



## OPEN ACCESS

## EDITED BY

Bruno Mesz,  
National University of Tres de Febrero,  
Argentina

## REVIEWED BY

Kevin Ryan,  
The University of Tennessee, Knoxville,  
United States  
Steve Goschnick,  
The University of Melbourne, Australia

## \*CORRESPONDENCE

Shuoyang Jasper Zheng  
✉ shuoyang.zheng@qmul.ac.uk

RECEIVED 12 February 2025

ACCEPTED 15 September 2025

PUBLISHED 07 November 2025

## CITATION

Zheng SJ, Xambó Sedó A and Bryan-Kinns N  
(2025) Exploring gestural affordances in audio  
latent space navigation.  
*Front. Comput. Sci.* 7:1575202.  
doi: 10.3389/fcomp.2025.1575202

## COPYRIGHT

© 2025 Zheng, Xambó Sedó and Bryan-Kinns.  
This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that  
the original publication in this journal is cited,  
in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Exploring gestural affordances in audio latent space navigation

Shuoyang Jasper Zheng<sup>1\*</sup>, Anna Xambó Sedó<sup>1</sup> and  
Nick Bryan-Kinns<sup>2</sup>

<sup>1</sup>Centre for Digital Music, Queen Mary University of London, London, United Kingdom, <sup>2</sup>Creative Computing Institute, University of the Arts London, London, United Kingdom

The latent space of generative AI models affords unique creative possibilities and broad design space for AI-enhanced digital music instruments. While interface designs for latent space navigation typically rely on sound-producing gestures that involve bodily motions and movements, the underlying subjective perception of these gestures remains underexplored. To understand how musicians perceive sound-producing gestures and tailor performance techniques in audio latent space, we present a user study workshop with an AI-enhanced digital music instrument with a tablet interface. Eighteen musicians were recruited to test out open-ended gestures and tasked to create musical scores. We report how they use sound-producing gestures in the latent space and develop performance techniques. We contribute findings from an embodied music cognition perspective of how subjective perception of gestures shapes musicians' technique development in audio latent space navigation. We discuss the implications of new gestural affordances discovered by participants in our workshop, aiming to elucidate new opportunities for digital musical instruments with audio latent space navigation.

## KEYWORDS

AI musical instruments, new interfaces for musical expression, neural audio synthesis, sound and music computing, embodied music cognition, digital musical instruments

## 1 Introduction

The latent space of generative AI models affords unique creative possibilities in the domain of creative AI for music and sound (Yee-King, 2022). Various attempts have been made to investigate how latent space can be integrated into Digital Music Instruments (DMIs) for musical ideas and expressions, such as to be effectively navigated with novel interfaces (Tahiroğlu et al., 2021; Privato et al., 2024), with explainable controls (Vigliensoni and Fiebrink, 2023; Kamath et al., 2024), or as an artistic material (Wilson et al., 2023; Shaheed and Wang, 2024). Observing how musicians' actions of navigating the latent space in these works relying on gestures with bodily movements, makes us believe that the perception of sound-producing gestures in the latent space is an important perspective for investigating its creative possibilities. However, there is little research focusing on the links between auditory and kinesthetic perceptions of gestural motions (Godøy, 2009) in audio latent spaces navigation. In this respect, how musicians discover gestural affordances in the latent space and tailor performance techniques based on these affordances remains underexplored.

According to Godøy and Leman (2009), musical gestures on an instrument are embodied movements concerning subjective bodily experiences. This links to embodied music cognition (Leman, 2007)[p. 95] that considers the role of kinesthetic and sensorimotor experiences in musical activities. Magnusson (2010) suggests that bodily

experience mediates the link between a musical instrument's expressive potential and a musician's engagement with it. This provides one of the key entry points for designing and evaluating New Interfaces for Musical Expression (NIME), that is, to investigate how musicians discover and test out the affordances of an instrument and develop them into a set of performance techniques (Rodger et al., 2020). This paradigm has guided the development of new musical instruments (Magnusson, 2009; Bertissolo, 2019). Given the recent development of generative AI for music and sound, gestural affordances open a prominent embodied perspective on investigating how audio latent space navigation can be harnessed into, or used as, new musical instruments. This motivated us to explore sound-producing gestures afforded in the latent space of AI audio synthesis models.

In this article, we designed a DMI with a stylus-tablet interface as a research probe (Tahiroğlu et al., 2020). The instrument embeds *Latent Terrain*, an adapted form of latent space of a neural audio synthesis model, inspired by wave terrain synthesis (Mitsubishi, 1982). One can navigate the latent terrain using gestures afforded by the stylus and tablet. We present a workshop in which 18 participants explored two latent terrains. Our aim is not to compare the effects of these two terrains, but rather to use them as prompts to invite musicians to actively test out open-ended gestural movements based on their capabilities and curiosities. With data from interviews, participants' documentary notes and demonstration of their musical creations, we aim to gain insight into musicians' exploration in audio latent space, and map out a complex relationship of gestural affordance, subjective perception, and the resulting deployment of techniques for musical expressions (Figure 1).

In summary, we make three contributions:

- Insights into how subjective perception of sound-producing gestures shapes musicians' performance technique development in audio latent space, to complement studies on gestural interfaces for latent space navigation (Lepri et al., 2024; Privato et al., 2024).
- Insights into gestural affordances discovered by participants, aiming to elucidate new opportunities for future developments of gestural interface for audio latent space navigation, and complement studies on new musical affordances of the latent spaces (Yee-King, 2022; Tahiroğlu and Wyse, 2024).
- A *documentary notes* method (see Section 4.3.1) adapted from the soma trajectories tool (Tennent et al., 2021) and body map (Anne Cochrane et al., 2022), in which participants take notes of in-the-moment experiences in their exploration process. Recorded notes are used by participants to recall their interaction trajectory.

The paper is structured as follows. Section 2 summarizes relevant literature to contextualize our study. Section 3 describes the design and configuration of the instrument and the encapsulated latent terrains. Section 4 outlines the method of our study. Section 5 presents the results and Section 6 discusses our findings in the context of the literature.

## 2 Background

Embodied music cognition highlights the role of bodily experience in shaping musical interaction, particularly bodily movements that engage with sound-producing gestures (Leman, 2007, p. 95). The bodily movement aspect of embodied music cognition has informed the design and evaluation of NIME (Bertissolo, 2019; Erdem and Jensenius, 2020; Mice and McPherson, 2022). It emphasizes how subjective experiences of the body, including the sense of kinesthetic movement (Godøy and Leman, 2009, p. 154), sensorimotor coupling (Godøy and Leman, 2009, p. 212), and listening, can shape both the performance and experience of musical gestures. The ecological aspect of a person's bodily subjective experience and the environment is highlighted (Leman, 2007, p. 51). This has motivated our research to approach embodied music cognition through the notion of affordance, to understand the relationship between body, experience, and sound.

### 2.1 Gestural affordances in NIMEs

The definition of affordances in Human-Computer Interaction (HCI) varies across the literature (López-Cano, 2006). Gibson (1979) took an ecological approach that defined it as actionable possibilities an environment offers to subjects. Building on this framework, in music perception, Clarke (2005), p. 204 argues that musical structures afford a range of interpretive and embodied responses. Similarly, Reybrouck (2005) elaborates that these action possibilities, especially in terms of sensorimotor engagement, can be offered by a listener's musical stimuli. In this respect, gestural affordances for musical expressions can be singular sound-producing actions such as hitting and blowing, or compound actions such as drumming rhythmic patterns (Reybrouck, 2005). We consider the gestural affordances perspective because it mediates embodied movements, musical expressions resulting from these movements, and the mental simulation of this coupling. In addition, de Vignemont (2015) develops the notion of affordances to the domain of bodily experience. She proposes that possibilities for movement and action are determined by the body's own structural and postural organization. In NIME design, musical and bodily affordances are interwoven to navigate the space of actionable possibilities (Dagleish, 2014; Nijs et al., 2024).

Elucidating the affordance space of NIME in an open-ended setting is a prominent entry point for analyzing how it is used or appropriated in musical practice. Magnusson (2010) suggests that the affordances of a musical instrument should be considered as a configuration of properties that constrain the instrument's expressive potential. While NIME designs can be shaped by the effectivities of affordances, it is common that musical practice on an instrument goes beyond the constraints and guidelines that it was originally intended (Dix, 2007). This calls for the need to consider open-ended affordances that are flexible and tailorable to individual musicians' skills and needs (Xambó Sedó, 2023). Practically, the evaluation of NIMEs following this paradigm encompasses the investigation of musicians' "exploratory information seeking" (Rodger et al., 2020) process, in which affordances are actively

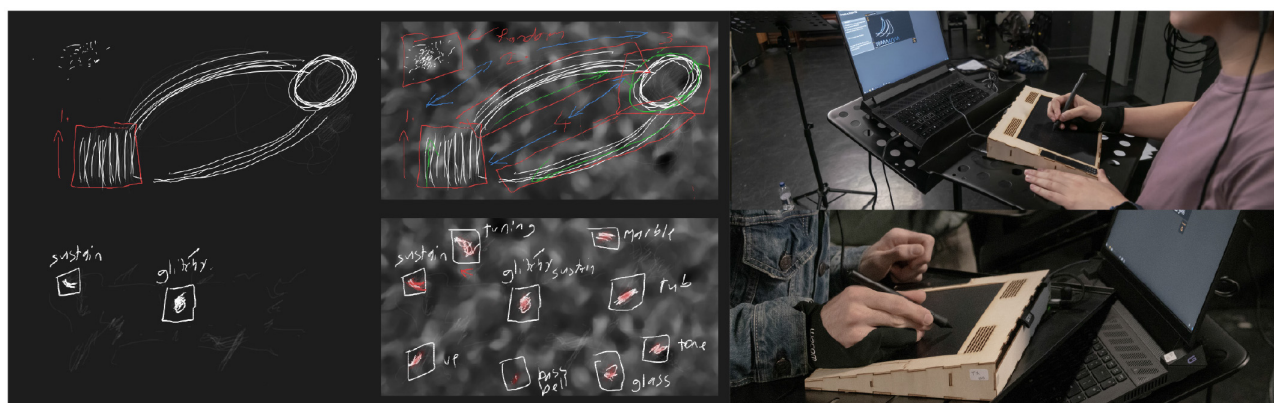


FIGURE 1

Work-in-progress musical scores created by two workshop participants (left); completed musical scores superimposed on the latent terrain (middle); workshop participants demonstrating their scores on the prototyped musical instrument (right).

tested, tailored, and deployed into strategies to perform. This paradigm has guided a number of works in the design and evaluation of NIMEs (Zappi and McPherson, 2018; Mice and McPherson, 2022).

Research methods and tools have been designed to investigate gestural affordances in NIMEs. Rodger et al. (2020) propose the viewpoint from musicians' *exploratory* and *performatory* modes of engagement. In particular, musicians actively test out the affordances of the instrument based on their skills and curiosities in the exploratory mode, and develop a bundle of sound-producing gestures in the performatory mode (Stapleton et al., 2018). Affordances in these modes are "not linear sums of sonic capacities" (Rodger et al., 2020), instead, they may be repurposed or discarded, and new affordances may emerge along the sense-making process. To put this into an operative notion, Godøy (2006)'s concept of *gestural sonic objects* describes snippets of sound-based musical materials in the 0.5–5 seconds duration range (Godøy, 2018) and the sound-producing gestures that perform it. The concept of gestural sonic objects is a useful tool for analyzing instruments in musical practice because it mediates auditory and kinesthetic perceptions of the instruments (Visi et al., 2024), and allows the investigation of musical techniques and repertoires to enter from an embodied perception perspective (Godøy, 2006).

