Exploring AI Audio Models in Soundwalking with Broader Audiences

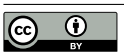
Jiatong Liu
Creative Computing Institute, University of the Arts London
London, United Kingdom
jiatong.liu37@outlook.com

Shuoyang Jasper Zheng
Creative Computing Institute, University of the Arts London
Centre for Digital Music, Queen Mary University of London
London, United Kingdom
z.shuoyang@arts.ac.uk

Figure 1: Project overview image.

Abstract

The development of generative AI offers new possibilities for music composition and production. However, it could undermine the creative process and obscure the generation steps between input and output, and lead to homogenous results. To enable broader audiences to engage with generative AI models, we adopt soundwalking—an embodied listening and compositional practice — as a design principle, and apply meaningful human control to its AI integration. Together, these frameworks can encourage the player's exploration and understanding of the sonic outcome. This demo presents an immersive soundwalking experience created using an AI audio synthesis model. The audience walks in the virtual environment as their movements transform into soundscapes in real time. The interactive system is built in an accessible and playful way for a broader gallery audience to engage with AI models through embodied listening.



C&C '26, London, United Kingdom

ACM ISBN 979-8-4007-2583-8/2026/07
https://doi.org/10.1145/3803784.3816887

CCS Concepts

• **Applied computing** → **Media arts**; • **Human-centered computing** → *Interactive systems and tools*; • **Computing methodologies** → *Artificial intelligence.*

Keywords

Soundwalking, Artistic Installation, Autoencoder, AI as Material, Machine Learning

ACM Reference Format:
Jiatong Liu and Shuoyang Jasper Zheng. 2026. Exploring AI Audio Models in Soundwalking with Broader Audiences. In *Creativity and Cognition (C&C '26), July 13–16, 2026, London, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3803784.3816887

1 Introduction

Recent developments in AI and audio offer a range of creative possibilities for musicians in their practices, from quick prototyping [22] and tools for mastering [14] to novel musical affordances [12]. However, many AI audio tools fall under a "Big Red Button"—where the system receives a high-level input from the user and generates a polished outcome [21]. This reduces the interaction with AI models

to a "black box" [4], raising various concerns such as artists' understanding of the creative artifact, homogeneity in generated outputs, and the resemblance to the training data [7].

In this demo we explore a creative application of generative AI models with the practice of soundwalking. Soundwalking is an act of combining human mobility and listening [2]. It can offer a way of experiencing space mediated and transformed by computation [6].

Our approach is grounded in practice-based research, where the first author, a sound artist, approaches generative AI from an artistic perspective. The methodology focuses on knowledge in the practice rather than the technical artifact, with iterative reflection across build sessions.

The demo takes the form of a virtual soundwalking experience in which the sound space is crafted with an AI sound synthesis model. The source audio materials to create this sound space are field recordings of soundscapes from Beijing's Hutongs—a type of traditional northern Chinese courtyard house. The audience wanders through and hears the real-time generated soundscape according to their movements. Our aim is to use soundwalking's embodied, co-creative nature to promote understanding and exploration of generative AI models.

2 Related Works

The principles being applied are those of soundwalking, a listening and composition practice used in both real-world and computational environments. We will outline its background, rationale, and related artworks integrating soundwalking and AI.

2.1 Soundwalk: Background and Rationale

Soundwalking is any excursion whose main purpose is listening to the environment [23], foregrounding the relationship between a specific form of human mobility — walking — and a specific way of sensory attention — listening [2]. It reframes the listener's role, shifting the audience from a passive observer to a "composer-performer" [18], as each soundscape requires the active participation of the listener. In artistic installations, soundwalking can be used to enable artist-audience communication [15].

Soundwalking has also been applied in computational contexts, where it can facilitate the exploration and understanding of digital systems [6]. Computational soundscapes enforce reduced listening — removing sound from its source [17]. Thus, the relationship between sound and origin is reframed [9], presenting the audience with a novel and unfamiliar sonic experience. Concurrently, the movement-to-sound feedback loop introduces human embodiment in digital environments. Embodiment is positioned as central to musical understanding in cognitive science, as perceiving and physically making music recruit the same parts of the brain [3]. Recently, embodiment has been applied as a tool for explainability in education, used in a museum setting to improve AI literacy [8]. Finally, the artistic nature of soundwalking could help develop AI literacy for the public [11]. As an interactive installation, it serves as an educational medium, explaining the computational materiality of AI under intuitive and engaging formats [5]. Together, these qualities suggest soundwalking could serve as a framework for both creative exploration and explainability in AI systems.

