narratives.

For the audio processing of Peking Opera, we selected the video titled "Qin Xianglian" [73] which tells the tragic story of Qin Xianglian seeking justice after being betrayed by her husband. We used SISA audio processing workflow (Figure 3) to work on the selected audio clip and split it into 10-second segments. The extension from 5-second segments in Prototype 1 to 10-second segments in the SISA Final Prototype reflects an optimization decision aimed at balancing fine-grained exploration with meaningful comprehension. This adjustment preserves the internal essence of the original performance components (DR-1 and DC-1) while supporting cognitive manageability (DC-2)—enabling users to identify character types, instrumental patterns, and emotional qualities that define Peking Opera through appropriately sized modular audio segments (DR-1 and DR-2). For instance, considering the line [73]:

> Turning my head, I will say to Xianglian: (timestamp 00:24:55-00:25:00)
> I advise you to shatter your empty hopes (timestamp 00:25:00-00:25:05)

In our previous 5-second configuration, this meaningful lyric would be fragmented, potentially confusing users encountering only partial phrases. The 10-second segment would provide more sufficient contextual integrity for proper comprehension while still maintaining manageable cognitive processing. Based on feedback from our Phase 2 Study highlighting preference with mixed audio and the need for longer segments, we refined our parameters when implementing, we modified the configuration in applying SISA Audio Processing and Spatial Mapping Workflow (Figure 3 & 4) to map the audio segments in the virtual environment. We generated visual environment using SISA Visual Environment Generation Workflow (Figure 5) , incorporating the preference of having human figures within the scene. The system overview of SISA Final Prototype is illustrated in Figure 6. The clustering results for Peking Opera in our final prototype, similar to our Prototype 1, revealed characteristic patterns consistent with UNESCO documentation [62]. Unlike Kunqu Opera, Peking Opera demonstrated distinctive acoustic distribution patterns, particularly in how instrumental sections integrated with character performances within the cluster space when using the mixed audio configuration. This comparison illuminates how different operatic traditions establish unique structural

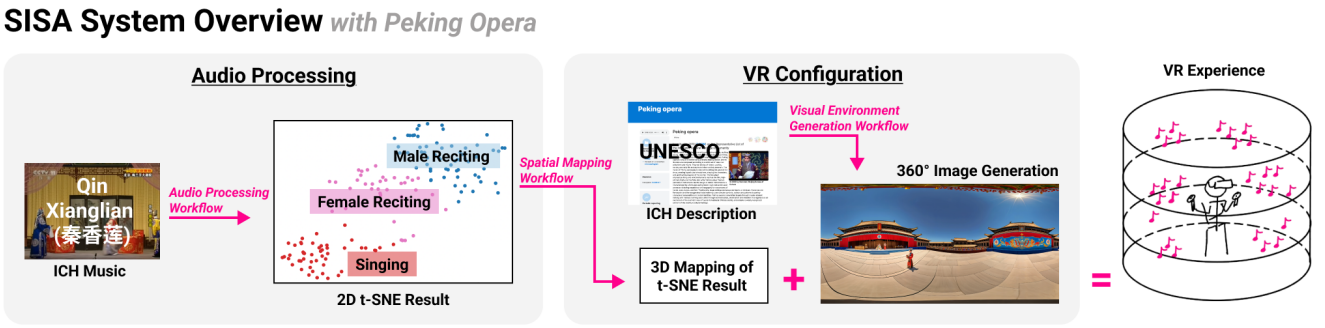

Figure 6: System Overview of the SISA illustrated with Peking Opera genre. The audio processing takes the ICH music as input and perform t-SNE algorithms after segmentation and MFCC feature extraction. The 2D result is then translated into 3D mapping and with the generated virtual environment for a culturally relevant immersive environment. Both components feed into the SISA system giving us a VR experience that enable to spatially explore the ICH performance arts.

relationships between vocal and instrumental elements, suggesting possibilities for genre-specific spatial mapping approaches in future applications, as elaborated in our discussion section.

6.2 Participants

We conducted the user study of our final prototype with 16 participants with divergent knowledge levels of Peking Opera, age groups (11 participants aged 20-27, 4 participants aged 28-35, and 1 participants aged 35+), gender (4 female and 12 male), cultural background (16 with Chinese background, which 11 with mixed cultural backgrounds, 5 with solely Chinese background), and mixed VR usage history. Table 2 provides an overview of the participants' demographics, along with their prior experience with Peking Opera. Each participant was equipped with a VR headset (Quest 2) and an iPad to complete surveys, a phone to record videos, and another phone to record think-aloud[12] and interview audio. Prior to participation, each participant provided informed consent, acknowledging that their survey responses, interview data, and videos during the workshop would be collected anonymously for analysis and that they could withdraw from the study at any time for any reason.

6.3 Procedure

With intentionality, we limit the entire experience to less than half an hour[25], with the total estimated VR usage time not exceeding 15 minutes. Moreover, we plan to allocate 10 minutes for onboarding activities, and less than 15 minutes for total duration of the temporal audio. The process began with obtaining informed consent from each participant. participants were informed that they could withdraw from the study at any time for any reason without penalty. Participants were reminded that they could pause or stop the experience at any point if they felt uncomfortable or wished to take a break. Participants were equipped with a VR headset (Quest 2) and given instructions on its proper use. They were directed to enter our prototype URL in the VR browser to access the experience. Upon opening the VR scene, participants were instructed to wait 10 seconds for the environment to load before starting their exploration. Participants were then guided to click the VR icon in the bottom right corner of the browser tab to switch from WebVR mode to full VR mode. They were taught to interact with audio points by directing the controller at them, triggering automatic playback without clicking. The color-coding system was explained. Throughout the experience, participants were encouraged to think aloud, providing real-time verbal feedback. Participants were informed they could exit a scene at any time. After completing each scene, they were instructed to press the circle button on the right-hand controller to return to the browser version and select "Download Trajectory." This process was repeated for all four scenes. The study concluded with a semi-structured interview.