In addition, capturing and assessing subjective perceptions is important in analyzing gestural affordances in musical practice because bodily experiences can guide musicians' embodied exploration (de Vignemont, 2015) and their development of techniques (Mice and McPherson, 2022). A growing interest in HCI focuses on tools for articulating subjective in-the-moment experience for research purposes (Núñez-Pacheco, 2021). One commonly used approach, body map (Anne Cochrane et al., 2022), aims to capture implicit bodily sensations by human subjects' self-reported documentation. In addition, the concept of *interaction trajectories* (Fitzpatrick, 2003, p. 120) emerged in HCI to explain how the experience changes at different points in the process of interaction. Benford et al. (2009) proposed conceptual frameworks to distill knowledge into design guidelines and patterns. In the context of embodied interaction, Tennent et al. (2021) developed

the soma trajectories tool to help human subjects capture the progression of their bodily experiences, and it was widely used in evaluating embodied experience with NIME (Avila et al., 2020). In our research, we adapt the soma trajectories tool for recall analysis of music performance experiences, and follow Visi et al. (2024)'s use of gestural sonic objects to ensure validity and rigor when articulating subjectivities (Ståhl et al., 2021).

## 2.2 Latent space navigation for musical sound creation

The latent space of generative AI models can be seen as a set of control parameters learned from a large corpus of data (Goodfellow et al., 2016, p. 501). Despite being difficult for humans to interpret the meanings of each parameter (Bryan-Kinns et al., 2024), they still offer unique creative possibilities in the domain of creative AI for music and sound due to their ability to encode raw audio data into a significantly smaller number of parameters (Yee-King, 2022). Notably, Neural Audio Synthesis (NAS), a method for generating audio waveforms using deep learning AI models, such as Realtime Audio Variational autoEncoder (RAVE) (Caillon and Esling, 2021), tackles the audio modeling task by (i) encoding a fragment of audio waveform into a sequence of vectors in the latent space, which typically has a low sampling rate, and a vector dimension ranges from 8 to 32 after regularization. Then (ii) decoding the sequence of latent vectors into a fragment of audio waveforms, which should reconstruct the input fragment.

The decoder of a trained NAS model can be used as an audio synthesizer with a parametric interface if the latent vector in the latent space is manually manipulated. This has been referred to as *latent space navigation* by practitioners in HCI, NIME, and Generative AI (Scurto and Postel, 2023; Wan and Lu, 2023; Bryan-Kinns, 2024).

Latent space navigation has become a broad design space for the human control interface. Typical approaches for navigation have explored reducing the latent space dimensionality to an appropriate

number manageable by a musician's cognitive bandwidth (Zappi and McPherson, 2018), and embedding it into parametric control mechanisms such as sliders or XY-pads (Roma et al., 2019). In the field of explainable AI, Bryan-Kinns et al. (2024) and Vigliensoni and Fiebrink (2023) explored ways of making the navigation understandable. Beyond parametric control methods, Zheng et al. (2024b) proposed using more abstract materials and Shaheed and Wang (2024) explored live coders' improvisation as ways of steering the navigation. We present our adaptation to the latent space to support navigation on a low-dimensional control space in Section 3.

## 2.3 Gestural affordances in latent space navigation

Latent space as a platform for discovering gestural affordances, despite being nascent, has yielded a broad design space for applications in sound, movement, and musical expression (Yee-King, 2022). However, research in AI and musical sound has been dominantly focused on technocentric aspects for solving and assisting musical tasks (Huzaifah and Wyse, 2021). The interactivity, in particular, the sound-producing gestures and their motion-sound experience (Godøy, 2018), remains underexplored.

Gestural affordances in an audio latent space are the results of a complex relationship between sounds, gestural perceptions, the navigator's sensorimotor skills, and their musical intentions and goals. Given a plain multi-dimensional control space (e.g., a 2D touchpad or a 3D accelerometer) without auditory outputs, there are very few constraints on how it can be navigated. To name a few, it affords impulsive strikes, sustained steering, or rapid back-and-forth motions. However, when these actions are used as sound-producing gestures in a latent space, the perceived sound outcomes further define their musical affordances (Rodger et al., 2020), and are kept in musicians' memory and used for the next action (Leman, 2012).

Here we illustrate how gestural affordances might occur when latent space navigation is used in musical practices by introducing examples of research on musicians' explorations of audio latent spaces. Tahiroğlu et al. (2021) designed a non-rigid and stiff physical interface to explore hand-held musical gestures, to study the use of various levels of pressure, rate of change in pressure, and bending. Privato et al. (2024) designed a board interface with magnetic attractors to explore the gestures of arranging the magnetic objects, and found that algorithmic adaptation in the latent space can affect the perception of sound-producing gestures. Scurto and Postel (2023) explored using spatial coordinates in a 3D virtual environment to navigate the latent space, and reflected on the social and aesthetic implications of embodied listening along the navigation.

Observing how these works explore distinctly different materials, modes of interaction, and scales, we note that the design of the control interface, inscribed with designers' assumptions and theoretical backgrounds (Kuzmin et al., 2024), has provided rich guidelines and constraints on what gestural affordances are perceived. However, how can we study latent space navigation in a purposefully simple, open, and less constrained way to enable tailorable and flexible developments of performance techniques?

In addition, challenges that emerged in these works are typically centered around latent space's nature of being complex, high-dimensional, and difficult for humans to interpret (Bryan-Kinns et al., 2024). Although dimensionality reduction methods can effectively help the navigation to be simple and utility, they raise the longstanding discussion on the balance between utility and expressiveness in Creativity Support Tools (CSTs) (Jacobs et al., 2017). In simple terms, linear mappings between gestures in a 2D touchpad or 3D accelerometer's inputs to each latent dimension are not enough to fully exploit the creative possibilities of the latent space (Tahiroğlu and Wyse, 2024).

## 2.4 Summary and research question

Gestural affordances in audio latent spaces remain underexplored, and existing practices of latent space navigation remain constrained in terms of how sound-producing gestures should be used. These motivate us to propose mapping out a complex relationship between gestural affordance, gestural perceptions, and the expressive potential of latent spaces. Therefore, we ask the research question:

**How do musicians perceive gestural affordances when navigating audio latent space and tailor them into performance techniques for musical expression?**

To approach this question, we used a DMI-as-research-probe approach (Hutchinson et al., 2003; Tahiroğlu et al., 2020). We created a DMI with two configurations, as described in the next section, for a user study workshop. We recruited 18 musicians to actively discover and test out open-ended gestural affordances, take notes, and create musical scores. We aim to analyze how perceived gestural affordances in a short-term engagement with the instrument can be adapted, repurposed, and discarded. We also aim to capture the formation and development of musicians' techniques, and analyze how the subjective perception of the sound-producing gestures and the sonic capacities of the latent space together contribute to this formation and development.

## 3 Musical instrument design

In light of the aim of our research, the design of our musical instrument probe aims to be minimal and flexible in terms of the use of gestures. And the interface's constraints should be open-ended so that musicians can tailor it to their needs. In addition, for the purpose of the user study, it should provide an easy way of tracing musicians' sound-producing gestures.

Given the above considerations, the idea emerged to use a stylus and tablet interface, in which one can use any drawing gestures with the stylus as input for latent space navigation. We chose line drawing to study gestural perception because: (i) line drawing with a stylus is a straightforward way of embodied interaction, and it can encode open-ended gestural movements for musicians to explore (Casey, 2018); (ii) previous work in musical timbre perception (Löbbers et al., 2023), graphic scores (Banar and Colton, 2022), and gestural sonic objects (Godøy, 2006) have shown that simple line and shape notations are able to express rich musical ideas; (iii)



we hope to contribute to broader research field of pen-based music control devices (Hinckley et al., 2014; Zheng et al., 2024a).

Section 3.1 presents our design principles to clarify assumptions and theoretical backgrounds that may have been brought into the design, and highlight some of the critical design decisions we made; Section 3.2 describes the configurations of our neural audio synthesis model and its training data; Section 3.3 presents how we designed our approach for latent space navigation; Section 3.4 describes the physical and software implementation of the instrument.

### 3.1 Design principles

The tablet interface captures the real-time spatial location of the stylus's pen tip on the canvas as a pair of  $(x, y)$  coordinates and the pressure  $p$  applied on the canvas, together as a 3-dimensional vector  $(x, y, p)$ . This poses the challenge of designing the mapping strategy to access the 8-dimensional latent vectors using  $(x, y, p)$ . We describe two mapping strategies used in Section 3.3. Another design decision we made was whether to map the speed of gestures as a parameter to control the latent space navigation. We decided not to consider this temporal dimension of the mapping to keep our intervention minimal, in this way, speeding up or slowing down the movement naturally results in faster or slower sound outputs. We suggest future works to explore the temporal dimension due to its importance in human-AI co-creation (Bryan-Kinns, 2024).

We consider the following two design principles when designing the dimensionality reduction strategy:

- **Balancing ambiguity and control:** As suggested by Françoise et al. (2022), the degree of clarity between sound-producing gestures and resulting sounds should be a sweet spot for performance techniques to develop. Therefore, the instrument's responses to gestures should give musicians freedom for openness and creative uses, while allowing for slow sculpting of the sound. Françoise et al. (2022) also mentioned this balance in their music improvisation with audio latent space.
- **Balancing surprise and repetition:** Unexpected results can be a positive factor in prompting musicians' exploration of an instrument (Kvifte, 2008). They can form gesture repertoire (Leman, 2012) that allows one to re-enact a gesture based on their mental simulation of both the action and its effect. This indicates how reproducibility can distinguish "pleasant surprise" from unexpected results that close down possibilities. However, the "exact" reproducibility of the sound is not always praised (Jordà, 2004). Therefore, finding a balance between surprising and repetitive is suggested.

### 3.2 Neural audio synthesis model

We used the Realtime Audio Variational autoEncoder (RAVE) (Caillon and Esling, 2021) as the encapsulated neural audio synthesis model given its ability to respond to real-time continuous

inputs (Françoise et al., 2022) with generated sample-based sound textures.

In order to follow our design principle of moderating the amount of surprise in the model's outputs, we created a customized guitar-plucking audio dataset with lower complexity in terms of timbre, playing techniques, and notes. We created a collection of guitar-picking sounds, mostly recorded dry on an acoustic guitar and an electric guitar, peaking around  $-6\text{dB}$ , played by the first author. The resulting dataset has a length of 2h after silences were trimmed. We used this customized dataset to train a RAVE using the "v2 causal<sup>1</sup>" configuration for 2M steps. The trained RAVE takes an 8-dimensional latent vector  $z$  as input, producing sequences of various guitar-picking sounds. The instrument has roughly 55ms latency between inputs and outputs, matching the results (52ms) reported in the literature (Caillon and Esling, 2022). We acknowledge Chris and Michael (2004)'s result on rhythmic inflections, which shows that delays of 14 ms and above tend to have a "slowing down" effect on rhythm perception.

### 3.3 Latent terrains

We designed a mapping model called *Latent Terrain* (hereby referred to as the terrain). A terrain is a set of one-to-one mapping between a given pair of coordinates  $(x, y)$  on the tablet's screen to an 8-dimensional latent vector  $z$ . Therefore, a terrain can be sampled by a closed interval, as a 2D plane of latent vectors tiled on each pixel location. When the stylus moves on the canvas, the terrain immediately retrieves a latent vector  $z$  corresponding to the stylus'  $(x, y)$ .