2.2 Soundwalking with AI Sound Synthesis Models

Several AI-driven artworks demonstrate the reciprocal relationship between movement and listening, central to soundwalking [2].

"Exploring Gestural Affordances in Audio Latent Space Navigation" [24] investigates the creative potential of navigating generative AI latent spaces through gesture. Zheng et al. describe a workshop in which musicians tested open-ended gestures and composed scores using an AI-enhanced digital instrument, identifying emergent composition techniques such as sampling, looping, and synthesis.

"Embodied Exploration of Deep Latent Spaces in Interactive Dance-Music Performance" [16] examines the use of generative AI models in interactive dance performance. The performer's movement is connected to an AI model via a motion sensor. The study explores various movement-sound mapping strategies, and evaluates the dancer's perception of each, which shifted from feeling "inside the machine" to being a partner with it.

"Soundwalking Deep Latent Spaces" [19] is a computational soundwalk that connects a player's coordinates to the latent space of an AI model. Unlike gesture-based interactions, which operate on a short timescale, Scurto and Postel investigate how soundwalking can reveal the structure of sound across larger scales of time and space, implying different embodied listenings. Furthermore, the uncanny qualities of AI-generated sound could foster new AI practices without impinging on existing musical labor.

3 Technology

3.1 Design Approach

Our design applies **meaningful human control** to the virtual soundwalk, building on Akten's framework for applying deep learning models as an artistic medium for performative creative expression [1]. The notion requires three necessary and sufficient conditions: *Intent* — the system carries out the human's intent in its outcome; *Predictability* — the process from input to outcome is understandable to the user; and *Accountability* — users interacting with the system should receive accountability for the outcome, not give it solely to the algorithm.

In the context of this demo, the final soundscape should preserve the theme and overall feeling of the original audio (intent), remain partially recognizable to its source material (predictability), and be tied directly to the player's movement decisions (accountability). This gives the player direct influence and control over the sonic outcome, encouraging creative exploration and understanding of the system.

3.2 Autoencoders, Latent Space and Latent Terrain

An autoencoder is a type of AI model that consists of an encoder that compresses input data into a small and compact representation and a decoder that reconstructs it to its original form [13]. Autoencoders offer unique musical affordances and opened up a plethora of interactive sonic strategies in the field of NIME [25].

The layer inside the autoencoder with the most compact representation is called the *latent space*, a multi-dimensional vector space

learned from a corpus of audio data [10]. However, attributes in a latent space are encoded in ways that are not easy to understand by humans [4]. Existing practice of engaging with the latent space therefore typically relies on establishing meaningful human control over the sound generation.

Latent Terrain is a method to tailor latent spaces into corpus-based sound spaces [25]. It maps two-dimensional coordinates to high-dimensional vectors in the latent space. When the coordinates move on the canvas, the system immediately retrieves its latent vector, generating its corresponding sound [24]. The mapping of position and sound makes Latent Terrain well suited for implementing meaningful human control in a soundwalk.

4 Realisation

4.1 Material Collection

The audio material consists of field recordings and focused recording sessions of traditional Hutong soundscapes captured by the first author at historical sites in eastern and western Beijing. Additional sound samples were sourced from open-access archives where suitable recordings of specific Hutong sounds were not feasible to self-record. The visual material consists of 3D scans captured by the first author at the same sites.

4.2 Latent Space Mapping

Latent terrain is utilised to map audio into a 2D plane: it maps audio into lines — its length corresponding to its duration — which is then evenly laid out onto the plane in the form of a spiral. If the player were to perfectly walk from the beginning to the end of the spiral, the sound will play chronologically. In practice, the sound will morph between its own sections (for example second 5 to second 1), not playing linearly but rather spatially. Each sound has its own region, corresponding to what the player will see in the virtual environment.