6.4 Data Collection and Analysis

Our study utilized semi-structured interviews, think-aloud protocols[12], and observations as the main data sources, supplemented by system logs to provide additional supportive validation. During the VR experiences, we employed the think-aloud method[12], where participants provided real-time verbal feedback as they engaged with the four VR scenarios. This approach offered immediate insights into participants' thought processes and reactions, capturing their unfiltered responses to the virtual environments. To further enrich our qualitative data, we used a phone to record videos of the participants' physical interactions, their think-aloud protocol[12], and any spontaneous comments. These observations provided valuable visual data on how participants physically engaged with the VR technology and responded to the cultural content. The interviews gathered in-depth insights into participants' experiences, preferences, and comparisons with other audiovisual media transmission methods, forming the primary basis for our analysis. [7, 44] of interview transcripts and think-aloud[12] data was conducted to uncover insights into user experiences, exploration patterns, impact of spatial interactions on cultural understanding, and comparisons with other media transmission methods. The transcripts of the interviews and think-aloud data were coded by two researchers separately [7, 44] . The two researchers then revised the codes and discussed these themes with the entire research team. This process was repeated iteratively until final consensus was reached among all team members. The observations allowed us to directly observe

Participant	Gender	Age Group	Cultural Background	Experience with Peking Opera
1	Male	20-27	Chinese, American	Once or twice in a lifetime.
2	Female	28-35	Chinese, Canadian, American	Never heard of Peking Opera.
3	Male	28-35	Chinese, Canadian, American	Never heard of Peking Opera.
4	Female	20-27	Chinese, American	Frequently. More than 10 times in a lifetime.
5	Male	20-27	Chinese	Never heard of Peking Opera.
6	Female	20-27	Chinese, American	Several times. 3-10 times in a lifetime.
7	Male	20-27	Chinese	Several times. 3-10 times in a lifetime.
8	Male	20-27	Chinese, American	Several times. 3-10 times in a lifetime.
9	Male	20-27	Chinese, Australian	Once or twice in a lifetime.
10	Male	20-27	Chinese	Never heard of Peking Opera.
11	Male	20-27	Chinese	Never heard of Peking Opera.
12	Male	20-27	Chinese, American	Never heard of Peking Opera.
13	Female	20-27	Chinese, American	Never heard of Peking Opera.
14	Male	28-35	Chinese, American	Once or twice in a lifetime.
15	Male	28-35	Chinese, American	Once or twice in a lifetime.
16	Male	35+	Chinese, Canadian	Several times. 3-10 times in a lifetime.

Table 2: Demographic Information of Phase 3 Study Participants (N=16)

audience behaviors and interactions in real-time. During these sessions, we systematically captured field notes. All observational data were collected and organised on Miro, for visualising and analysing patterns in audience interaction patterns. Lastly, we captured trajectory data through system logs, which recorded user interactions within the VR environment, including dwell time at each audio segment and overall experience duration, and sequence of audio point IDs that participants interacted with. These information provided objective measures to support and contextualize our primary findings.

6.5 Phase 3 Study Results: User Interaction Patterns

Our analysis of 16 participants revealed rich insights into how users navigate and make sense of spatially-arranged ICH audio content in SISA.

6.5.1 Progressive Exploration Pattern. Our prototype offers a new approach to listening to auditory ICH, requiring participants spatially interacting with auditory ICH segments in the immersive environment. Participants all reported their exploration advanced from an 1) initial disorientation through 2) active cognitive engagement to develop 3) a systematic exploration to 4) pattern-seeking and awareness of spatial audio clusters, transforming traditional passive auditory ICH content consumption into a more active and engaging experience. This progression is illustrated by an example shown in Figure 7. This progression from disorientation to pattern recognition suggests that SISA facilitates an understanding of the genre's musical structure through spatial exploration. Users gradually identify distinctive genre elements—such as recitation, singing, or character-specific sonic signatures — that would typically require extensive listening to discern.

1) Initial Disorientation. Participant 2 initially experienced disorientation when encountering the SISA environment, and commented *"Em, the first time I saw everything, I wasn't sure what was going on"*. This disorientation stemmed from the unfamiliarity of the spatial representation of traditionally temporal music content and was attributed to participants' accustomed expectations of traditional temporal forms of music content consumption. Our prototype's design, which segmented temporal music into spatial audio segments, challenged the audience's typical passive content consumption style.

2) Active Cognitive Engagement. Despite initial confusion, as participants progress, this new exploration method triggers their curiosity, motivating them to actively engage and explore different points, seeking to understand the relationships between the spatially arranged audio segments. This indicates that the spatial arrangement encouraged analytical thinking and deeper engagement with the ICH material. Participant 5 remarked on the perceived intentionality of the design: *"I think these design elements may have a certain distribution. I guess this distribution might be intended to guide me to watch and listen simultaneously."* These responses indicate that the spatial-audio environment successfully engages users to move beyond passive listening to active exploration, interaction and interpretation of the auditory ICH content.

3) Systematic Spatial Exploration. To navigate the unfamiliar environment, participants developed systematic spatial exploration strategies. Participant 2 described their approach: *"It was a 360-degree picture, so I just went in a circle between different altitudes and usually checked the top as well to make sure I saw everything."* Other participants (P4,14,15) reported similar methodical approaches, such as *"First exploring the audio segments in front of me, then listening to the nearby audio segments around this segment."* Participant 14 elaborated further: *"Then I would look at the audio segments further away to see if they were very different from the closer segments."* These strategies demonstrate that participants actively tried to understand the relationships between spatially arranged audio segments while exploring the meaning of audio segments placed far apart.

4) Pattern-seeking and Awareness of Spatial Audio Clusters. As participants continued their exploration, many began to engage