We used two algorithmic strategies to generate two terrains, respectively: a Variational AutoEncoder (VAE) (Kingma and Welling, 2013) and a Fourier-Compositional Pattern Producing Networks (Fourier-CPPN) (Tsfaldet et al., 2019). A terrain is fixed after it is generated. While the technical details and the procedure used for generating the two terrains can be found in the Supplementary material, here we visualize their differences in Figure 2 left and middle in the same way as Roma (2023): The width and height of the grayscale rectangles correspond to the width and height of the canvas. In each rectangle, the brightness at one spatial location represents the value of one latent vector dimension at that location. Given that our RAVE has 8 latent space dimensions, we can visualize a terrain by a stack of 8 grayscale rectangles. The functional difference between the two terrains is that the second terrain offers higher spectral complexity than the first one. That is, when the stylus moves for the same distance in the same amount of time on both terrains, the resulting snippet of sound from the second terrain is typically richer and more varied.

### 3.4 Hardware and software implementations

The instrument's physical interface, shown in Figure 3 left, is constructed from wood boards, embedded with a medium Wacom

<sup>1</sup> <https://github.com/acids-ircam/RAVE>

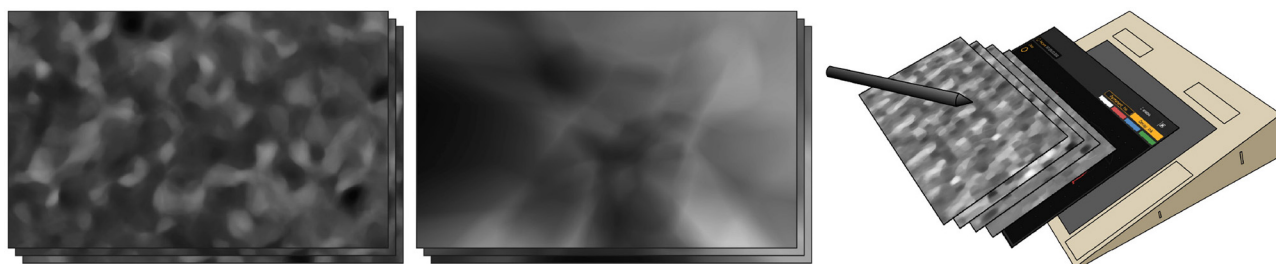


FIGURE 2

Visualization of three dimensions in the latent terrain 1 (left); Visualization of three dimensions in the latent terrain 2 (middle); Illustration of how a latent terrain is embedded in the tablet's canvas and accessed by the stylus (right).

Intuos<sup>2</sup> drawing tablet (21.6 × 13.5cm canvas size) and a Bela Trill Bar<sup>3</sup> slider. The slider bar next to the tablet was designed to be modular and can be quickly installed on either the right or left side to accommodate both left-handed and right-handed users. The drawing tablet connects to a laptop via USB. The slider runs on an ESP32<sup>4</sup> microcontroller powered by a portable charger, sending slider inputs to the instrument's software via the OpenSoundControl (OSC)<sup>5</sup> protocol. We implemented the latent terrain in C++ as a Cycling '74 Max 8<sup>6</sup> external: *nn\_terrain*,<sup>7</sup> facilitating stylus inputs and rendering drawings in the canvas. It encapsulates ACIDS-IRCAM's *nn\_tilde*<sup>8</sup> to load pre-trained RAVEs. As shown in Figure 2 right, the terrain is embedded in a part of the Graphical User Interface (GUI).

The GUI, as shown in Figure 3 right provides utility controls, including a “clear” button to clear drawings from the canvas, two buttons to switch between decayed ink (old drawings will fade out while new drawings appear) or permanent ink (drawings will never decay unless “clear” is clicked), and four buttons to switch ink colors. These utility controls are purely for annotation purposes to facilitate one to take notes and draw scores (see Section 4.2), and the modes of decay and ink colors do not affect how the sound is produced. The hardware slider's inputs control a pitch shifter to post-process the neural synthesis model's output.

To help readers understand the sonic capacities of the instrument, a video of the first author demonstrating the instrument is recorded.<sup>9</sup>

## 4 Study method

We held 6 recurrent 90-minute workshop sessions, each with 3 musicians. Our study was inspired by artist-led methods (Benford et al., 2013; Bryan-Kinns and Reed, 2023). In particular, musicians

lead the content creation on our instrument probe presented in Section 3, and we aim to study what the instrument affords them to do, and how they develop performance techniques on it. All workshop sessions took place in person in a performance room at Queen Mary University of London (QMUL). The study was approved by QMUL's Research Ethics Committee.<sup>10</sup>

### 4.1 Participants

Recruitments for participants were sent to postgraduate research groups interested in music and AI in London, the UK. A total number of 18 participants were recruited for the study. They were each reimbursed with a £20 (GBP) voucher. They were divided into 6 groups of three, each group allocated to one workshop session. Two participants opted out of the data analysis, 3 participants' data was incomplete due to late arrival, and 1 was incomplete due to a system crash. Therefore, the study collected 12 participants' data (6 Females, 5 Males, and 1 Non-Binary), and they all provided written consent for data collection, data analysis, and displaying their creations. Although we did not limit participants' handedness when distributing the call for participation, all participants were right-handed. Each participant was given an ID from P1 to P12.

To understand the participants' background in musical instruments, tablet and pen-based interfaces, AI-enhanced musical instruments, bodily experiences in musical instruments, and DMI design, we gathered self-report measures of their familiarity with these items. In a pre-task survey, participants were asked to rate on a 5-point Likert-type scale (Strongly Disagree - Disagree - Neutral - Agree - Strongly Agree) on six question statements listed in Table 1. The survey also gathered participants' self-reported primary instruments and years of practice.

Participants' responses to each question statement are described in the box plot shown in Figure 4, showing median (green dots), first and third quartile (edges of the box), extreme values (whiskers), outliers (outlined dots), and mean value (orange double-dot). Participants' primary instruments and years of practice are shown in Table 2. We interpret the results as low familiarity with

2 <https://www.wacom.com/en-gb/products/pen-tablets/wacom-intuos>

3 <https://learn.bela.io/products/trill/about-trill/>

4 <https://www.espressif.com/en/products/socs/esp32>

5 <https://opensoundcontrol.stanford.edu/>

6 <https://docs.cycling74.com/legacy/max8>

7 [https://github.com/jasper-zheng/nn\\_terrain](https://github.com/jasper-zheng/nn_terrain)

8 [https://github.com/acids-ircam/nn\\_tilde](https://github.com/acids-ircam/nn_tilde)

9 <https://bit.ly/latent-terrain-1>

10 Reference number: QMERC20.565.DSEEC524.068.

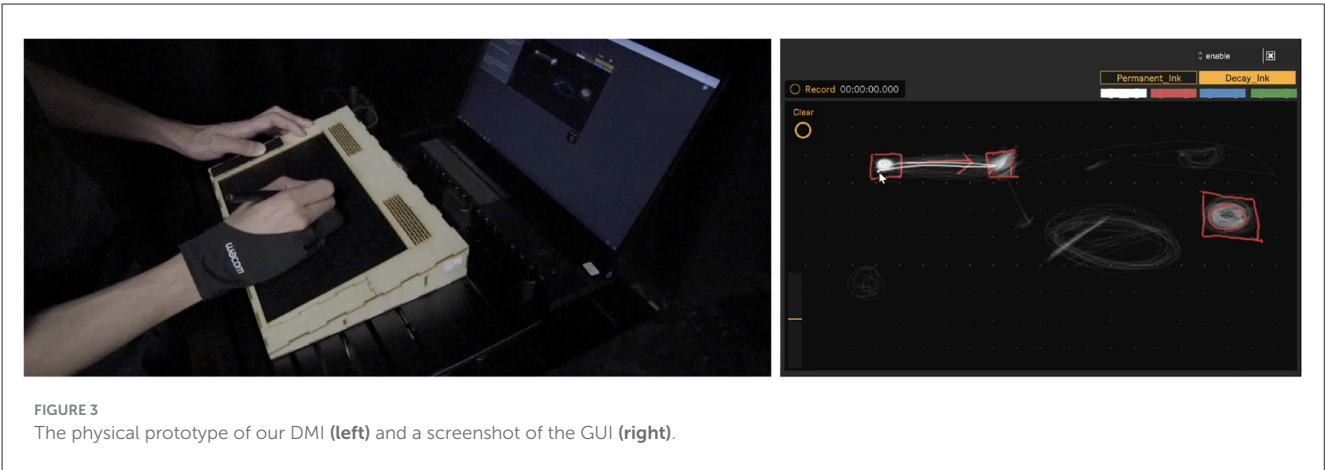
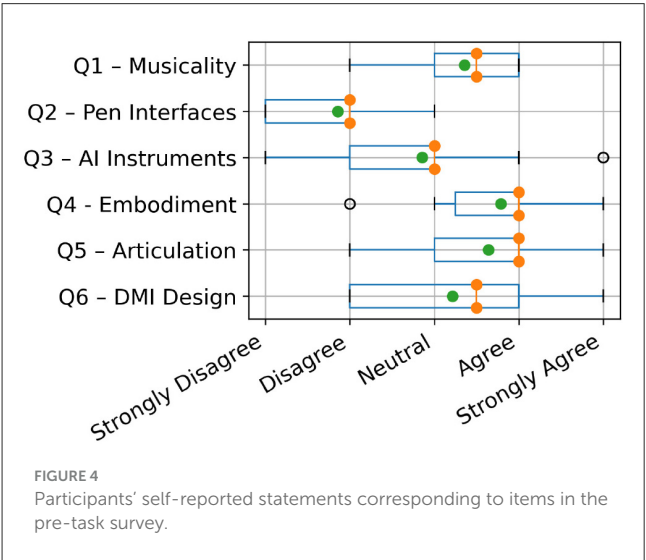


TABLE 1 Questions in the pre-task survey to gather participants' backgrounds.

| Question            | Question statement                                                                                    |
|---------------------|-------------------------------------------------------------------------------------------------------|
| Q1 – Musicality     | I would consider myself a musician.                                                                   |
| Q2 – Pen Interfaces | I am familiar with pen-based musical instruments, such as a tablet.                                   |
| Q3 – AI Instruments | I am familiar with AI-enhanced musical instruments, such as a neural synth.                           |
| Q4 – Embodiment     | I am able to identify what embodied (bodily-related) actions are involved when playing an instrument. |
| Q5 – Articulation   | I am able to talk about the bodily experiences of playing an instrument.                              |
| Q6 – DMI Design     | I am able to design digital musical instruments for others/myself to use.                             |



pen-based interfaces and AI-enhanced instruments, reasonable familiarity with musicality and ability to identify and articulate body-related experiences.

TABLE 2 Participants' self-reported background in musical instrument.

| ID  | Gender     | Primary instrument | Years of practice | Handedness |
|-----|------------|--------------------|-------------------|------------|
| P1  | Female     | Flute              | 5                 | Right      |
| P2  | Male       | Drum kit           | 12                | Right      |
| P3  | Female     | /                  |                   | Right      |
| P4  | Male       | Piano              | 7                 | Right      |
| P5  | Female     | Guitar             | 5                 | Right      |
| P6  | Non-binary | Piano              | 10                | Right      |
| P7  | Female     | Cello              | 20                | Right      |
| P8  | Female     | Hulusi *           | 4                 | Right      |
| P9  | Male       | DJ deck            | 13                | Right      |
| P10 | Male       | Guitar             | 22                | Right      |
| P11 | Female     | /                  |                   | Right      |
| P12 | Male       | Piano              | 10                | Right      |

\*Hulusi: a reed wind instrument also known as the gourd flute.  
/: None or "prefer not to say."

## 4.2 Procedure

The workshop was advertised as “Soundwalking Workshop with an AI Musical Instrument,” and we called the navigation of the latent space “soundwalking activities” as an analogy for freely exploring the spatial coordinates in a space (in our case, the latent terrains) while focusing one’s attention on bodily movements and sonic materials inside it, the same as [Scurto and Postel \(2023\)](#). We have clarified to the participants that this protocol differs from conventional soundwalking as physical walking ([Eckhardt, 2022](#)) to avoid false expectations.

