What Does It Sound Like? Reconstructing and Reimagining Cultural Soundscapes

Bingqing Chen[a], Xuehua Fu[b], Yue Li[a], and Marcel Zaes Sagesser[c*]

[a]School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China
[b]ENES Bioacoustics Research Laboratory, University of Saint-Étienne, Saint-Étienne, France
[c]University of Applied Sciences St. Pölten, St. Pölten, Austria

ABSTRACT

This study explores how generative artificial intelligence (GenAI) can support human co-creation in reconstructing and reimagining cultural soundscapes. Participants were invited to a local heritage site, *Xiyuan Temple* (Suzhou), where they collected field recordings and photographs, then collaboratively generated new ambient sounds through AI using text-based prompts. The process revealed that participants did not merely reproduce the existing auditory environment but actively idealized it - introducing imagined elements such as temple bells or animal sounds to express a personal sense of place. This phenomenon highlights the interplay between reconstruction and idealization in human-AI collaboration and reveals how generative AI can serve as a tool for cultural reinterpretation. Drawing on soundscape theory and human-AI co-creation, this work contributes to the field of AI-mediated sonic co-creation for cultural contexts, where human perception, imagination, and algorithmic models interact to reshape cultural and sensory heritage. The findings highlight the potential of GenAI as a tool for creative cultural engagement, expanding the AI4Culture discourse through embodied listening and participatory imagination.

Keywords: Large Language Model, Cultural Heritage, Soundscape, Human-AI Co-creation

1. INTRODUCTION

Generative artificial intelligence (GenAI) is increasingly used for cultural applications, from image generation to interactive storytelling. However, most works have focused on visual media, while the role of generative models in sonic and embodied cultural experiences is still underexplored. In particular, we know little about how people might use Artificial Intelligence Generated Content (AIGC) to reconstruct and reimagine the soundscapes of cultural heritage sites. In soundscape research, environmental sound is treated not merely as "noise", but as a resource that shapes people's perception of place and quality of life.[1,2] Recent human-centered projects have combined soundscapes theory with participatory and citizen science approaches, enabling lay participants to record, annotate, and creatively rearrange urban soundscapes using mobile phones and interactive archives.[3,4] These systems foreground subjective listening and qualitative interpretation, rather than relying solely on purely quantitative noise metrics. Recent work has developed mobile recording tools and web-based sound archives for participatory listening in high-tech urban environments such as Shenzhen,[5] showing how citizens can document and reflect on complex urban sonic identities through distributed systems. In this paper, we extend this line of research from urban to heritage soundscapes, and from documentation to co-creation with GenAI, focusing on how visitors collaborate with generative models to reinterpret a site's acoustic environment. By situating our work in a local heritage site in Suzhou, China, the *Xiyuan Temple*, our guiding research questions are,

RQ1 How do users explore and perceive *Xiyuan Temple*'s soundscape through our system?

RQ2 How do users reflect on and reimagine *Xiyuan Temple*'s soundscape through our system?

RQ3 How do users use the GenAI module to compose their imaginary soundscape?

*Corresponding author: Marcel.Sagesser@ustp.at

International Conference on AI-Generated Content (AIGC 2025), edited by Feng Zhao, Duoqian Miao, Proc. of SPIE Vol. 14183, 141830L
 · 0277-786X · doi: 10.1117/12.3108901

We conducted a study with four participants in which they visited the *Xiyuan Temple* in Suzhou. Results revealed that participants selectively attended to temple-specific and natural sounds as key markers of place, and used the interface not only to reconstruct their walks but also to idealize the temple soundscape by removing unwanted noise and introducing imagined elements using GenAI. GenAI was thus positioned as a supportive co-creator for filling gaps and materializing personal imaginaries — while real recordings remained the primary carriers of presence, warmth, and authenticity. At the same time, participants raised concerns about the relative flatness and mechanical quality of the generated audio. These findings both demonstrate the potential of AIGC for creative cultural engagement and foreground design challenges around balancing reconstruction and idealization, preserving a sense of authenticity, and supporting embodied listening in AI-mediated soundscape systems.