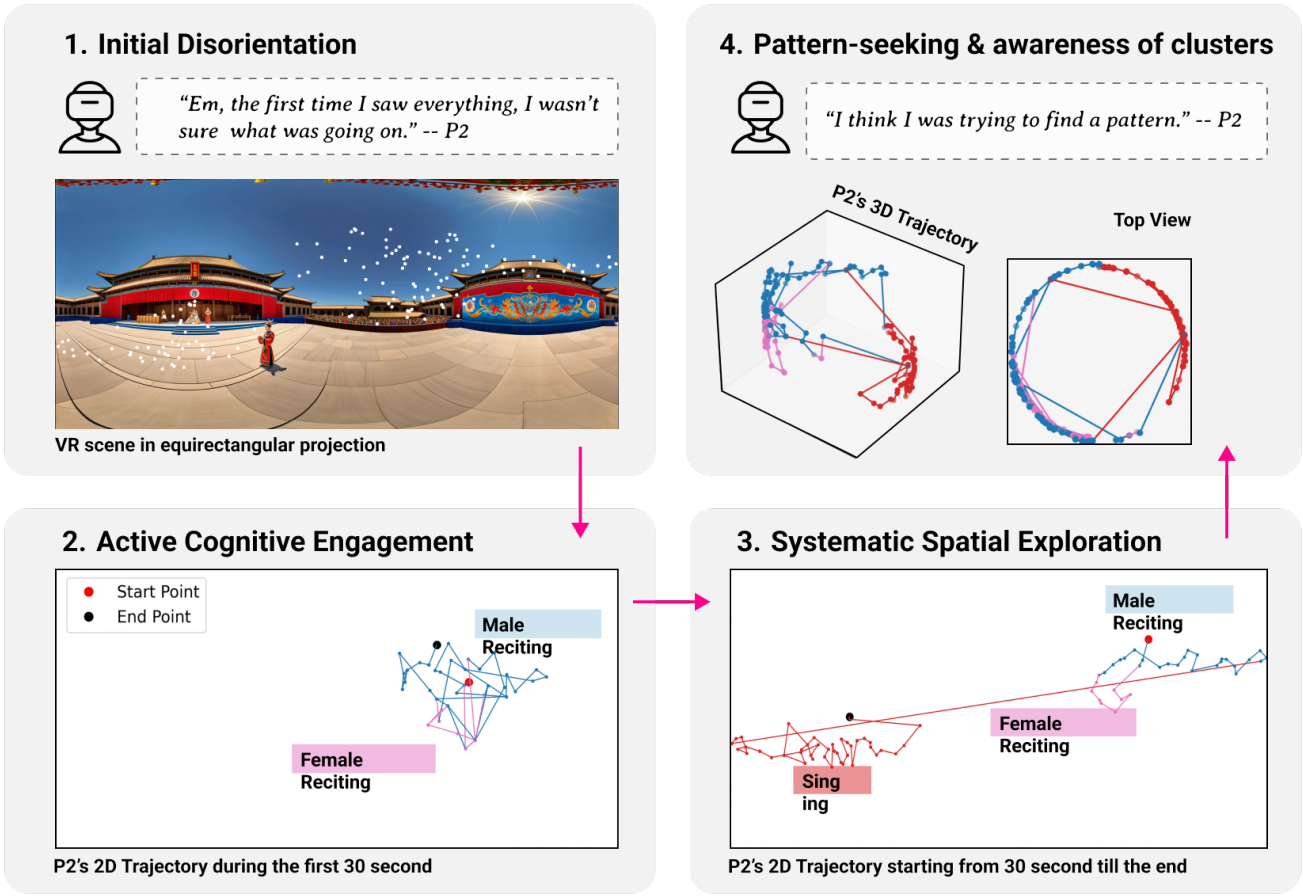

Figure 7: Progressive Exploration Pattern. The sequence illustrates Participant 2's transformation from 1) initial disorientation through 2) active cognitive engagement to develop 3) a systematic spatial exploration to 4) pattern-seeking and awareness of spatial audio clusters. (Top Left) Equirectangular Projection of the Peking Opera VR scene. (Bottom Left) The 2D trajectory of Participant 2's engagement with audio points during the first 30 seconds, showing disoriented exploration between the two reciting clusters. (Bottom Right) The trajectory of Participant 2's remaining experience after the initial 30 seconds. P2 has shown patterns to adopt a more systematic exploring between clusters.(Top Right) The 3D trajectory and top view illustrate how Participant 2 identified audio clusters, with transitions between singing and reciting clusters in the 360-degree spatial environment, indicating an emerging understanding of their distinctions.

in pattern-seeking behavior, attempting to discern intentional relationships between the spatially arranged audio segments. For instance, Participant 2 reflected on their experience: *"I think I was trying to find a pattern."* This indicates that the spatial arrangement encouraged analytical thinking and deeper engagement with the ICH material. Participant 5 remarked on the perceived intentionality of the design: *"I think these design elements may have a certain distribution. I guess this distribution might be intended to guide me to watch and listen simultaneously."* These responses indicate that the spatial-audio environment successfully engages users to move beyond passive listening to active exploration, interaction and interpretation of the auditory ICH content.

6.5.2 Adaptive Exploration Strategy: Transitioning from Visual to Audio-Driven. Participants initially relied on visual cues to navigate the spatial audio environment but gradually shifted towards using audio cues as their primary guide. This transition, illustrated in Figure 8, marked a significant change in their exploration strategy and engagement with the content. Participant 8 clearly articulated this shift: *"At first I tried to use visual cues to guide through my track, but later I found it might be more useful to use audio cues."* This transition from visual to audio-driven exploration was also evident in other participants' behaviors. For example, Participant 4's observations marked a shift towards more focused auditory recognition: *"This area is all musical bangu sounds. That area is all reciting."* This comment illustrates how participants began distinguishing between

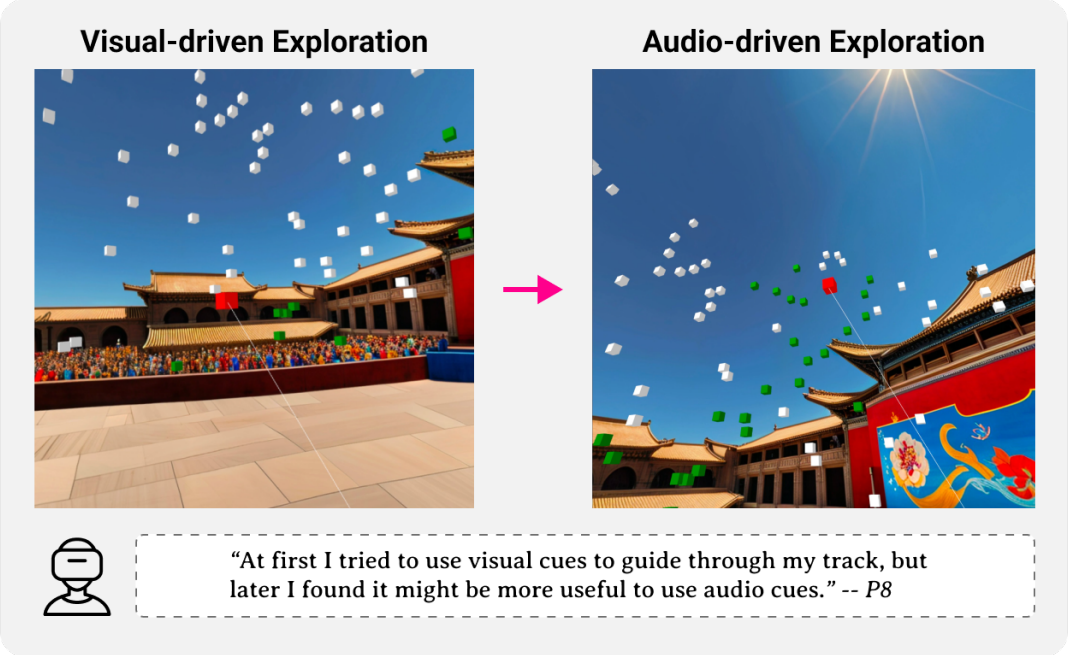

Figure 8: Adaptive exploration strategy: transition from visual-driven to audio-driven exploration in the SISA environment. (Left): Initial visual-driven interaction in the VR experience. (Right) Evolved and adapted audio-driven approach, where Participant 8 identified audio clusters independent of visual cues. This shift, exemplified by P8's comment, demonstrates the participants' adaptation to SISA environment.