In a workshop session, each participant wore a pair of headphones and stood in front of the tablet instrument and a laptop, both placed on a stand. The laptop displayed the workshop user interface (described in Section 4.3) in full-screen mode with no other user interface objects visible. Participants were required to wear a drawing glove on the pen-holding hand to prevent smudges and minimize friction from hand to tablet. Three participants stood



FIGURE 5

Workshop space from investigator's point of view (left); Participants drawing on the tablet interface (right).

side by side, facing the workshop investigator (the first author), as shown in Figure 5. Benches were provided for each participant to sit during cutscene between activities to minimize fatigue.

The following subsections describe the detailed workshop activities. Each activity has a GUI displaying relevant information, instructions, and the instrument's user interface. The full GUIs and a protocol script loosely followed by the facilitator can be found in the [Supplementary material](#). The investigator's laptop controls the switching between GUIs and switching between terrains in participants' instruments through OSC.

#### 4.2.1 Pre-task activity

**Pre-task survey:** Upon arrival, participants were asked to complete the pre-task survey mentioned in Section 4.1. The survey was embedded in the instrument's software user interface, and the participants were asked to fill it out using the same tablet and stylus they would be using for the rest of the workshop. We hope to use this pre-task survey not only as a method for data collection, but also as a warm-up activity to help participants familiarize themselves with the tablet interface and pen-drawing movements.

**Warm-up and introduction:** In a 3-min warm-up section, the investigator introduced and demonstrated a few line-drawing techniques to the participants as prompts, including hatching, contouring, stippling, and large and small scribbling. Then, the investigator introduced the workshop's aim and the instrument, and explained the upcoming activities.

**Body-scan:** Then, as suggested by other body-related and movement-related studies (Stahl et al., 2021; Tennent et al., 2021), the investigator led a 5-min "body scan," a closed-eyes sensitizing activity in which the participants stood still and focused their attention to their hands, wrists and arms, and then tested out various speeds, pressures, and intensities of sketching techniques.

#### 4.2.2 Exploration activity

Participants spent 20-min exploring and experiencing two latent terrains in the instrument. The first terrain (generated by the VAE) and the second terrain (generated by the Fourier-CPPN) were presented to the participants at random orders. The user interface for this section is shown in Figure 6. Participants were instructed

to (i) test out different line-drawing techniques on different spatial locations on the canvas, (ii) fill out an entry in the note template (see Section 4.3.1) for each technique they used to document their in-the-moment experience, and (iii) complete around 5 note entries for each terrain. We embedded the note template into the interface instead of using a printed copy to minimize disruption caused by note-taking and keep participants in their flow while exploring. We encouraged participants to explore various expressions, including gestural techniques, speed, intensity, and pressure. Visualizations of the terrains (see Section 3.3) are not revealed until the end of the workshop to prevent participants from relying on visual cues. After the exploration section, we conducted the group interview described in Section 4.3.2.

#### 4.2.3 Score creation and demonstration activity

Then, participants spent 20-min exploring and finding pieces of sound or music in both terrains, each for 10-min, and drew a "score" on the canvas to help them perform these pieces at a later time. They were encouraged to be creative and use any kind of notation (graphical/textual) in the score. We did not limit the duration of the pieces.

Participants who completed at least one score were asked to demonstrate their creations to the investigator and the other two participants. All demonstrations were limited to 8 minutes. Other participants were free to give comments or discuss with each other.

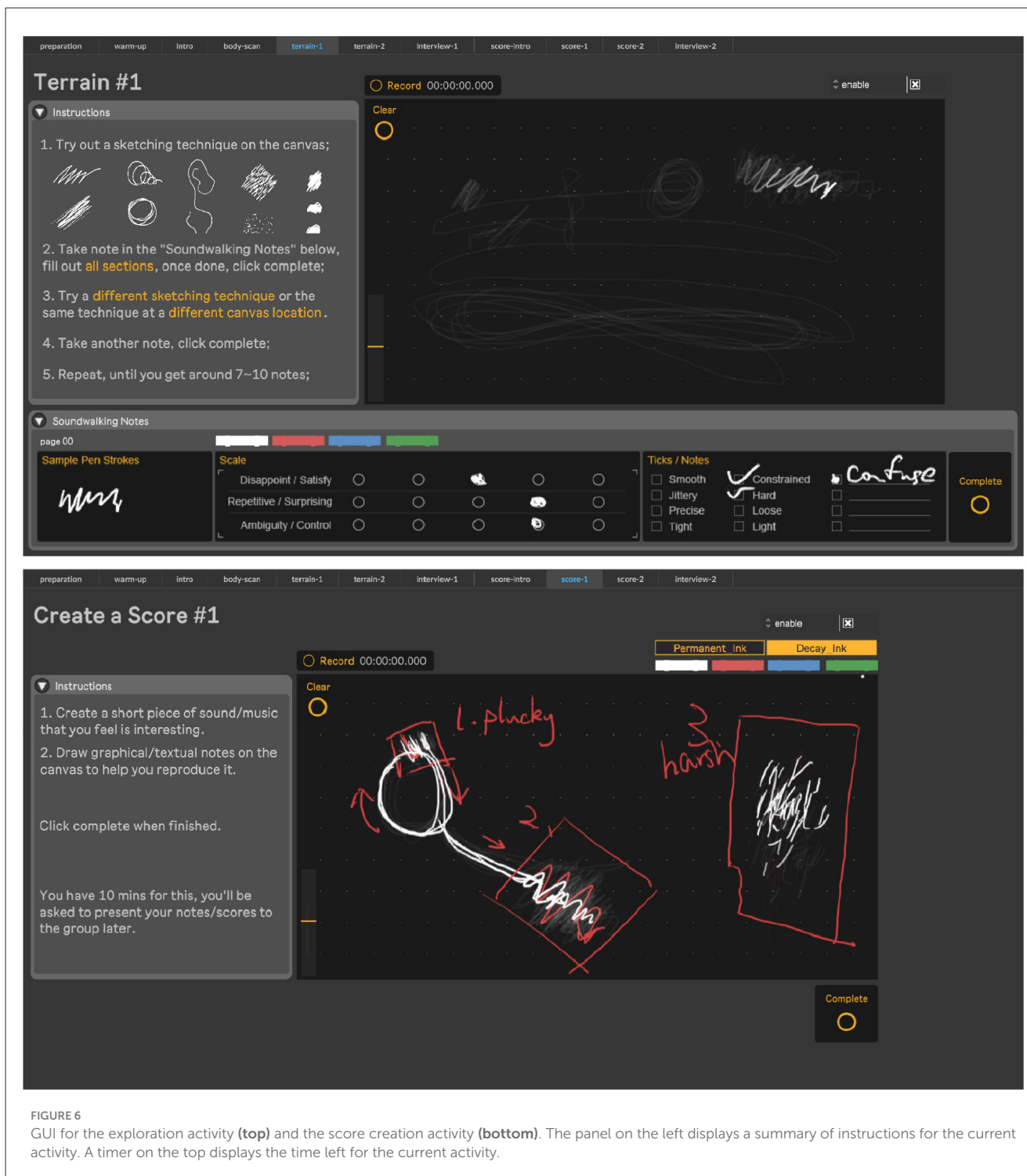
### 4.3 Data collection

In the activities above, we used a mixed data collection approach including interviews, demonstrations, and participants' documentary notes. It aims to capture how participants interact with the instrument and their in-the-moment experience, as described below.

#### 4.3.1 Documentary notes

The exploration activity described in Section 4.2.2 used a documentary note template that we developed, inspired by the





soma trajectories tool (Tennent et al., 2021) and body map (Anne Cochrane et al., 2022). We did not limit how this note template should be used by the participants, and we encouraged creative ways of articulation. Each note entry, shown at the bottom of the interface in the first screenshot in Figure 6, has three elements to fill out:

- The left column has a space to fill out a sample of the drawing technique.
- The middle column has a 5-point Likert-scale to self-report one's experience regarding the design principles *reproducibility* we described in Section 3.1, and an additional *satisfaction* for overall experience. Following the suggestion by Weijters et al. (2010) that a numbered scale with fewer categories helps respondents to orientate themselves more easily, we replaced the slider scale in the soma trajectories tool (Tennent et al., 2021) with the five-category Likert scale. To clarify, the purpose of these ratings is not to rigorously

measure participants' experience in a quantitative manner. Instead, they were used as cues for participants to recall their "interaction trajectory" (Benford et al., 2009) later during the interview section.

- The right column has a space to fill out one's bodily experience related to hand mobility and movements. We adapted the convention in body map (Núñez-Pacheco, 2021) of providing vocabularies to help users of this template articulate their bodily experiences. They could either tick on existing ones or add their own notes.

### 4.3.2 Group interviews and demonstrations

The group interview after the exploration activity aims to gather participants' first-person descriptions of their subjective experiences while they were exploring the instrument. The investigator guided each participant in reviewing all entries of their documentary notes. For each entry, participants were asked to (i) describe the technique they used and the sound it produced, (ii) explain each rating responding to the design principles, and (iii) describe the bodily experience while performing this technique. Participants were free to use the instrument to demonstrate their findings.

The demonstration after the score creation activity aims to gather participants' explanations of what they intended to do and what techniques they ended up using. Participant was asked to describe: (i) what they have created, (ii) how they would use their scores, and (iii) how they came up with their scores. After their demonstration, the investigator started open-ended conversations to elicit descriptive narratives of why they chose to use the techniques in their score and their overall experience of the instrument.

Both the group interviews and the demonstrations were video and audio recorded. The video recording focused only on participants' hands and wrist areas and the laptop's screen. All scores created by participants were collected in image format.

## 4.4 Data analysis

We gathered participants' documentary notes, scores, video recordings of interviews and demonstrations. We took a narrative analysis approach (Sparkes and Smith, 2008) to compile participants' exploration of gestures and their development of techniques into longitudinal descriptions. This was inspired by previous works in HCI and artist-led research on DMI design (Sturdee et al., 2021; Saitis et al., 2024) to interpret observational data from mixed methods in a systematic way centering around a research question. In practice, instead of transcribing the recordings, the first author reviewed video recordings and wrote a narrative in third-person view for each participant. Each narrative includes (i) paraphrases of participants' verbal words that are relevant to our research questions, (ii) direct quotes when necessary, (iii) text descriptions of participants' hand movements as they demonstrate, and (iv) vignettes of observations (Bryan-Kinns and Reed, 2023) extracted from the video recordings. To ensure our interpretation is neutral and aligned with the standard ways

of reporting sketching gestures, we studied literature on sketching techniques (Lohan, 2012, p. 15–25) and stylus-tablet interfaces (Hinckley et al., 2014).

We also aim to use this narrating process to clarify and refine any unclear self-reported descriptions of drawing gestures and pen strokes. As described in Section 4.1, participants typically have low familiarity with pen-based interfaces and, therefore, are likely to have limited vocabulary in articulating relevant gestural movements such as hatching and scribbling. Therefore, as we encouraged our participants to demonstrate their movements in action as they were being interviewed, we were able to refine participants' descriptions of their movements according to the video recording.

We coded and categorized sound-producing gestures based on participants' notes and scores, with reference from the literature on sketching techniques (Lohan, 2012, p. 15–25) and gestural sonic objects (Godøy, 2006). The purpose is to showcase how participants' choice of gestures progresses from their initially perceived affordances to the final set of techniques used on their scores. This is inspired by related works on embodied cognition that capture techniques adaptations in musical instruments (Mice and McPherson, 2022).