2. RELATED WORK

2.1 Soundscapes and Cultural Heritage

The concept of the soundscape describes the acoustic environment as it is perceived by listeners, rather than as a purely physical signal. Schafer's foundational work framed environmental sound as "our sonic environment" and advocated soundwalk as a way to cultivate critical listening to everyday places.[1] Subsequent soundscape research has shifted from a narrow concern with noise reduction toward human-centred notions of acoustic quality, restoration, and place identity and has begun to recognise environmental sounds as part of intangible cultural heritage, especially in sites such as historic districts, markets, and religious spaces.[6] Building on this perspective, participatory and mobile soundscapes tools have invited citizens to document and assess their own acoustic environments. For example, the Hush City app enables users to crowdsource, map, and evaluate quiet areas in cities through a citizen-science approach, producing open-access sound maps that can inform planning.[7] UrbanRemix combines location-based mobile recording with web-based remix tools, allowing participants to explore and recompose the "acoustic identity" of their communities through workshops and public art events.[8] While these systems demonstrate how participatory listening and basic remixing can deepen engagement with everyday soundscapes, they have rarely been applied to heritage settings and generally focus on documentation, mapping, or public art rather than on visitors' fine-grained processes of reconstructing and imaginatively transforming a specific cultural soundscape.

2.2 GenAI for Cultural Interpretation

Recent advances in AIGC have opened up new possibilities for cultural interpretation, especially in the visual domain. Zhang et al.[9] propose a pipeline that couples large language models with diffusion models and LoRA fine-tuning to generate images of Chinese traditional culture from Chinese prompts, arguing that such systems can support the "inheritance and revitalization" of cultural motifs. Other AI4Culture work explores generative tools for rural cultural creativity, intangible cultural heritage dance, and digital exhibitions, positioning AI as a collaborator that helps non-experts experiment with traditional aesthetics and narratives.[10] These studies highlight how AIGC can mediate between complex cultural knowledge and accessible creative practice. However, most existing AI4Culture projects focus on visual media, and pay limited attention to environmental soundscapes as a cultural resource. Very few studies examine how visitors might use text-to-audio models to co-create site-specific ambience or how generative audio reshapes their perception of authenticity, place, and embodiment. The tension between reconstructing what was heard and idealizing what a cultural site should sound like remains largely unaddressed in current AIGC and cultural heritage research.

3. METHOD

3.1 System Design & Implementation

Our prototype consists of a web-based mobile recorder for in-situ documentation and a desktop human-AI co-creation interface that supports multi-track mixing and sound generation (as shown in Figure 1).