different sound clusters, setting the stage for more nuanced auditory analyses. Expanding further, Participant 14's analysis reflects a sophisticated engagement with the material: *"I feel like these points might be from the same part of an opera piece. Because the sound and singing style are similar. I tried to connect them, so I listened to all the surrounding ones."* This shift from a primarily visual-oriented approach to one that is more driven by audio cues demonstrates the participants' adaptability and the design's effectiveness in emphasizing auditory over visual information, guiding their exploration and connection-building between segments. The shift from visual to audio-driven exploration also came with a learning curve, as noted by Participant 15: *"But once you figure it out, it becomes pretty intuitive after that."*

7 DISCUSSION: DESIGN INSIGHTS FOR SISA

Our findings highlighted promising opportunities where temporal-to-spatial SISA approach could enhance traditional auditory ICH engagement. Through the lens of Schaeffer's four modes of listening [55] — Écouter (listening to identify sources), Ouïr (passive perception), Entendre (selective hearing), and Comprendre (comprehending meaning) — we can interpret how the observed Progressive and Adaptive exploration patterns suggest potential pathways through increasingly sophisticated levels of auditory engagement with ICH content. In this section, we synthesize the design insights distilled from our findings. These insights reflect a combination of design considerations identified through expert evaluation, interaction patterns that emerged from our user study, and broader implications for the future of ICH preservation and engagement enhanced by machine learning algorithms and immersive technologies. They are intended to provide actionable guidance for designers and researchers working on interactive system design and development for ICH engagement and preservation, particularly in the context of auditory ICH such as performing arts.

7.1 Design Insight 1: Re-frame auditory ICH engagement, a temporal-to-spatial approach

Our research demonstrates how the SISA prototype fundamentally transforms traditional temporal-based auditory ICH experiences into spatially-oriented audio-driven interactive engagements. This transformation, anchored in our modular presentation principle (DR-1) and accessibility enhancement objectives (DC-2), also addresses critical ICH sustainability challenges by decoupling cultural engagement from location constraints. The temporal-to-spatial transformation enables accelerated comprehension of genre-specific characteristics through interactive exploration. By spatially clustering similar audio properties, SISA creates a navigable taxonomy that reveals essential patterns and variations defining Peking Opera. This approach allows even novices to distinguish between character types, vocal techniques, and emotional qualities through progressive engagement levels, while preserving cultural authenticity. Phase 3 findings identified a distinct progressive exploration pattern that aligns with Chi and Wylie's ICAP framework

[13]. Users typically advanced from initial spatial disorientation through increasingly sophisticated engagement levels: from passive observation to active exploration, constructive pattern recognition, and ultimately interactive meaning-making through audio cluster interpretation. This progression provides empirical validation of the ICAP hypothesis, which posits enhanced learning outcomes through hierarchical engagement advancement [13, 52].

7.2 Design Insight 2: Enhance engagement with auditory ICH, from visual to audio-driven

Our SISA prototype introduces an approach that transforms auditory ICH into interactive spatial experiences, advancing beyond traditional VR applications that primarily focus on visual elements [10, 75]. Our work provides designers with clear design rationale and steps for converting temporal-based auditory ICH content into engaging spatially interactive experiences.

The adaptive exploration pattern emerged from our phase 3 user study demonstrate users' natural transition from visual-driven to audio-driven exploration strategy, validating the effectiveness of our temporal-to-spatial transformation approach. This adaptivon supports our scaffolded exposure approach (DR-2) while aligning with UNESCO's emphasis [60] on active participation in cultural preservation. The framework extends the HCI toolkit for designing engaging interactive systems for ICH preservation, offering insights that can be adapted across various cultural contexts.

Both adaptive and progressive exploration patterns from our findings directly address documented limitations of passive reception in traditional auditory ICH experiences [40, 42, 50, 53, 71]. Through prioritizing auditory over visual engagement and facilitating spatial exploration, we establish a novel approach grounded in modular presentation (DR-1) that enhances accessibility (DC-2) while providing scaffolded exposure (DR-2) to auditory ICH."

7.3 Design Insight 3: Balance exploration and guidance

Our analysis demonstrates the importance of balanced navigation systems in ICH interaction design. These systems must support unstructured exploration while integrating strategic guidance mechanisms. Based on empirical observations from Phase 2 and 3 studies, we identified that effective ICH engagement requires two key elements: support for natural exploratory behaviors and guidance toward essential cultural elements, addressing our design consideration for scaffolded exposure (DC-2).

The implementation strategies should combine visual and auditory guidance, aligning with our multi-sensory design rationale (DR-3). In the SISA final prototype, visual feedback is provided through color-coded audio segments, differentiating between explored and unexplored cultural components. For auditory guidance, our SISA approach implements spatially-clustered audio segments that direct users toward significant modular elements (DR-1) of the selected ICH. This integrated multi-sensory approach (DR-3) facilitates a progressive revelation of content that corresponds to users' exploration patterns.

7.4 Design Insight 4: Balance engagement and learning

Our methodology, validated through observed progressive exploration patterns, establishes a systematic approach to embedding contextual information within immersive environments. Systems should integrate modular presentation (DR-1) of cultural knowledge, creating interconnected learning pathways that support various engagement levels while preserving cultural authenticity (DC-1). The effectiveness of this approach is evidenced by participants' successful transition from initial exploration to deeper engagement with cultural content, as observed in our Phase 3 study.