See Figure 2 for the sonic region map.

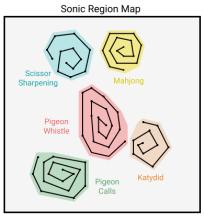


Figure 2: Sonic region map showing the spatial distribution of sounds across the virtual environment.

4.3 Localising Sound

Moving within a certain region of the soundwalk will trigger its corresponding soundscape, whereas leaving that area results in silence. To achieve this, we encapsulated sounds to their region by adding silent latent vectors around the existing sound spirals, acting as a sonic barrier to the area. The silent latent vectors are visualised in Figure 3 on the sonic region map.

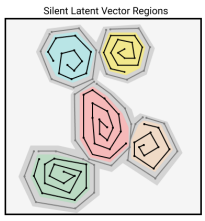


Figure 3: Visualization of the silent latent vectors on the sonic region map.

A helper tool called Latent Terrain Brush was developed under this research to remove any unwanted noise from the system. Latent Terrain GUI visualizes the sonic output to a 2D plane, then Latent Terrain Brush is used to erase the noise graphically: the user uses the paint tool in the GUI to create a mask. If the coordinates fall under the mask, the sound becomes muted.

This was achieved by building a paint tool in Jitter — MAX's open-ended toolkit for graphics. The coordinates of the paint tool have been aligned to the coordinates of Latent Terrain GUI. The amplitude of the sonic outcome is scaled to the luminosity of the pixel the user is on.

Figure 4 showcases the UI of Latent Terrain Brush, with Latent Terrain GUI on the left and the painted mask on the right.

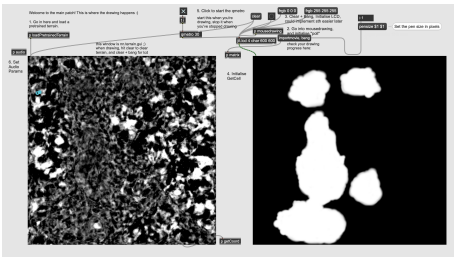


Figure 4: UI of Latent Terrain Brush.

4.4 Virtual Environment

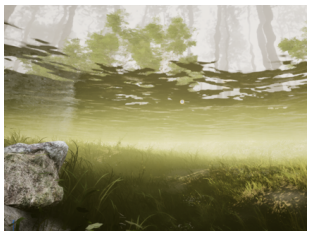
The audio system was connected to Unreal Engine via OSC, where the player character's coordinates were mapped to the coordinates in Latent Terrain.

Each region in the soundwalk has a visual theme, corresponding to the sonic region plan, with 3D scans of the local area working as visual assets in the game environment. See Table 1 for the background and rationale behind each sonic theme.

5 Demonstration Description

The demonstration will be an interactive installation. The audience will stand in front of a monitor and put on studio headphones. Using keyboard (WASD keys) and mouse, they can choose to walk, run, stop or turn around in the virtual environment. As this happens, they will hear the corresponding soundscape in real-time. A companion sound map (see Figure 5) will be provided alongside the installation, providing details about the sound type, their story and

Table 1: Level Design for Soundwalk

Visual	Sound Type	Background and rationale
	Scissor Sharpening	Street merchants make scissor sharpening sounds to call out residents to sharpen household blades. The area was designed with 3D scans of Hutong streets, emulating the trajectory of the merchant.
	Mahjong	Mahjong is played in Beijing as a hobby or in times of celebration, presented here with a self-collected 3D scan.
	Pigeon Whistle	The sound of whistles attached to pigeons at flight. A lake was made to match its fluid, pad-like sonic quality.
	Katydid	Katydids are kept as pets by retired men. Flower bushes are laid out, where katydids play and nest.
	Pigeon calls	Pigeon keeping is a communal hobby. The pigeon calls recorded in courtyard birdhouses are spatially mapped onto a cliff populated with pigeons.

location in the virtual world. The map design is inspired by Johana Knowles's leaflet from the "Sensing the Forest project"[20].

Future iterations will move beyond keyboard-and-mouse input towards alternative modalities, such as physical controllers or full-body tracking, better aligning with soundwalking's embodied, walking-based principles.