Then we gathered all the narratives and performed a thematic analysis (Clarke and Braun, 2017). The analysis involved (i) reading through the data and identifying segments that connect to our research question, (ii) generating initial codes, (iii) organizing the codes into themes, and (iv) iteratively reviewing and refining themes. Themes emerged from iterative analysis of the data, with a focus on our overarching research question, and inspiration from literature on movement-sound interaction (Françoise et al., 2022) and body motion in DMI (Jensenius, 2022).

## 5 Results

Twelve participants completed the study and opted for data analysis. All 12 participants who completed the exploration activities were interviewed. Ten out of the 12 participants [P1–P5, P7–P11] completed scores for both terrains and demonstrated their creations. We gathered and analyzed 138-min video recording in total including the group interviews and the demonstrations. Screen and audio recordings of participants' exploration, creation and demonstration of their scores are synced and displayed at our interactive web repository.<sup>11</sup> This section reports the progression of participants' choice of sound-producing gestures, creation and demonstration of musical scores, and themes emerged around participants' subjective perception when navigating the latent space.

### 5.1 Progression of sound-producing gestures

In Table 3, we list drawing strokes resulting from participants' sound-producing gestures. We distilled this list from participants' notes and scores based on our visual observation. And we

<sup>11</sup> [https://jasper-zheng.github.io/stylus\\_analyser/](https://jasper-zheng.github.io/stylus_analyser/)

documented and categorized them with reference from the literature on pen sketching techniques (Lohan, 2012, p. 15–25). We observed six categories: impulsive/iterative line strokes, M-strokes, scribbles, stipples, loops, and outlines. We divided the size of the drawing strokes into short/small, medium, and long/large. Figure 7 illustrates these categories and divisions.

We use this collection to show how participants' choice of techniques changed between the exploration and score creation activities. Two main inventive techniques that the investigator did not initially demonstrate to the participants but used by them were: (i) "M" shaped strokes, a technique similar to hatched line strokes but have larger distances between each line, typical examples are the zig-zag lines that appeared in P8's score on Terrain 2; (ii) wavy outlines, a technique similar to curved outlines but more structured in terms of the shapes. Compared to the techniques we initially demonstrated, both inventive techniques focus more on the drawing strokes' shape in the visual aspect instead of the pen-stroking movements in the gestural aspect. Especially the wavy outlines from P8, which they described as "quite meditative to draw." In addition, P7 and P10 opted to limit the size of their movement to small areas because they felt that the sound produced by smaller shapes was consistent every time they drew them, while the sound produced by larger shapes seemed to be random. Notably, all techniques used by P10 in their score were either short or small strokes while they experimented with medium size shapes and outlines.

We documented and grouped techniques according to our visual observation of the resulting drawing strokes. However, we found that the visual appearance of drawing strokes cannot comprehensively represent the diverse movements participants have used. Various expressions were used in the same type of drawing stroke. For instance, P2 and P4 both used curved line strokes, but P2's movements tended to be careful and steady, while P4's tended to be fast and scrawled. Therefore, we use the techniques collected here to complement the individual demonstration data in Section 5.2, which describes the diverse and nuanced differences in how participants enact their techniques.

## 5.2 Creations and demonstrations of musical scores

Here we present participants' diverse approaches to creating and demonstrating the musical scores. We summarize our results from the aspects of postures and pen grips (Section 5.2.1), ways of notating the scores (Section 5.2.2), and techniques for musical expressions in the scores (Section 5.2.3). Figure 8 shows the scores from P1 to P5, Figure 9 shows the scores from P7 to P11. The summary presented here has been edited to reduce the length and center around our research question. We invite readers to trace the full third-person narratives in the [Supplementary material](#) for more in-depth descriptions.

### 5.2.1 Postures and pen grips

Participants used various postures and pen-gripping styles to engage their hands, wrists, and arms with the instrument in their

demonstration, as shown in Figure 10. We observed that this variation of postures and pen grips is typically associated with how they use sound-producing gestures in their scores.

For instance, P1 tended to hold the pen's upper part, keep their fingertips far away from the tablet's surface, and avoid contact between their hand and the surface. They explained that this is not how they usually hold a pen, but they felt comfortable using it because it allows them to "[use the pen to] visit the entire canvas easily." They demonstrate this "visit the entire canvas" action by fixing their forearm and rotating their wrist in small circles while the pen's tip softly path through the four corners of the canvas. As a result, P1's score for the second terrain is composed of various circular movements and has a lower level of pressure applied to the tablet's surface.

Postures and pen grips could result from how a participant intended to use gestures. For instance, P4 held the very top part of the pen in a posture similar to holding a baton. When questioned about this posture, they explained that they intentionally held the pen this way to use it as a baton for gestural control, and distinguish playing on the instrument from the regular handwriting. A relevant observation is that P4 tended to use fast and scrawled techniques, with a focus on maintaining the overall gestures' shape and direction. In contrast, P7 held their pen very close to the pen's tip and tended to lift their elbow while moving their forearm. This helped them to create nuanced and precise articulations of the drawing strokes' length.

Postures and pen grips could also affect how a participant ended up using gestures. For instance, P2 pressed and fixed their wrist firmly against the tablet and constrained their movement to wrist rotation, in particular, rotating their hand toward or away from the side of their thumb to navigate the pen on the canvas, shown in Figure 10. As a result of this constrained posture, P2 tended to deploy repetitive line-hatching gestures over the same region on the canvas, and search for other aspects, such as rhythm and intensity of the gestures, to vary the sound.

### 5.2.2 Notations

Participants approached the creation of scores in various ways. Most participants (except P4 and P11) used the scores as a "map" to mark down spatial regions in the canvas as notations of where they should enact their sound-producing gestures. P2, P3, P9, and P10 specified a technique for each region. For instance, in P2's score on the second terrain, they casually traced horizontal lines from bottom to top in the box labeled "1." In contrast to the "map" approach, P4 and P11 tended to follow the shapes and patterns of lines in their score without specifying where to perform them. P4 assigned a vertical area on the right side of the canvas, in which they notated sample gestures. And they reproduced these sample gestures in a sequence on a larger scale, covering almost the entire canvas.

The trajectories of gestures played an important role in participants' notations of the scores. P1, P5, and P8 tended to have a strong intention of following the trajectories of lines they had drawn. For instance, P1's score for the first terrain begins by carefully and softly tracing the long outlines on the upper-left area of the canvas from bottom-left to top-right. Their score

TABLE 3 Techniques for sound-producing gestures used by participants in the exploration and score creation activities.

| Techniques                           | P1 | P2 | P3 | P4 | P5 | P6* | P7 | P8 | P9 | P10 | P11 | P12* |
|--------------------------------------|----|----|----|----|----|-----|----|----|----|-----|-----|------|
| Short impulsive line strokes         | ⊗  |    | ○  | ○  |    |     | ⊗  | ⊗  |    | ○   |     |      |
| Medium impulsive line strokes        |    | ○  |    |    |    |     | ○  |    |    |     |     |      |
| <b>Short iterative line strokes</b>  | ⊗  |    | ×  |    |    | ×   |    |    | ○  | ○   | ⊗   |      |
| <b>Medium iterative line strokes</b> | ○  | ×  | ⊗  |    | ×  | ×   | ×  | ×  | ○  | ×   | ⊗   | ×    |
| Long iterative line strokes          |    | ⊗  |    |    |    |     |    |    |    |     |     |      |
| Short iterative M-strokes            | ⊗  |    |    | ○  |    |     | ○  | ○  | ○  |     | ○   |      |
| Medium iterative M-strokes           |    |    |    |    | ×  |     |    | ○  |    |     | ○   |      |
| <b>Small scribbles</b>               | ×  |    | ○  |    | ⊗  | ×   |    | ⊗  |    | ⊗   | ○   | ×    |
| Medium scribbles                     |    | ×  | ○  | ○  |    |     | ○  | ○  | ×  |     | ○   | ×    |
| <b>Stippled lines</b>                |    |    | ⊗  | ○  | ⊗  | ×   |    | ⊗  | ⊗  | ⊗   | ×   | ×    |
| <b>Stippled dots</b>                 | ⊗  | ⊗  |    |    |    |     |    |    | ○  | ×   | ⊗   |      |
| Curve loops                          | ○  | ×  |    |    |    |     | ⊗  |    |    |     | ○   |      |
| <b>Small spiral loops</b>            | ⊗  |    | ○  |    |    |     |    | ×  |    | ×   | ○   |      |
| <b>Medium spiral loops</b>           | ○  | ⊗  | ⊗  | ○  |    | ×   | ○  | ×  |    |     | ⊗   | ×    |
| Large spiral loops                   | ○  |    |    | ○  | ⊗  | ×   | ×  |    | ⊗  |     |     |      |
| Wavy outlines                        | ⊗  | ×  | ⊗  |    | ⊗  |     | ⊗  | ○  | ⊗  |     | ○   |      |
| <b>Curved outlines</b>               | ×  | ⊗  | ⊗  | ○  |    | ×   | ⊗  | ×  |    | ×   | ⊗   | ×    |
| Jagged outlines                      | ×  | ×  |    |    | ×  |     |    |    |    |     |     |      |
| Straight outlines                    | ×  |    | ×  | ○  |    |     |    |    | ×  | ×   | ○   |      |

×: Used this technique in the exploration session only.  
○: Used this technique in their score (except for annotation purposes).  
⊗: Used this technique in both the exploration session and their score.  
**Techniques in bold text** were originally demonstrated by the investigator to the participants.  
**Participants with asterisk (\*) ID** have not completed both scores.  
Six cells with **gray backgrounds** are to be reflected and discussed more in-depth in Section 6.2.1.

for the second terrain started by tracing the largest white circles in the middle and then moved to the adjacent smaller circles in an arbitrary order to maintain each circle’s shape. In P7’s score for the first terrain, they employed a gradual change from strictly following the lines to loosely following the lines. They traced the white color pattern using the line hatching technique, with precise and intentional control of the length of each line to maintain the pattern’s shape, and gradually gave up this intention and changed to carelessly hatching wavy lines that loosely followed the pattern.

Textual notations have been used for various purposes. P2 and P4 wrote texts to describe how a particular gesture should be enacted. P10 annotated text to describe the sound produced by specific regions they defined. P2, P5, P8, P9, P10, and P11 used numeric indexes to sequence the gestures in their scores.

5.2.3 Techniques for musical expressions

We observed three notable techniques to create musical expressions in the scores. These techniques explored various aspects of the latent space’s sonic capabilities and the articulation of gestures.