7.5 Design Insight 5: Balance technology and authenticity

Our observation reveals that the current time-based segmentation approach (5 or 10 seconds) in SISA prototypes can be sometimes disruptive to the natural flow of some performances, suggesting the necessity for more culturally informed segmentation methodologies. While time-based segmentation provides technical consistency, it occasionally fragments natural performance structures. We propose that future implementations should prioritize segmentation based on the "internal essence" (DR-1) of performances—specifically aligning with Peking Opera's inherent architecture by respecting natural boundaries defined by lyrical phrases, rhythmic patterns, and dramatic transitions. This culturally sensitive approach would preserve semantic integrity while enhancing accessibility through modular presentation. Its implementation requires extensive collaboration with ICH practitioners to identify meaningful content boundaries that maintain the traditional performance experience while minimizing disruptive interventions. The implementation of such culturally sensitive segmentation methods requires extensive collaboration with ICH practitioners to establish meaningful content boundaries that preserve traditional performance elements while minimizing disruptive interventions in the cultural experience.

7.6 Design Insight 6: Opportunities for fine-tuning t-SNE algorithms in auditory ICH preservation and engagement

Our work extends the application of t-SNE algorithms into a novel domain: the clustering of auditory ICH content to facilitate interactive spatial engagement experiences. The efficacy of our audio-clustering approach is validated through empirical observation of users successfully identifying and navigating distinct audio clusters, suggesting promising avenues for developing technologically enhanced preservation methods that maintain cultural authenticity (DC-1) while improving accessibility (DC-2). As an early attempt, we focus more on investigating how people listening to and interact with the SISA prototypes rather than developing a systematic method for fine-tuning the t-SNE algorithm specifically for ICH content. Future research should focus on developing systematic methods for optimizing clustering algorithms specifically for ICH content while maintaining the balance between technological innovation and cultural preservation.

7.7 Contributions and Implications

The above insights highlight the contributions of our work to auditory ICH interaction design. Our research advances understanding of the methodology for designing sustainable cultural heritage interactions through a structured three-phase process: participatory co-design, iterative prototyping, and comprehensive user testing.

The SISA system's development framework suggests the potential of temporal-to-spatial transformation in cultural heritage preservation, offering a potentially replicable methodology that reconceptualizes how temporal performances might be spatially preserved and experienced.

Our work contributes to the discourse on sustainable digital heritage preservation. The transformation of traditionally resource-intensive cultural experiences into digital environments provides insights into an environmentally conscious framework while seeking to ensure cultural authenticity. The methodological framework offers researchers and practitioners protocols for developing and implementing interactive systems in the ICH domain. This systematic approach establishes a foundation for methodological consistency and explores the role of interactive systems as integral components in supporting the long-term resilience of ICH preservation initiatives.

7.8 Limitations

7.8.1 Cultural genre Limitations. Our focus on Peking Opera constrains generalizability, embedding insights within Chinese cultural contexts. This specificity affects audio segmentation and clustering functionality, as different traditions employ varying rhythmic structures and tonal patterns requiring different clustering approaches. The t-SNE algorithm may perform differently when applied to other genres such as Indian classical music's tala patterns or African polyrhythmic traditions.

7.8.2 Sampling and Representation Challenges. Our participant sample (n=16), predominantly individuals with Chinese cultural backgrounds (11 mixed, 5 solely Chinese), limits the analytical scope for broader SISA framework applications. This culturally homogeneous sampling, while appropriate for our Peking Opera focus, creates a specific interpretive lens that may not account for how culturally unfamiliar audiences would navigate and make sense of spatially transformed ICH content. This limitation is particularly significant given that our design rationale explicitly included accessibility enhancement (DC-2) to address how to "help people who don't speak Chinese understand the deeper value of our art, including its cultural significance and abstract elements" (CoD3). The absence of participants without Chinese cultural knowledge means we cannot fully evaluate how effectively our approach bridges cultural divides—a critical consideration for ICH preservation and transmission in an increasingly global context.

7.8.3 Algorithmic Parameter Exploration. Our implementation of the t-SNE algorithm for audio clustering, while demonstrating a proof of concept, represents a constrained exploration of the potential parameter space. The efficacy of t-SNE is highly dependent on parameters such as perplexity and learning rate, which significantly influence the resulting spatial organization of audio segments. As noted in our methodology, we selected parameters based on observed outcomes where "minimum similarity across segments < 0.9 and KL divergence < 1, ideally < 0.5," but did not systematically investigate how different parameter configurations might affect user experiences and engagement patterns. This analytical gap limits our understanding of how algorithm optimization might enhance or detract from cultural authenticity (DC-1), particularly when considering the balance between technological innovation and preservation of a genre's "internal essence" (DR-1).

7.8.4 Visual Environment Generation limitation. While our AI generated environment creation approach aims to produce contextually relevant visual scenes, we recognize the inherent limitations of AI in representing authentic cultural spaces. The generated environments, though visually compelling, cannot fully capture the architectural nuances, historical accuracy, and atmospheric qualities of traditional performance venues. The visual fidelity of AI-generated environments and their alignment with genre characteristics were not assessed in this paper, as it falls outside the scope of our study. The AI-generated imagery raises questions about authenticity that align with our design consideration for cultural authenticity preservation (DC-1). As echoed by CoD3, traditional opera serves as "a living fossil of our culture," suggesting that visual representation requires advanced examination and preservation care.

7.8.5 Methodological Study Limitations. Our study design brings some methodological constraints affecting interpretation of SISA prototype interactions. The inconsistent exploration time among participants — resulting from our naturalistic, self-guided approach — created variability in engagement levels that complicates comparative analysis and potentially obscures patterns in spatial audio comprehension development. Our reliance on think-aloud protocols, while yielding rich qualitative data, likely altered participants' natural exploration patterns and introduced artificial reflective processes. The absence of control groups comparing traditional linear listening with our spatial approach further limits quantitative assessment of spatial transformation's impact on engagement and learning outcomes.

7.9 Implications

The SISA approach demonstrates significant potential for adaptation across various ICH dissemination contexts beyond immersive environments. The temporal-to-spatial transformation methodology established in this study offers a conceptual foundation that could enhance other interactive media platforms while addressing UNESCO's call for making ICH "truly alive" [60]

Our SISA approach could be effectively integrated into serious games, where players navigate cultural soundscapes as part of gameplay mechanics. Similar to how Ch'ng et al. demonstrated improved learning outcomes through gamified cultural heritage experiences [14], SISA design insights could transform passive musical content into interactive game elements where progression depends on recognizing and categorizing traditional performance features. For cinematic and documentary presentations, the SISA clustering methodology could inform interactive film interfaces where viewers navigate between narrative segments based on acoustic similarities,

creating personalized engagement pathways analogous to Lu et al.'s implementation of livestreaming for cultural transmission [43].