6 Conclusion

In response to concerns raised over the development of generative AI and sound, we presented a virtual soundwalk as an alternative to the "Big Red Button" approach to sound generation. We applied meaningful human control to soundwalking, using AI as an artistic medium for creative expression. We selected Latent Terrain as the central technology, enabling us to map and localize field recordings to a virtual environment so that each sound corresponds to what the player sees. In a playful way, the interactive system could increase

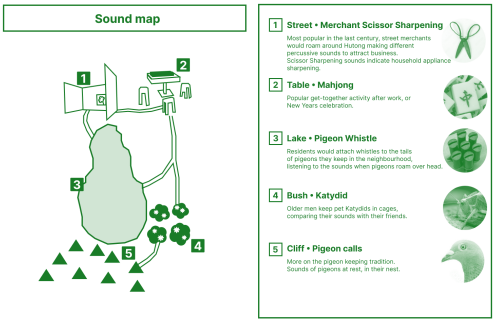


Figure 5: Installation sound map inspired by Johana Knowles's soundwalking map from the "Sensing the Forest project"[20].

a player's exploration and understanding of the AI-assisted sonic outcome for a broader audience.

References

[1] Memo Akten. 2021. *Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression.* Doctoral thesis. Goldsmiths, University of London. https://research.gold.ac.uk/id/eprint/30191/

[2] Frauke Behrendt. 2019. Soundwalking. In *The Routledge Companion to Sound Studies*, Michael Bull (Ed.). Routledge, New York, NY, 249–257. doi:10.4324/9781315722191-28

[3] Wayne Bowman. 2004. Cognition and the Body: Perspectives from Music Education. In *Knowing Bodies, Moving Minds*, Liora Bresler (Ed.). Landscapes: The Arts, Aesthetics, and Education, Vol. 3. Springer, Dordrecht, 29–50. doi:10.1007/978-1-4020-2023-0_3

[4] Nick Bryan-Kinns, Bingyuan Zhang, Songyan Zhao, and Berker Banar. 2024. Exploring Variational Auto-encoder Architectures, Configurations, and Datasets for Generative Music Explainable AI. *Machine Intelligence Research* 21, 1 (2024), 29–45. doi:10.1007/s11633-023-1457-1

[5] Nick Bryan-Kinns, Shuoyang Jasper Zheng, Francisco Castro, Makayla Lewis, Jia-Rey Chang, Gabriel Viglianson, Terence Broad, Michael Paul Clemens, and Elizabeth Wilson. 2025. XAIxArts Manifesto: Explainable AI for the Arts. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25).* Association for Computing Machinery, New York, NY, USA. doi:10.1145/3706599.3716227

[6] Miguel Carvalhais and Rosemary Lee. 2019. Soundwalking and Algorithmic Listening. In *Proceedings of RE:SOUND 2019 – 8th International Conference on Media Art, Science, and Technology (Electronic Workshops in Computing (eWiC)).* BCS Learning and Development Ltd., Aalborg, Denmark, 51–56. doi:10.14236/ewic/RESOUND19.8

[7] Adam Cole. 2022. *Old Sights, New Visions: Controlled Uses of Diffusion Based Image-to-Image Translation for Generative Video.* Technical Report. University of the Arts London, London. https://ualresearchonline.arts.ac.uk/id/eprint/20056/ Unpublished.

[8] Hasti Darabipourshiraz, Dev Ambani, and Duri Long. 2024. DataBites: An embodied and co-creative museum exhibit to foster children's understanding of supervised machine learning. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) *(C&C '24).* Association for Computing Machinery, New York, NY, USA, 550–555. doi:10.1145/3635636.3664247

[9] Joanna Demers. 2010. *Listening Through the Noise: The Aesthetics of Experimental Electronic Music.* Oxford University Press, New York.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* MIT Press. http://www.deeplearningbook.org.