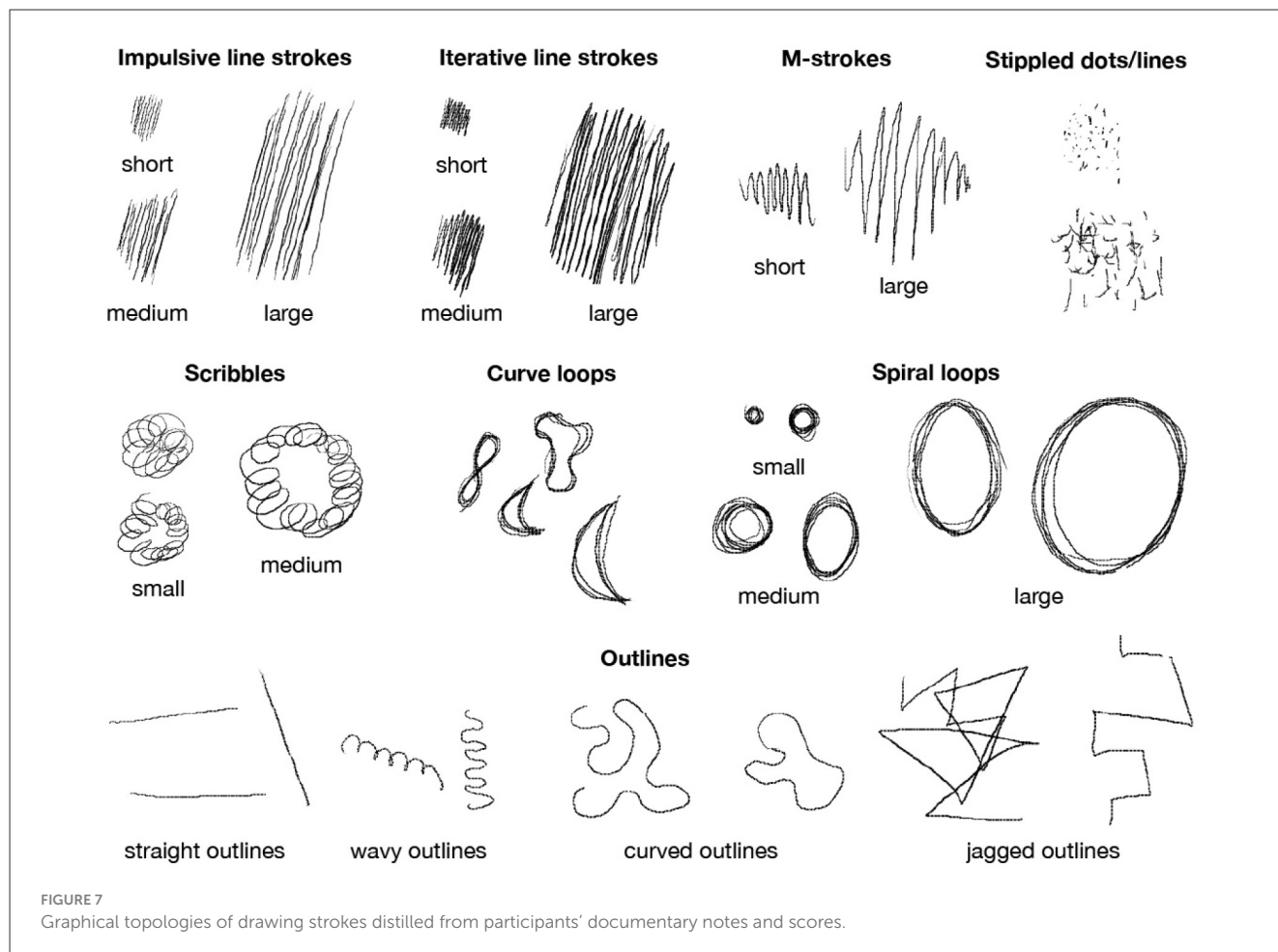
P2 described the red box labeled “1” in their second score as a “space that produces a variation of plucking sounds,” in which their

movements are similar to using the pen to pluck a piece of string—a soft pen-down movement to enter the canvas, quickly adding pressure to it and a sudden lift to exit the canvas. This gesture was performed typically in a rhythmic way, resulting in sequences of guitar-picking sounds with the same pitch. Another example is in P3’s score on the second terrain, where they placed short straight lines in steady rhythm.

P4’s score for the first terrain has a white outline section, in which they started by quickly contouring a horizontal line that path through the canvas from left to right, then abruptly stopped and restricted their movements to the small scribbling technique in a jittery and trembling way. This sequence of gestures first results in a slightly random fragment of audio when pathing through the canvas, followed by a tremolo sound with sustained timbre when the pathing stops.

P10 attempted to create contrasts between various sounds discovered in the latent space. They explained the labeled text “basic” in their score for the first terrain as a similar idea to the root of a chord, which they chose to start from, and then went to other labels such as “harsh” and “scrab” to create tension, and then return to the root by returning to “basic” or “soft.” When demonstrating techniques in boxes labeled “harsh” and “scrab,” they used a burst of scribbling





that starts with casual and chaotic gestures and immediately converges to a point, resulting in drawing strokes with tornado shapes. Then they moved on to the boxes labeled “soft” and “calm.” Their gestures were softer and less aggressive in these two boxes.

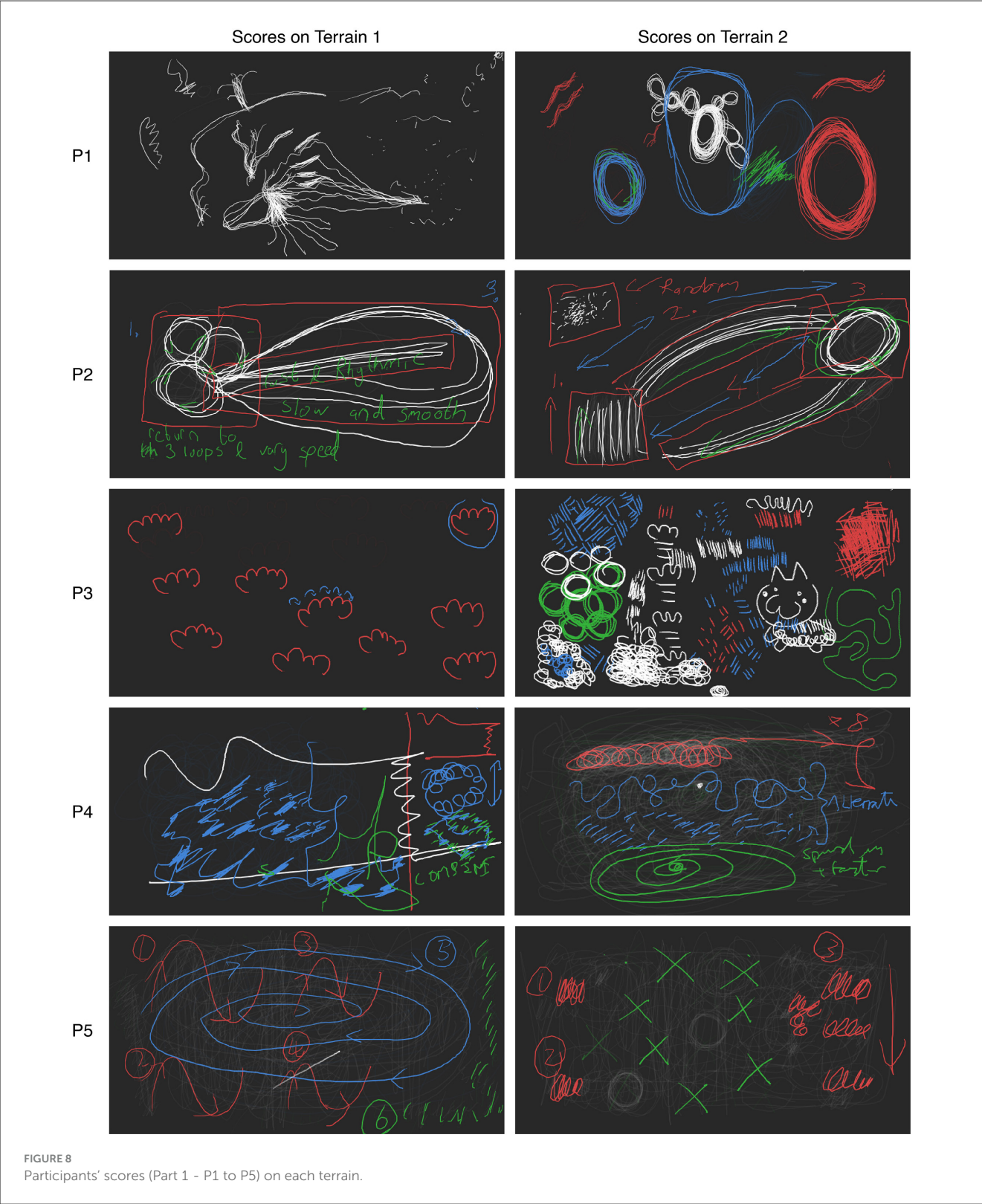
### 5.3 Subjective perception of latent space navigation

Codes from our thematic analysis were organized into six themes around how subjective perception of latent space navigation contributes to the formation and development of techniques: T1 - Movements, Postures, and Attending to the Body; T2 - Repetition and Persistence; T3 - Techniques and Repertoires; T4 - Action-Sound Mapping; T5 - Timbre and Sound Characteristics; and T6 - Visual Cues and Visual Interpretation. The order of the themes is based on our interpretation of how close the theme is to the embodied experiences of participants. The codebook in Table 4 shows the themes, codes contributing to each theme, the number of times each participant commented on each code, and the overall counts of each theme and code. Detailed definitions of each theme can be found in the Supplementary material.

#### 5.3.1 T1: movements, postures, and attending to the body

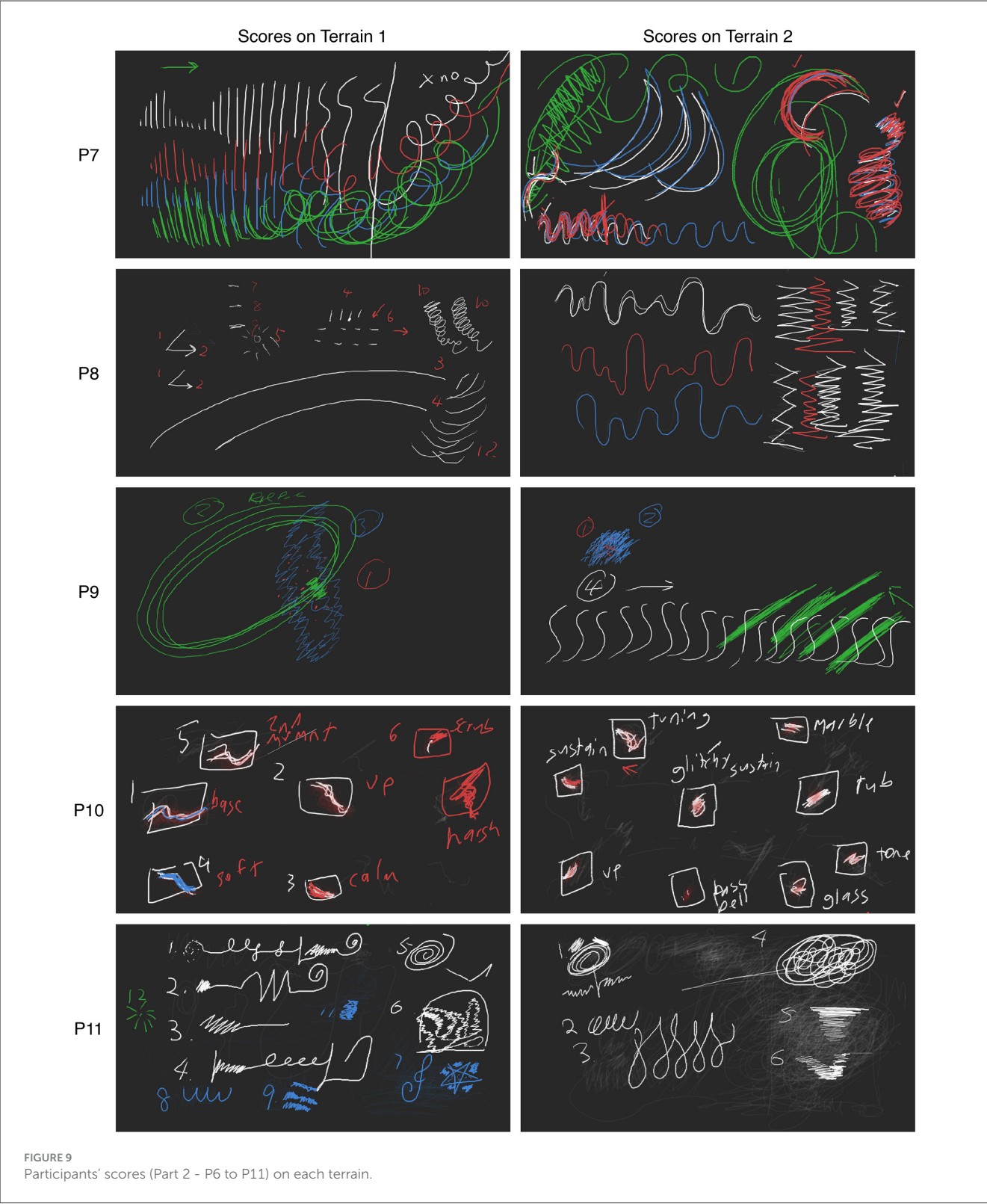
Eleven out of the 12 interviewees described their experience in relation to the movements of their sound-producing gestures. Aspects of descriptions include movements' size, speed, and intensity. *Size* describes how far one's pen will travel on the canvas in one sketching action, while *speed* defines how fast it travels. For instance, P1 tended to scale up the size of their movements to “visit the entire canvas,” while P9 managed to keep the speed of their movement consistent while sketching ellipses with different sizes. *Intensity* describes how much pressure one may apply to the canvas while navigating.