The SISA system's practical deployment would be particularly effective in museum contexts as "interactive soundscape stations." Visitors could use lightweight VR headsets at designated exhibition areas to explore traditional Peking Opera elements through spatial navigation, with each station focusing on specific performance aspects (vocal techniques, instrumental sections, character types). Similar to Tsita et al.'s VR museum implementation [59], these stations would function as self-contained exploration pods requiring minimal staff supervision while providing rich engagement opportunities. The interface would offer graduated interaction levels—from guided exploration for novices to free navigation for experienced users—addressing the accessibility enhancement considerations (DC-2) identified in our design framework. The experience resembles "tuning on the radio" but with spatial dimensions, creating an intuitive entry point for unfamiliar cultural content.

7.10 Future Work

These limitations provide valuable directions for future research, including the development of content-aware segmentation algorithms, creation of validated ICH engagement measurement tools, and conducting longitudinal studies with larger, more diverse samples. Future iterations could also incorporate eye-tracking to precisely map attention patterns during the transition from disorientation to engagement. Additionally, future studies should consider including control groups and standardizing exploration times to more robustly assess the SISA prototype's impact.

8 CONCLUSION

The SISA approach presents a novel perspective on auditory ICH engagement by transforming temporal listening into active spatial interaction within immersive VR environments. Through a structured, three-phase research process—co-designing with stakeholders, iterative prototyping, and user testing—we identified critical design elements and explored the potential of spatial audio exploration. Our findings suggest SISA's potential to foster engagement, particularly with at-risk genres like Peking Opera. By addressing specific challenges in ICH engagement and contributing to discussion around UNESCO's vision of keeping ICH vibrant, this study offers insights toward a methodological foundation for future exploration of ICH interactive systems.

REFERENCES

[1] A-Frame. 2021. A-Frame. https://github.com/aframevr/aframe

[2] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics* 8 (2014), 71792.

[3] Emilie Maria Nybo Arendttorp, Heike Winschiers-Theophilus, Kasper Rodil, Freja B. K. Johansen, Mads Rosengreen Jørgensen, Thomas K. K. Kjeldsen, and Samkao Magot. 2023. Grab It, While You Can: A VR Gesture Evaluation of a Co-Designed Traditional Narrative by Indigenous People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 308, 13 pages. https://doi.org/10.1145/3544548.3580894

[4] Svetlana Badrajan and Diana Bunea. 2023. Safeguarding and researching the intangible musical heritage in the context of contemporary digital technologies. *Valorificarea și conservarea prin digitizare a colecțiilor de muzică academică și tradițională din Republica Moldova* (2023). https://doi.org/10.55383/digimuz2023.02

[5] Wilfred C. Bain. 1969. Performing Arts: The Economic Dilemma. *Journal of Research in Music Education* 17, 1 (1969), 170–172. https://doi.org/10.2307/3344206 arXiv:https://doi.org/10.2307/3344206

[6] Francesco Beritelli and Rosario Grasso. 2008. A pattern recognition system for environmental sound classification based on MFCCs and neural networks. In *2008 2nd International Conference on Signal Processing and Communication Systems*. 1–4. https://doi.org/10.1109/ICSPCS.2008.4813723

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] Michael Brown. 2005. Heritage Trouble: Recent Work on the Protection of Intangible Cultural Property. *International Journal of Cultural Property* 12 (02 2005), 40 – 61. https://doi.org/10.1017/S0940739105050010

[9] Venkata Naresh Burla and Sudarshan Yadav. 2022. REVALUATION OF TRADITIONAL PERFORMING ARTS IN THE POST-INDEPENDENT INDIAN THEATRE. *ShodhKosh: Journal of Visual and Performing Arts* (2022). https://doi.org/10.29121/shodhkosh.v3.i2.2022.180

[10] Yunyan Cai and Cao Yang. 2023. The Application and Research of VR Animation Technology in Intangible Cultural Heritage: –Take Danzhai Miao Batik as an example. *Proceedings of the 2023 8th International Conference on Information and Education Innovations* (2023). https://doi.org/10.1145/3594441.3594467

[11] Mara Cerquetti and Concetta Ferrara. 2018. Marketing research for cultural heritage conservation and sustainability: Lessons from the field. *Sustainability* 10, 3 (2018), 774.

[12] Elizabeth Charters. 2003. The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Education Journal* 12, 2 (July 2003). https://doi.org/10.26522/brocked.v12i2.38

[13] Michelene TH Chi and Ruth Wylie. 2014. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist* 49, 4 (2014), 219–243.

[14] Eugene Ch'ng, Yue Li, Shengdan Cai, and Fui-Theng Leow. 2020. The Effects of VR Environments on the Acceptance, Experience, and Expectations of Cultural Heritage Learning. *J. Comput. Cult. Herit.* 13, 1, Article 7 (feb 2020), 21 pages. https://doi.org/10.1145/3352933

[15] Alan S Cowen, Hillary Anger Elfenbein, Petri Laukka, and Dacher Keltner. 2018. Mapping 24 Emotions Conveyed by Brief Human Vocalization. *American Psychologist* 117, 4 (2018), 1924–1934. https://doi.org/10.1037/amp0000399

[16] Alan S Cowen, Xia Fang, Disa Sauter, and Dacher Keltner. 2020. What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences* 117, 4 (2020), 1924–1934. https://doi.org/10.1073/pnas.1910704117

[17] Blanca de Miguel-Molina and Rosina Boix-Doménech. 2021. Introduction: Music, from Intangible Cultural Heritage to the Music Industry. In *Music as Intangible Cultural Heritage*, Blanca de Miguel-Molina, Victoria Santamarina-Campos, Marina de Miguel-Molina, and Rosina Boix-Doménech (Eds.). Springer International Publishing, 3–8. https://doi.org/10.1007/978-3-030-76882-9_1

[18] Li Dong, Jiangping Kong, and Johan Sundberg. 2014. Long-term-average spectrum characteristics of Kunqu Opera singers' speaking, singing and stage speech. *Logopedics Phoniatrics Vocology* 39, 2 (2014), 72–80.

[19] Stéphane Dupont, Thierry Ravet, Cécile Picard-Limpens, and Christian Frisson. 2013. Nonlinear dimensionality reduction approaches applied to music and textural sounds. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[20] Zeynep Erol, Zhiyuan Zhang, Eray Özgünay, and Ray LC. 2022. SOUND OF(F): Contextual Storytelling Using Machine Learning Representations of Sound and Music. In *ArtsIT, Interactivity and Game Creation*. Springer, 332–345.