[11] Drew Hemment, Morgan Currie, SJ Bennett, Jake Elwes, Anna Ridler, Caroline Sinders, Matjaz Vidmar, Robin Hill, and Holly Warner. 2023. AI in the Public Eye: Investigating Public AI Literacy Through AI Art. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23).* Association for Computing Machinery, New York, NY, USA, 931–942. doi:10.1145/3593013.3594052

[12] Purnima Kamath, Fabio Morreale, Priambudi Lintang Bagaskara, Yize Wei, and Suranga Nanayakkara. 2024. Sound Designer-Generative AI Interactions: Towards Designing Creative Support Tools for Professional Sound Designers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 730, 17 pages. doi:10.1145/3613904.3642040

[13] Daniel Manz and Mick Grierson. 2025. Brave: Designing an Embedded Network-Bending Instrument, Manifesting Output Diversity in Neural Audio Systems. In *Proceedings of the 16th International Conference on Computational Creativity (ICCC '25).* São Paulo, Brazil. https://ualresearchonline.arts.ac.uk/id/eprint/24221/

[14] Marco A. Martínez-Ramírez, Wei-Hsiang Liao, Giorgio Fabbro, Stefan Uhlich, Chihiro Nagashima, and Yuki Mitsufuji. 2022. Automatic Music Mixing with Deep Learning and Out-of-Domain Data. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference.* Bengaluru, India, 411–418. https://archives.ismir.net/ismir2022/paper/000049.pdf

[15] Andra McCartney. 2014. Soundwalking: Creating Moving Environmental Sound Narratives. In *The Oxford Handbook of Mobile Music Studies, Volume 2*, Sumanth Gopinath and Jason Stanyek (Eds.). Oxford University Press, 212–237. doi:10.1093/oxfordhb/9780199913657.013.008

[16] Sarah Nabi, Philippe Esling, Geoffroy Peeters, and Frédéric Bevilacqua. 2024. Embodied exploration of deep latent spaces in interactive dance-music performance. In *Proceedings of the 9th International Conference on Movement and Computing* (Utrecht, Netherlands) *(MOCO '24).* Association for Computing Machinery, New York, NY, USA, Article 12, 9 pages. doi:10.1145/3658852.3659072

[17] Pierre Schaeffer. 2012. *In Search of a Concrete Music.* University of California Press, Berkeley, CA.

[18] R. Murray Schafer. 1977. *The Tuning of the World.* Alfred A. Knopf, New York.

[19] Hugo Scurto and Ludmila Postel. 2023. Soundwalking Deep Latent Spaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '23)*, Miguel Ortiz and Adnan Marquez-Borbon (Eds.). Mexico City, Mexico, 232–235. doi:10.5281/zenodo.11189166

[20] Sensing the Forest. 2023. Sensing the Forest: Visitor Leaflet. https://sensingtheforest.github.io/. AHRC-funded project.

[21] Nicholas Shaheed and Ge Wang. 2024. I Am Sitting in a (Latent) Room. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* Utrecht, Netherlands, 333–338. doi:10.5281/zenodo.13904872

[22] Bob L. Sturm, Oded Ben-Tal, Úna Monaghan, Nick Collins, Dorien Herremans, Elaine Chew, Gaëtan Hadjeres, Emmanuel Deruty, and François Pachet. 2019. Machine Learning Research That Matters for Music Creation: A Case Study in Musical Palette Transfer. *Journal of New Music Research* 48, 1 (2019), 36–55. doi:10.1080/09298215.2018.1515233

[23] Hildegard Westerkamp. 1974. Soundwalking. *Sound Heritage* 3, 4 (1974), 18–27. Revised 2001. Available at: https://www.sfu.ca/sonic-studio-webdav/WSP_Doc/Booklets/SHWesterkamp.pdf.

[24] Shuoyang Jasper Zheng, Anna Xambó Sedó, and Nick Bryan-Kinns. 2025. Exploring Gestural Affordances in Audio Latent Space Navigation. *Frontiers in Computer Science* 7 (2025). doi:10.3389/fcomp.2025.1575202

[25] Shuoyang Jasper Zheng, Keigo Yoshida, Nico García-Peguinho, Jiatong Liu, Dan Hearn, Anna Xambó Sedó, and Nick Bryan-Kinns. 2026. Latent Terrain: Adapting Neural Audio Autoencoders as Design Materials in NIME. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* London, UK.