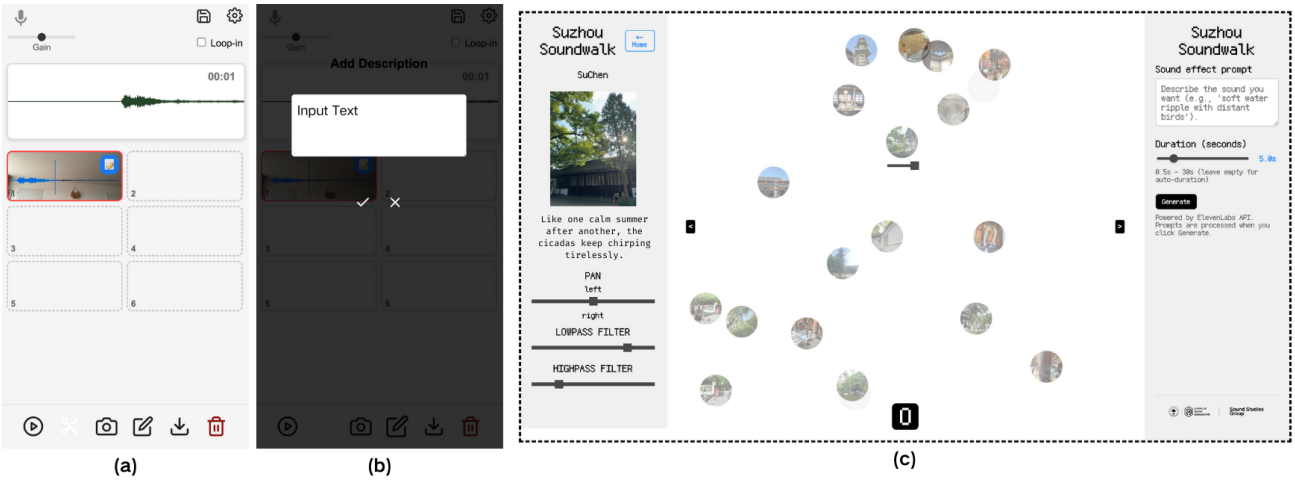

Figure 1. System Interfaces. (a) Web-based mobile recorder used in the field to capture short audio clips and associated photos. (b) Text input panel for adding descriptions to each recording. (c) Desktop co-creation interface showing recorded sound objects on a canvas (center) with mixing controls and a text-to-audio prompt panel (right) for generating sounds.

3.1.1 Mobile Recorder

The mobile recorder is a lightweight web application that runs in the smartphone browser and allows participants to capture short audio clips, take photographs, and add free-text notes for each recording. Technically, the recorder is implemented using HTML5 and JavaScript. Audio capture is handled via the Web Audio and MediaStream APIs, with each recording limited to approximately 30 seconds to keep file sizes manageable and to encourage focused selection. Once a clip is recorded, the interface presents basic controls to play it back and either keep or discard it. The browser's camera API is then used to attach a contextual photograph, and a text input field allows the participant to write a short note about why they chose this sound or how it made them feel. Each element, including audio file, image, and text, is stored locally as a JSON structure containing metadata. At the end of the field session, all JSON files are exported as a compressed archive and transferred to the lab computer, where they form the basis of the subsequent co-creation stage (see Figure 2).

3.1.2 Co-creation Interface with GenAI

The interface is implemented as a browser-based application using JavaScript and the Web Audio API. Each imported sound object appears as a circular thumbnail on a 2D canvas, with the on-site photograph shown as its icon. Clicking a circle plays or pauses the corresponding clip. Additionally, multiple circles can be active simultaneously, allowing for layered sound mixtures. For each active clip, a small control panel provides standard parameters such as playback, per-track volume, stereo panning, and simple biquad filters (low- and high-pass), all realized within a single Web Audio context. To integrate generative AI, we added a dedicated text-to-audio module. A text box in the interface allows participants to type prompts in English that describe the sounds they wish to add (e.g., "distant temple bell"). When they were not comfortable writing in English, they wrote in Chinese and used Google Translate to obtain the English prompt. When a prompt is submitted, the client sends an HTTPS request to the ElevenLabs API*, which we use as an off-the-shelf environmental sound generator. The API returns a short audio file, which is decoded into an AudioBuffer and added to the same Web Audio context as the field recordings (see Figure 2).

3.2 Site, Participants and Procedure

Xiyuan Temple is a historic Buddhist temple in Suzhou, known for its large garden, koi ponds, free-roaming cats and Buddha statues (see Figure 3). The site combines religious activities with everyday tourism, making it a rich sonic environment. We recruited a total of four participants aged from 21 to 31 ($M = 25.250, SD = 4.646$).