[21] Leon Fedden. 2017. Comparative Audio Analysis With Wavenet, MFCCs, UMAP, t-SNE and PCA. https://medium.com/@LeonFedden/comparative-audio-analysis-with-wavenet-mfccs-umap-t-sne-and-pca-cb8237bfce2f

[22] Kexue Fu, Yixin Chen, Jiaxun Cao, Xin Tong, and RAY LC. 2023. "I Am a Mirror Dweller": Probing the Unique Strategies Users Take to Communicate in the Context of Mirrors in Social Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3544548.3581464

[23] Kexue Fu, Ruishan Wu, Yuying Tang, Yixin Chen, Bowen Liu, and RAY LC. 2024. "Being Eroded, Piece by Piece": Enhancing Engagement and Storytelling in Cultural Heritage Dissemination by Exhibiting GenAI Co-Creation Artifacts. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2833–2850. https://doi.org/10.1145/3643834.3660711

[24] Ramy Georgious Zaher Georgious, K Elfeky, RR Elrazky, et al. 2022. Scalar, phasors and vector magnitudes for electric and electronic engineering. *Encyclopedia of Electrical and Electronic Power Engineering: Volumes 1-3* (2022).

[25] Eugy Han, Kristine L. Nowak, and Jeremy N. Bailenson. 2022. Prerequisites for Learning in Networked Immersive Virtual Reality. *Technology, Mind, and Behavior* 3, 4 (2022). https://doi.org/10.1037/tmb0000094

[26] Zhiting He, Jiayi Su, Li Chen, Tianqi Wang, and RAY LC. 2025. "I Recall the Past": Exploring How People Collaborate with Generative AI to Create Cultural

Heritage Narratives. *Proceedings of the ACM on Human-Computer Interaction* 9, CSCW 108 (April 2025), 30. https://doi.org/10.1145/3711006

[27] Hong Kong SAR Government. 2022. Government announces appointment of Postmaster General. Press Release. https://www.info.gov.hk/gia/general/202211/09/P2022110900138.htm

[28] Hilary Hutchinson, Wendy Mackay, Bosse Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, Björn Eiderbäck, Sinna Lindquist, and Yngve Sundblad. [n. d.]. Technology Probes: Inspiring Design for and with Families. ([n. d.]).

[29] Ben Jameson. 2021. SongDonkey.AI. https://songdonkey.ai/

[30] Hari Krishna Kanagala and V.V. Jaya Rama Krishnaiah. 2016. A Comparative Study of K-Means, DBSCAN and OPTICS. , 6 pages. https://doi.org/10.1109/ICCCI.2016.7479923

[31] Holtzblatt Karen and Jones Sandra. 2017. Contextual inquiry: A participatory technique for system design. In *Participatory design*. CRC Press, 177–210.

[32] Sungyoung Kim, Doyuen Ko, Miriam A. Kolar, and Xuan Lu. 2022. Aural heritage preservation and access: Methodological explorations from data collection to immersive multimodal virtual reality. *The Journal of the Acoustical Society of America* (2022). https://doi.org/10.1121/10.0016251

[33] Marina Kliuchko, Elvira Brattico, Benjamin Gold, Mari Tervaniemi, Brigitte Bogert, Petri Toiviainen, and Peter Vuust. 2019. Fractionating auditory priors: A neural dissociation between active and passive experience of musical sounds. *PLOS ONE* 14 (05 2019), e0216499. https://doi.org/10.1371/journal.pone.0216499

[34] Cuiting Kong. 2024. Digital Diabolo: A Virtual Reality Game for the Presentation of Intangible Cultural Heritage Through Participatory Design. In *Proceedings of the Participatory Design Conference 2024: Situated Actions, Doctoral Colloquium, PDC Places, Communities - Volume 3* (Sibu, Malaysia) *(PDC '24)*. Association for Computing Machinery, New York, NY, USA, 19–23. https://doi.org/10.1145/3661456.3666052

[35] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.

[36] AMT Lab. 2022. The challenge to keep Gen Z interested in long-form, high-quality content. Arts Management and Technology Laboratory (AMT Lab), Carnegie Mellon University. https://amt-lab.org/blog/2022/10/the-challenge-to-keep-gen-z-interested-in-long-form-high-quality-content

[37] Blockade Labs. 2024. Skybox AI. Blockade Labs, generating 360°panoramic images in glorious 8K resolution. https://skybox.blockadelabs.com/

[38] RAY LC. 2023. TOGETHER ENOUGH: Collaborative Constructions of Adaptations to Climate Futures. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference (DIS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 55–59. https://doi.org/10.1145/3563703.3596805

[39] Guanhong Liu, Xianghua Ding, Jinghe Cai, Weiyun Wang, Xinyue Wang, Yuting Diao, Jin Chen, Tianyu Yu, Haiqing Xu, and Haipeng Mi. 2023. Digital making for inheritance and enlivening intangible cultural heritage: A case of hairy monkey handicrafts. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–17.

[40] Zixiao Liu, Shuo Yan, Yu Lu, and Yuetong Zhao. 2022. Generating Embodied Storytelling and Interactive Experience of China Intangible Cultural Heritage "Hua'er" in Virtual Reality. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 439, 7 pages. https://doi.org/10.1145/3491101.3519761

[41] Charles Lo. 2012. Nonlinear dimensionality reduction for music feature extraction. *Tech. Rep. CSC2515* (2012).

[42] Fei Lu, Feng Tian, Yingying Jiang, Xiang Cao, Wencan Luo, Guang Li, Xiaolong Zhang, Guozhong Dai, and Hongan Wang. 2011. ShadowStory: creative and collaborative digital storytelling inspired by cultural heritage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1919–1928.

[43] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. 2019. " I feel it is my responsibility to stream" Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[44] Maria Lungu. 2022. The Coding Manual for Qualitative Researchers. *American Journal of Qualitative Research* 6, 1 (May 2022), 232–237. https://doi.org/10.29333/ajqr/12085

[45] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in python.. In *SciPy*. 18–24.

[46] Preben Mogensen. [n. d.]. TOWARDS A PROVOTYPING APPROACH IN SYSTEMS DEVELOPMENT. 4, 1 ([n. d.]).

[47] Felesia Mulauzi, Phiri Bwalya, Chishimba Soko, Vincent Njobvu, Jane Katema, and Felix Silungwe. 2021. Preservation of audio-visual archives in Zambia. *ESARBICA Journal: Journal of the Eastern and Southern Africa Regional Branch of the International Council on Archives* (2021). https://doi.org/10.4314/esarjo.v40i.4

[48] Reese Muntean, Alissa N Antle, Brendan Matkin, Kate Hennessy, Susan Rowley, and Jordan Wilson. 2017. Designing cultural values into interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6062–6074.