How a participant uses size, speed, and intensity of movements can relate to the posture of their hand, wrist, and arm, or their pen gripping style. For instance, we described the “holding upper part” pen grip of P1 and the “baton-like” pen grip of P4 described in Section 5.2.1. By incorporating these postures and pen-gripping styles, P1 and P4 engaged their entire arm for gestures while performing, and this typically results in long and sweeping strokes that fly across the canvas. Similarly, P5 experienced a change of technique from large-size hatching to short-line hatching when demonstrating their score. By hovering their forearm on the canvas and moving the entire forearm for large hatching gestures, P5 described this technique as “free



and loose.” In contrast, by pressing their wrist against the tablet and slightly flexing their fingers to move the pen around, P5 described this technique as “similar to a regular hand-writing posture so that it [the drawing] naturally feels precise and accurate.”

Eight out of the 12 interviewees were able to articulate their body experiences using either vocabulary from the note template or words of their own. For instance, P5 described their hand experience as “light and peaceful” while scribbling circles using wrist and finger movements, and “hard and



angry” while stippling lines using wrist movements. In addition, P3 commented that the body experiences in different hand movements have affected their choice of techniques. They indicated that performing line hatching in the top-right to bottom-left direction feels “natural” and “easy to control,” whereas the

top-left to bottom-right direction feels “disrupting” and “in a wrong direction.”

Seven out of the 12 interviewees indicated that at a particular moment, they paid active attention to their hand movements, and consciously thought about how and where to navigate next.



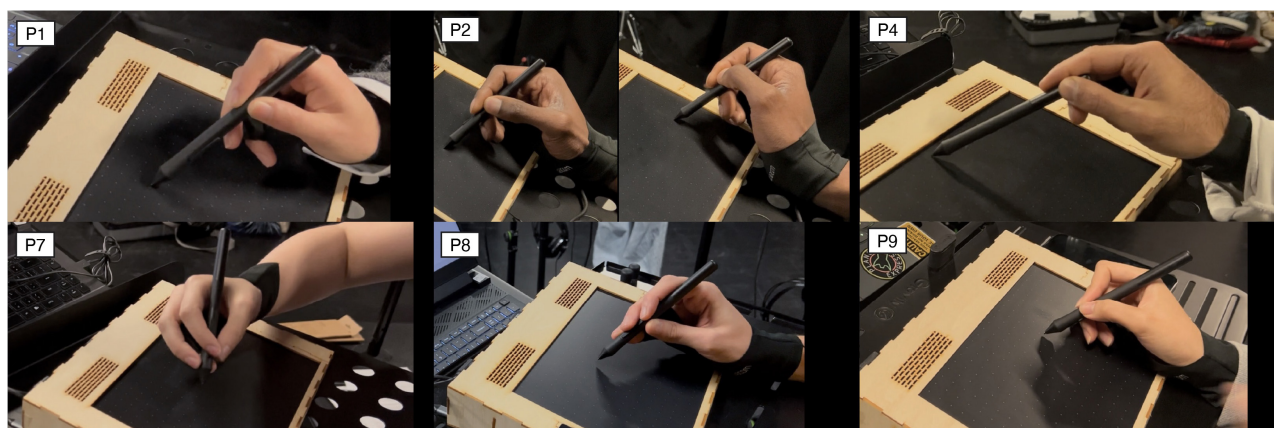


FIGURE 10

Frames extracted from videos recorded while participants were demonstrating their scores, showing various hand, wrist, and arm postures and pen-gripping styles.

For instance, P6 indicated that “[they] intentionally kept [their] movements smooth and round” when they were drawing a long, sustain, and random direction curved outline. This intention has let them feel that their movements were “constrained by their thoughts,” which is in contrast to the feeling of “relaxed and free” when they rely entirely on their instincts.

### 5.3.2 T2: repetition and persistence

Seven out of the 12 interviewees have attempted to perform particular patterns repetitively to learn and familiarize them. For instance, P8 and P11 explored drawing in loops of circles on various canvas locations in the exploration activity, P3 explored hatching short straight lines back and forth and gradually moving to the lines’ perpendicular direction, and P2 incorporated repeating patterns in both scores. We named this theme with the inspiration from [Jensenius \(2022\)](#) as they used “persistence” to describe a learning and adaptation process in performance.

P6, P7, P10, and P12 expressed the idea of finding a particular sound snippet they enjoyed and repeating the same movement on the same canvas locations in the hope of reproducing the sound. For instance, P6 attempted to memorize the trajectories of sketches and trace them repetitively. In the case of P10, who used this technique across both their scores, they described this process as “collect interesting spots on the canvas, then arrange and annotate them to perform repeated patterns on these spots.” P7 described this process as “digging around and finding points of interest, zoom-in on these points to inspect the sequences of melodies, and then retracing the sketches that produce these sequences.”

Some participants familiarized themselves with the instrument by repeating and learning a pattern. For instance, P10 commented that they explored hatching short and straight lines back and forth twice—the first attempt was the first sketching technique they tried on the instrument, and the second attempt was near the end of the exploration. During the second attempt, they were able to slow down their hand motions and limit the drawing to fixed-length short lines. They commented that this familiarity with the

movements made them feel “stabilized.” Similarly, P1 explained that after repeating the same pattern for a while, they could “soak into repetitive movements and move their full attention on the sound.” They also described this learning process as “forget about their hand movements and feel relaxed.”

### 5.3.3 T3: techniques and repertoires

In addition to Section 5.1, in which we described the drawing strokes resulting from sound-producing techniques, in this theme we summarize two high-level techniques participants used as repertoires for expressions.

First, 4 out of the 12 participants attempted to find rhythmic patterns in sketching movements. For instance, P2 commented “it was interesting to sketch lines between two spaces in a rhythmic fashion,” and they ended up identifying two spatial areas on the canvas and quickly hatching long parallel lines that connect these two areas in a rhythmic pattern. P9 also focused on the rhythmic pulse because “[their] first impression was that it feels like a percussive instrument.”

Second, 8 out of the 12 participants used contrast in movements, sounds, or feelings for musical expressions. For instance, P3 “consciously distinguish between fast and slow motions” as they hatch short straight lines. Similarly, P4, P5, and P9’s demonstrations all employed a sudden change in speed as a moment of tension in their scores. As described in Section 5.2.3, P4’s score for the second terrain has a moment in which they suddenly pause and freeze their movement to create a feeling of suspense. The idea of contrast has been used by P10 across their demonstration, they were also able to use the size of their movement to create contrast, as described in Section 5.2.3, the tornado-shaped drawing in which they quickly shrink their movement in one cycle.

### 5.3.4 T4: action-sound mapping

Ten participants indicated that their ratings on the overall satisfaction in the note template related to whether they can



**TABLE 4** Thematic analysis codebook showing themes, codes contributing to each theme, the number of times of each participant commented on each code.

| Themes (bold), Codes (regular)                            | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | Count     |
|-----------------------------------------------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----------|
| <b>T1: Movements, postures, and attending to the body</b> |    |    |    |    |    |    |    |    |    |     |     |     | <b>45</b> |
| Attention and conscious efforts                           | 1  |    | 1  |    |    | 2  | 1  | 2  | 2  |     | 2   |     | 11        |
| Feeling of movements                                      | 1  |    |    |    | 1  | 1  | 1  | 2  | 1  | 2   | 1   |     | 10        |
| Posture and gripping variability                          | 1  | 2  | 2  | 1  | 1  |    | 1  |    |    | 1   |     |     | 9         |
| Size, speed, and intensity of movements                   | 1  | 2  |    | 1  | 2  | 1  | 1  | 2  | 2  | 1   | 1   | 1   | 15        |
| <b>T2: Repetition and persistence</b>                     |    |    |    |    |    |    |    |    |    |     |     |     | <b>25</b> |
| Discover and reproduce points of interest                 | 1  | 1  |    | 1  | 1  | 1  | 2  |    | 2  | 2   | 1   | 1   | 13        |
| Familiarity of the instrument                             | 2  | 1  |    |    |    |    |    |    |    | 1   |     |     | 4         |
| Repeat particular patterns                                | 1  | 1  | 1  |    |    |    | 1  | 1  | 1  | 1   | 1   |     | 8         |
| <b>T3: Techniques and repertoires</b>                     |    |    |    |    |    |    |    |    |    |     |     |     | <b>26</b> |
| Combine multiple techniques                               | 2  |    |    | 1  |    |    |    |    | 1  |     | 1   | 1   | 6         |
| Create rhythmic patterns                                  |    | 2  |    |    |    |    |    | 1  | 2  |     | 1   |     | 6         |
| Expressions in movements and sounds                       | 1  | 1  | 1  | 2  | 2  |    |    |    | 1  | 2   |     | 1   | 11        |
| Variability in gesture and drawing                        |    |    |    | 1  | 1  |    |    |    |    | 1   |     |     | 3         |
| <b>T4: Action-sound mapping</b>                           |    |    |    |    |    |    |    |    |    |     |     |     | <b>27</b> |
| Approaches for dealing with ambiguity                     | 1  | 2  |    |    | 1  |    | 1  |    | 1  | 1   |     |     | 7         |
| Degree of clarity                                         | 1  | 1  | 1  | 2  | 1  |    | 1  |    | 1  | 1   | 1   | 1   | 11        |
| Sensory alignment                                         | 1  | 1  | 1  |    | 1  |    | 2  | 1  |    |     | 1   |     | 8         |
| Sound produced by physical interface                      |    |    | 1  |    |    |    |    |    |    |     |     |     | 1         |
| <b>T5: Timbre and sound characteristics</b>               |    |    |    |    |    |    |    |    |    |     |     |     | <b>21</b> |
| Create timbral variations                                 | 1  | 1  |    | 1  | 1  |    |    |    | 1  |     |     |     | 5         |
| Inspect sound capacity                                    |    | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1   |     | 1   | 10        |
| Spectral complexity of the sound piece                    |    | 2  |    |    |    | 1  |    |    | 1  | 1   |     | 1   | 6         |
| <b>T6: Visual cues and visual interpretation</b>          |    |    |    |    |    |    |    |    |    |     |     |     | <b>22</b> |
| Action guided by visual cues                              | 1  | 1  | 1  | 1  | 1  |    | 1  | 2  | 2  |     |     | 1   | 11        |
| Figurative understanding of drawings                      | 1  |    | 1  |    | 1  |    | 1  |    |    |     | 1   |     | 5         |
| Methods for annotating                                    |    | 1  |    | 1  | 1  |    |    | 1  |    |     | 1   | 1   | 6         |

gain control over the sound using their movements, specifically, the instrument's action-sound mapping (Jensenius, 2022) design. For instance, three participants were expecting readable one-to-one mapping between parameters of a gesture's spatial movement such as speed [P3, P7], intensity [P10], and location [P7]. As a result, six participants attempted to develop methods for dealing with ambiguity when they felt the sound was hard to control. For instance, P1 reduced the amount of improvisation in the second terrain because "the sounds come from improvisation often seem too random." P2 and P9 discovered that slowing down and stabilizing their hand was helpful in inspecting how the sound related to their hand movements.