*https://elevenlabs.io/

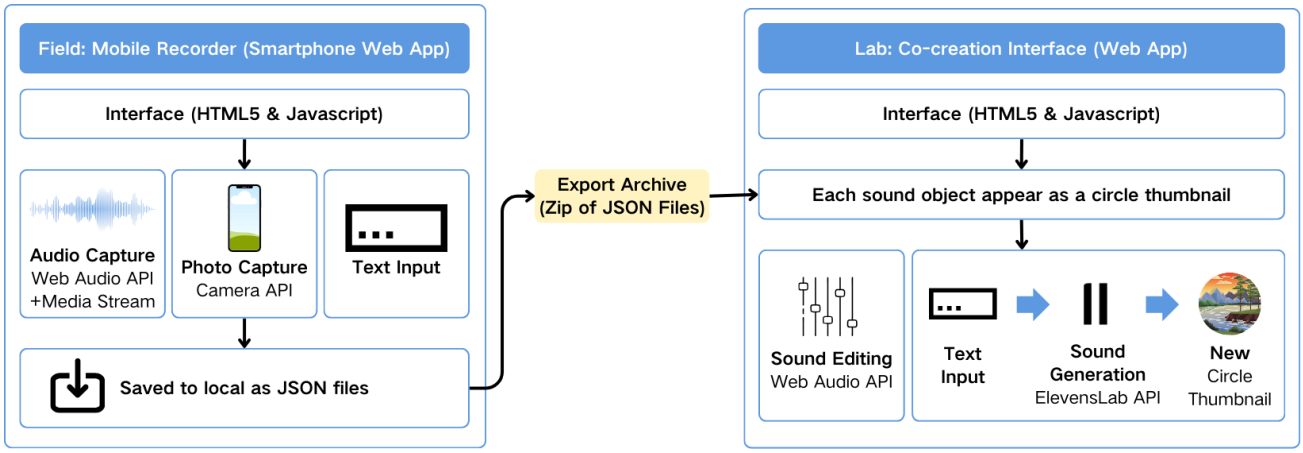

Figure 2. System design and implementation.

Figure 3. *Xiyuan Temple*. (a) Main courtyard and entrance. (b) Central pavilion and pond in the garden. (c) Buddha statues in the main hall.

All had lived in Suzhou for at least five years, and three identified as local residents. They reported visiting the temple approximately one to two times per year. The study followed a two-stage procedure conducted on the same day. First, in the field session at *Xiyuan Temple*, a moderator briefly introduced the study purpose, system, and tasks. Participants were then asked to walk freely through the temple grounds for about an hour, using the mobile recorder to capture up to six short audio clips that they felt best represented the site. For each clip, they also took a photo and wrote a short note, and at the end of the walk, they completed a brief soundscape questionnaire about their overall experience. In the lab-based session, participants were asked to use the desktop co-creation interface to listen back to their recordings, adjust basic parameters, classify the audio clips, and optionally generate additional sounds via the text-to-audio module. After completing their composition, each participant participated in a semi-structured interview. The procedure is illustrated in Figure 4.

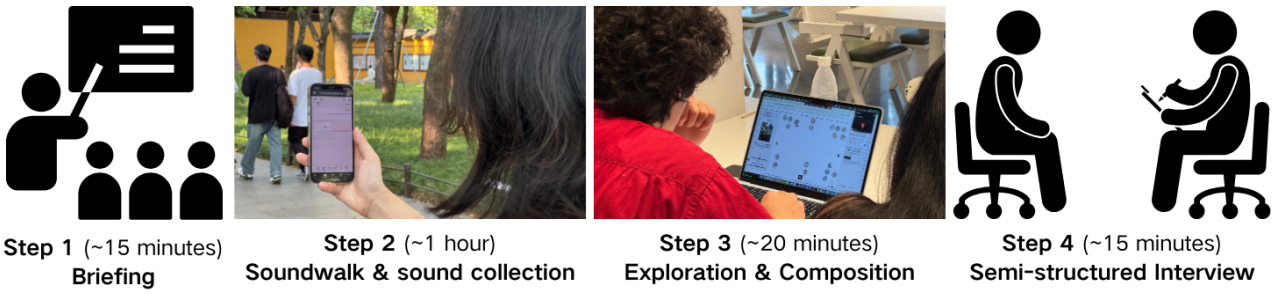

Figure 4. The experimental procedure.

3.3 Measures

We adapted a brief soundscape questionnaire from the ISO soundscape standard,[11] asking participants to rate the temple environment on core perceptual dimensions (e.g., pleasant–annoying, calm–chaotic, eventful–uneventful, vibrant–monotonous) using 5-point Likert scales. We also collected system logs that recorded the full co-creation process, including when participants clustered, arranged, and generated sounds. In addition, we conducted semi-structured interviews after the co-creation session to probe how participants selected and combined sounds, how and why they used the AIGC module, and how they evaluated the resulting soundscapes.