[49] Caroline Joelle Nwabueze. 2013. The role of intellectual property in safeguarding intangible cultural heritage in museums. *International journal of intangible heritage* 8 (2013), 181–190.

[50] Razifah Othman, Masitah Ahmad, Othman Ibrahim, Haziah Sa'ari, Siti Nuur–Ila Mat Kamal, and Aflah Isa Darami. 2021. Overview of UX-UI Via Virtual Reality Project in Preserving the Intangible Cultural Heritage of Negeri Sembilan, Malaysia. (2021), 180–185. https://doi.org/10.1109/ICSPC53359.2021.9689107

[51] Tamás Pál and Dániel T Várkonyi. 2020. Comparison of Dimensionality Reduction Techniques on Audio Signals.. In *ITAT*. 161–168.

[52] Yue Qin and Hassan A Karimi. 2024. Active and passive exploration for spatial knowledge acquisition: A meta-analysis. *Quarterly Journal of Experimental Psychology* 77, 5 (2024), 964–982.

[53] Junkai Rao, Feng Zhou, Ju Dai, Chi Li, and Yong Hu. 2024. FormationCreator: Designing A VR Dance Formation System for Intangible Cultural Heritage Dance. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[54] Gerard B Remijn and Haruyuki Kojima. 2010. Active versus passive listening to auditory streaming stimuli: a near-infrared spectroscopy study. *Journal of biomedical optics* 15, 3 (2010), 037006–037006.

[55] Pierre Schaeffer. 2017. *6. The Four Listening Modes*. University of California Press, Berkeley, 80–93. https://doi.org/doi:10.1525/9780520967465-012

[56] Alka Singh and Sureka Ghanges. 2016. SPEAKER RECOGNITION USING MFCC AND DELTADELTA MFCC AND CLASSIFICATION USING ARTIFICIAL NEURAL NETWORK. *2016 8th International Journal of Advance Research in Science and Engineering* (2016), 8354.

[57] Dongyan Sun and Chengping Wang. 2024. Application of AR Technology in Intangible Cultural Heritage and Cultural Tourism. In *Proceedings of the 3rd International Conference on Electronic Information Technology and Smart Agriculture* (Sanya, China) *(ICEITSA '23)*. Association for Computing Machinery, New York, NY, USA, 247–252. https://doi.org/10.1145/3641343.3641387

[58] Peng Tan, Damian Hills, Yi Ji, and Kaiping Feng. 2020. Case Study: Creating Embodied Interaction with Learning Intangible Cultural Heritage through WebAR. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3375199

[59] Christina Tsita, Maya Satratzemi, Alexandros Pedefoudas, Charalabos Georgiadis, Maria Zampeti, Evi Papavergou, Syrago Tsiara, Eleni Sismanidou, Petros Kyriakidis, Dionysios Kehagias, and Dimitrios Tzovaras. 2023. A Virtual Reality Museum to Reinforce the Interpretation of Contemporary Art and Increase the Educational Value of User Experience. *Heritage* 6 (05 2023), 4134–4172. https://doi.org/10.3390/heritage6050218

[60] UNESCO. 2003. *Basic Texts of the 2003 Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO, Paris. https://ich.unesco.org/en/convention

[61] UNESCO. 2008. Kun Qu opera - Intangible Cultural Heritage of Humanity. https://ich.unesco.org/en/RL/kun-qu-opera-00004.

[62] UNESCO. 2010. Peking Opera - Intangible Cultural Heritage of Humanity. https://ich.unesco.org/en/RL/peking-opera-00418.

[63] UNESCO. 2023. Meshrep. https://ich.unesco.org/en/USL/meshrep-00304. Accessed: 2024-08-29.

[64] UNESCO. n.d.. Dive into Intangible Cultural Heritage. https://ich.unesco.org/en/dive. Accessed: 2024-08-29.

[65] UNESCO Intangible Cultural Heritage. 2023. We Are Living Heritage: Photo Exhibition 2023. Online Exhibition. https://ich.unesco.org/en/we-are-living-heritage-photo-exhibition-2023-01331

[66] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[67] Artem Vasilev. 2023. audio-splitter 0.1.0. https://github.com/temavasilev/audio-splitter

[68] Google VR. 2019. Degrees of Freedom. https://developers.google.com/vr/discover/degrees-of-freedom

[69] Stephan Wensveen and Ben Matthews. [n. d.]. Prototypes and Prototyping in Design Research. In *The Routledge Companion to Design Research* (1 ed.), Paul A. Rodgers and Joyce Yee (Eds.). Routledge, 262–276. https://doi.org/10.4324/9781315758466-25

[70] Glenn A Withers. 1980. Unbalanced growth and the demand for performing arts: An econometric analysis. *Southern Economic Journal* (1980), 735–742.

[71] Zhihao Yao, Shiqing Lyu, Yao Lu, Qirui Sun, Hanxuan Li, Xuezhu Wang, Guanhong Liu, and Haipeng Mi. 2024. ShadowMaker: Sketch-Based Creation Tool for Digital Shadow Puppetry. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[72] YouTube. 2012. "1699 Peach Blossom". https://www.youtube.com/watch?v=CUPkUkLT69Y&t=2087s. Uploaded on January 15, 2012, with 115,724 views as of Jan 16, 2025.

[73] YouTube. 2014. "Qin Xianglian" [English Subtitles]. https://www.youtube.com/watch?v=5NzBu5v-ISE. Uploaded on July 18, 2014, with 454,924 views as of August 29, 2024.

[74] Minjing Yu, Meng Zhang, Chun Yu, Xiaoguang Ma, Xing-Dong Yang, and Jiawan Zhang. 2021. We Can Do More to Save Guqin: Design and Evaluate Interactive

Systems to Make Guqin More Accessible to the General Public. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 294, 12 pages. https://doi.org/10.1145/3411764.3445175

[75] Lufang Zhang, Yue Wang, Zhichuan Tang, Xia Liu, and Moran Zhang. 2023. A Virtual Experience System of Bamboo Weaving for Sustainable Research on Intangible Cultural Heritage Based on VR Technology. *Sustainability* (2023). https://doi.org/10.3390/su15043134

[76] Xuanmiao Zhang, Linqi Sun, and Shuo Yan. 2023. NVSHU: Virtual Reality Design and Narrative Popularization for Intangible Cultural Heritage Characters. In *SIGGRAPH Asia 2023 XR* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 2 pages. https://doi.org/10.1145/3610549.3614615