4. RESULTS

4.1 Soundscape Exploration (RQ1)

Questionnaire ratings indicated that the soundscape of *Xiyuan Temple* was experienced as positive: participants tended to rate it toward the pleasant ($M = 4.250$, $SD = 0.957$), calm ($M = 4.333$, $SD = 1.258$), vibrant ($M = 4.333$, $SD = 0.957$) and eventful ($M = 4.250$, $SD = 0.500$), and away from annoying ($M = 1.667$, $SD = 1.500$), chaotic ($M = 1.750$, $SD = 1.500$), and monotonous ($M = 1.250$, $SD = 0.500$). In interviews, they consistently described cicadas as the first and most salient element with "loud", "constant", and "accompanying the whole journey". Other natural sounds (birds, pond water, cats) and temple-specific sounds (coin tossing, chanting, group prayer) were highlighted as key markers that made the site feel distinct from a generic park or garden. Participants reported no fixed plan before the soundwalk. Instead, they recorded opportunistically whenever a moment caught their eye or felt emotionally meaningful. When asked which recording best represented *Xiyuan Temple*, they pointed to scenes such as people playing with cats, the quietest prayer hall, where only a fan was audible, collective chanting near the incense burner, or the sound of coins being thrown. Across cases, they framed sound not as background noise but as a trace of place that, together with visual impressions, defined their personal sense of the temple. These temple-specific and natural sounds were treated as markers of place.

4.2 Soundscape Reconstruction and Idealization (RQ2)

In the lab-based co-creation session, participants used the desktop interface to cluster their recordings into intuitive groups and to combine and arrange these sounds into a short soundscape. P3 reflected on them in terms of different emotions (see Figure 5a). P2 clustered recordings into meaningful groups—for example, "temple sounds" versus "people and animals connecting with the temple" (see Figure 5b). The interview suggested that participants found the interface easy to use and felt they had reasonable control over the resulting arrangement. Qualitative analysis revealed two main strategies. A reconstructive strategy aimed to approximate the remembered soundwalk, using basic controls mainly to clarify or balance existing recordings. An idealizing strategy used the interface to "clean up" the soundscape, reducing tourist noise and emphasizing ritual and natural elements to create a more serene, imagined temple atmosphere. In both cases, selecting, ordering, and combining sounds prompted participants to reflect on what they saw as essential to the temple's identity and how different sonic layers should coexist.

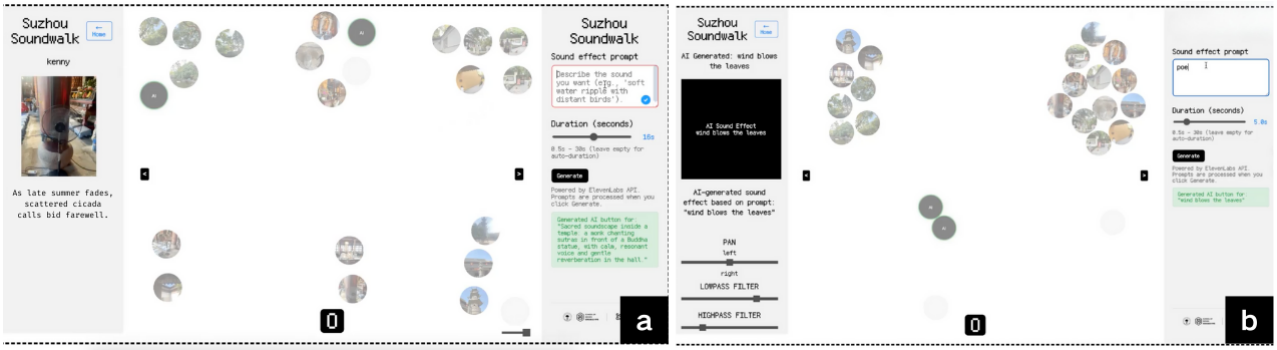

Figure 5. Example of clustered audio clips with generated sound.

4.3 AIGC as a Supplementary Co-creative Layer (RQ3)

Participants used the text-to-audio module sparingly but deliberately. Logs and interviews indicate two main uses: to fill gaps (e.g., replacing missing or low-quality recordings) and to materialise imagined elements that fit their mental image of a temple, such as distant bells or larger chanting groups. AI-generated clips were always mixed with, rather than replacing, field recordings and were seen as helping to make the soundscape more coherent or expressive. At the same time, participants questioned the quality and character of the generated audio, describing it as "flatter" and "more mechanical", with less subtle variation than in-situ recordings; synthetic chanting in particular was sometimes experienced as uncannily song-like. Overall, they positioned AIGC as a supporting layer for completing and stylising the arrangement, while treating field recordings as the primary carriers of presence, warmth, and authenticity.

5. CONCLUSION

We presented a study in which visitors to *Xiyuan Temple* co-created sonic interpretations of the site using their own field recordings together with a text-to-audio module. Participants selectively focused on temple-specific and natural sounds as key markers of place and used the interface both to reconstruct their walk and to idealize the soundscape. GenAI was mainly employed to fill gaps and realize imagined elements—such as clearer cicadas, temple bells, or amplified chanting—while real recordings remained the primary carriers of presence, warmth, and authenticity. Overall, the study suggests that GenAI can support creative cultural engagement when embedded in human-centred sound co-creation and highlights design opportunities for AI-mediated soundscape systems in cultural heritage contexts.

ACKNOWLEDGMENTS

We thank our participants for their time and efforts. This study is supported by the Key Laboratory of Intelligent Processing Technology for Digital Music (Zhejiang Conservatory of Music), Ministry of Culture and Tourism (Grant Number: 2023DMKLC006).

REFERENCES

[1] Schafer, R. M., [*The soundscape: Our sonic environment and the tuning of the world*] (1993).
[2] Brown, A. L., "Advancing the concepts of soundscapes and soundscape planning," in [*Acoustics 2011: Annual Conference of the Australian Acoustical Society*], Australian Acoustical Society (2011).
[3] Freeman, J., Disalvo, C., Nitsche, M., and Garrett, S., "Soundscape composition and field recording as a platform for collaborative creativity," *Organised Sound* **16**(3), 272–281 (2011).
[4] Radicchi, A., Henckel, D., Memmel, M., et al., "Citizens as smart, active sensors for a quiet and just city: the case of the "open source soundscapes" approach to identify, assess and plan "everyday quiet areas" in cities," *Noise mapping* **5**(1) (2018).
[5] Fu, X., Li, S., and Sagesser, M. Z., "Noise in the high-tech city: Listening to shenzhen through a participatory phone-based app," in [*Proceedings of 11th Convention of the European Acoustics Association*], 1–8 (2025).
[6] Brambilla, G. and Pedrielli, F., "Smartphone-based participatory soundscape mapping for a more sustainable acoustic environment," *Sustainability* **12**(19), 7899 (2020).
[7] Radicchi, A., "Hush city: a new mobile application to crowdsource and assess' everyday quiet areas' in cities," in [*Proceedings of invisible places: the international conference on sound, urbanism and the sense of place*], 7–9 (2017).
[8] DiSalvo, C., Freeman, J., and Nitsche, M., "Participatory art as inner city workshop: The urbanremix sound project,"
[9] Zhang, Y., Ji, N., Zhu, X., and Zhao, Y., "Inheritance and revitalization: Exploring the synergy between aigc technologies and chinese traditional culture," in [*International Conference on AI-generated Content*], 24–32 (2023).
[10] Lai, S., Tian, Y., and Zhang, Q., "The impact of ai-generated technologies-driven digital cultural heritage platforms on users' offline cultural participation intentions," *npj Heritage Science* **13**(1), 574 (2025).
[11] "ISO 12913-1:2014 Acoustics — Soundscape — Part 1: Definition and conceptual framework," (2014). Retrieved from https://www.iso.org/standard/52161.html.