We identified that seven participants' comments are related to sensory alignment (Marshall et al., 2019) between sound and their kinesthetic senses. Participants typically would highlight a sound when the movement that triggers this sound feels similar to how the sound should be produced in the real world. For

instance, P3 highlighted the hatching technique because the movement of quickly placing short lines "feels like strumming a guitar with a pick," while the resulting sound was also similar to plucking a string. P2's demonstration of their score for the second terrain also includes this idea. Similarly, P5's score for the first terrain has a section they described as "mimicking someone's footstep," they explained that they created this section because they "found that the sounds in this area feel wooden and noisy, and if I tap [stipple] on it, it sounds like stepping on a wooden floor." We gathered participants' descriptions about aligned kinesthetic senses and perceived or expected sounds in Table 5.

### 5.3.5 T5: timbre and sound characteristics

All participants described their experiences in relation to the sound characteristics, using either metaphor (e.g. "foggy" by

TABLE 5 Sensory alignment reported by participants.

| ID  | Primary instrument          | Years | Sound-producing gesture            | Kinaesthetic sense                  | Sound                                           |
|-----|-----------------------------|-------|------------------------------------|-------------------------------------|-------------------------------------------------|
| P2  | Drum kit                    | 12    | Individual medium line hatching    | “Plucking strings”                  | “Variation of plucky sounds”                    |
| P2  | Drum kit                    | 12    | Continuous scribbling              | “Rubbing something to a chalkboard” | “Harsh”                                         |
| P3  | None or “prefer not to say” | /     | Individual short line hatching     | “Strumming a guitar with a pick”    | “Guitar strumming”                              |
| P5  | Guitar                      | 5     | Short lines stippling              | “Mimicking footsteps”               | “Wooden and noisy”                              |
| P7  | Cello                       | 20    | Line hatching with flowy movements | “Playing a cello bow”               | Expecting the sound to be “sustained and flowy” |
| P7  | Cello                       | 20    | Short continuous line hatching     | “Plucking strings”                  | “Plucking strings”                              |
| P10 | Guitar                      | 22    | Short continuous line hatching     | “Stirring jar of marble balls”      | “Shaking a jar of marble balls”                 |
| P11 | None or “prefer not to say” | /     | Continuous scribbling              | “Drawing with a chalk”              | “Similar to drawing with a chalk”               |

P6), timbre descriptors (e.g., pitch, “percussive” by P3, “harsh” by P10), or emotions (e.g., “feel a bit down” by P8). Ten out of the 12 indicated that they attempted to inspect the sound characteristics in different parts of the terrains. And 6 of them either identified that the second terrain had higher spectral complexity or expressed similar ideas with different terminologies. For instance, P1 explained that they used simple and repetitive loops of circles as the primary technique on the second terrain because the sound produced by a loop already has a certain level of complexity.

Participants’ perception of sound can also affect how they perform movements. For instance, P10 described their creation process as “discovering the characteristic of sounds in a [canvas] location and then coming up with a movement that amplifies the sound characteristic.” To illustrate this idea, they discovered that the side to the right of the canvas sounded harsher than the left side, so they came up with the “burst of scrambling” technique on the right side to make the results even harsher.

### 5.3.6 T6: visual cues and visual interpretation

P1, P5, and P7 tried to read from the sketches for the figurative understanding of the sound. For instance, P1 expected a rounded shape to produce a “rounder and smoother” sound. In Section 5.2.2, we have described how participants used notations in the scores as visual cues to guide their navigation trajectories.

## 6 Discussion

In the study, we gained insight into participants’ progression of techniques in relation to the sound-producing gestures they explored, their subjective perception, and sonic capacities afforded in the latent space. This section discusses our study’s results in relation to our research question: **How do musicians perceive gestural affordances when navigating audio latent space and tailor them into performance techniques for musical expression?**

### 6.1 Discovering gestural affordances with various skills and capacities

Participants’ skills and individual capacities to engage with the latent space can shape the gestural affordances they perceive. On the broader landscape of affordances, [Rietveld and Kiverstein \(2014\)](#) suggested that the rich variety of physical and sociocultural backgrounds coordinate how actionable possibilities are discovered. This aligns with our observation, as described in Section 3, the two terrains encapsulated in the instrument were designed to have distinct spectral complexities - the second terrain is deliberately more complex than the first one, and the sound it produced is sensitive to microscopic movements. In this setting, participants’ interpretations of the spectral complexity varied. As described in T2 in Section 5.3.2, P2, P7, P9, and P10 were able to discover that slowing down their movements allows steady control over fine-level details of the sound. Since the constraints on how the control interface is used were minimal in our study, we see an inclination to rely on initial skills and capacities to navigate the actionable possibilities. For instance, P7, as a cellist, explicitly mentioned that they attempted to sculpt the sound with microscopic movements, whereas P2, who has a background in performing on drum kits, emphasizes the rhythmic aspect of their use of gestures. In this respect, considering physical constraints for the movements could be an important point in adding design interventions to guide how affordances are perceived.

We also found various ways of interpreting the movement of “navigation” in the latent space. This aligns with the ecologies and processes aspect of musical instruments ([Rodger et al., 2020](#)). That is, the affordances of an instrument are not a linear sum of its sonic capacities, instead individual capacities affect how they subvert the design brief of an instrument and discover the hidden affordances ([Parkinson, 2013](#)). In our observations, participants extended the initial sound-producing gestures we demonstrated to them, and discovered their constellation of affordances in the latent space. Given the results of our narrative analysis, we derive the following ways of how participants in our study use gestures for latent space navigation. In particular, we interpret them as our participants’ expectations of what the instrument affords musically, elucidating some of the future design spaces of DMIs with latent space navigation:

- *Using gestures to activate sonic materials:* For instance, in P10's score for the first terrain, they typically used bursts of short movement to trigger 0.5–1 second sonic objects (Godoy, 2006), and arranged these objects into sequences for compositions. The latent space serves as an unknown space in which musical materials are placed in different spatial locations, and the role of gestures is to visit these spatial locations to activate them.
- *Using gestures to define the trajectory of sonic objects:* For instance, P1 and P2's scores involve repetitive closed circles or lines, to annotate the trajectory of their gesture. Musical materials in this case are slightly longer 1–2 second sonic objects, often used as a loop. The role of gesture in their navigation is to define and trace the trajectory of sonic objects. This option is similar to a modern audio sampler, but the displacement of samples can extend from a one-dimensional timeline to a two-dimensional plane or higher-dimensional space.
- *Using gestures' coordinates as an XY-pad:* For instance, P4 and P10 annotated timbre descriptors to specific regions on the canvas, or attempted to figure out the effect of the X and Y coordinates. And P4 stopped their movement and let the sound freeze for a few seconds. This is similar to an electronic synthesizer with parametric controls, in which latent space navigation is to modify the value for each parameter to sculpt the sound. The role of gesture here is to drag the handle on an XY-pad.
- *Using properties of gestures to control attributes of sound:* For instance, P4 and P11's ways of using gestures, in contrast with the others, do not rely on absolute spatial locations of the canvas. In particular, they tend to compose a sequence of gestures and re-enact without a fixed canvas location, and use the system as a gestural control interface in which high-level properties, such as the gesture's speed or entropy, should be mapped to specific attributes of sound.

## 6.2 Developing tailorable techniques in the latent space

The control interface of the latent space hosts an initial set of affordances, in our case, the stylus and tablet. Musicians' kinesthetic and sensorimotor perceptions of gestures largely depend on their postures, such as pen grips (Hinckley et al., 2014). Meanwhile, the latent space further defines the auditory perception by its action-sound mapping. This combination of gestural and auditory perception forms the subjective experience of latent space navigation. Previous research has suggested that this subjective bodily experience plays a central role in bundling perceived affordances into a repertoire of musical skills (Zappi and McPherson, 2018; Mice and McPherson, 2022; Bang and Fdili Alaoui, 2023). In our findings, we also observed this formation and deployment of performance techniques in the latent space, and how gestural and auditory experience can guide musicians' choices of techniques. According to our results, the role of subjective perceptions in the latent space is twofold: (i) it helps one rely on *sensory alignment* to reconcile unknowns in the latent space, and

(ii) it constrains one's choice of techniques by *bodily and cognitive effort*. Here we elaborate on these two aspects.

### 6.2.1 Sensory alignments

We identified that musicians tend to rely on sensory alignments (Marshall et al., 2019) to reconcile unknowns in the latent space with familiar auditory (i.e., how it sounds) and kinesthetic (i.e., how the hand moves) combinations. This can be observed from our Thematic Analysis in T4: Action-Sound Mapping, Section 5.3.4 that shaped 7 out of the 12 participants' choices of techniques. In particular, as the NAS model encapsulated in our instrument was trained on recordings of guitar picking, four participants mentioned the motion of strumming or plucking strings. In addition, closely inspecting the eight techniques that involve sensory alignment and musical background described in Table 5, we noticed that these sensory alignments can involve either abstract feelings or realistic depictions, and can go beyond their previous experience and musical training. For instance, P4, who primarily used the piano, mentioned “flowy” to describe fluidly moving a cello bow for a sustained legato, and they expected a sustained sound from this movement. In contrast, P2 and P3, who did not mention the guitar as their primary instrument, highlighted the short impulsive line hatching techniques because the movement and the resulting sounds are both close to picking a guitar string.

We observed examples of participants' progression of techniques affected by sensory alignments. For instance, in Section 5.3.5, we quoted how P10 refined their performance techniques to align their kinesthetic sense to expressions they perceived in the auditory domain. Another example is that in their exploration activity, P2 initially used the “iterative line hatching” technique, which is to place lines in a back-and-forth manner without lifting their pen from the canvas. However, they switched to “impulsive line hatching” when creating the score for the second terrain, and they explicitly stated that the gesture of “impulsive line hatching” felt similar to “plucking strings.” This change of technique can be seen in Table 3 that P2's record of “medium iterative line strokes” only appeared in the exploration activity, whereas “medium impulsive line strokes” only appeared in their scores. Similar results also appear in the record for P3 (“short iterative line strokes” changed to “short impulsive line strokes”) and P7, who both also mentioned the same sensory alignment in Table 5.

We suggest that investigating musicians' longer-term engagement with the latent space is an important direction for future study. As described in T4 Section 5.3.4, we observed that participants' familiarity with the movements changed as they practiced. However, due to the time limitation (90 minutes) of our research workshop, we are not able to gain more in-depth insight into how this familiarity with sensory alignment changes in longer-term practice. Other studies in sensory alignment that last for days (Tennent et al., 2020) have revealed valuable insights into participants' adaption of the body and ways of interaction. In addition, longer-term engagement is an important aspect of fostering reflection (Morreale and McPherson, 2017) and gaining insights that are deeply situated in creative practice with new musical instruments (Ford et al., 2024). Therefore, studies over

a longer period would have better uncovered how participants' techniques evolve with their subjective experience.

We also suggest future work to design for sensory alignment or sensory misalignment between the kinesthetic movements in the control interfaces and the sonic responses in the latent space. Marshall et al. (2019) suggested the use of sensory alignment to reconcile the unknowns, and the use of sensory misalignment for previously impossible experiences. Given that latent spaces as platforms for musical expressions have introduced a broad and open-ended design space for their control interfaces (Tahiroğlu and Wyse, 2024), it can be a useful material to build technology probes (Hutchinson et al., 2003) to look into the design and evaluation of experience, usability, or acceptance of the human-AI interactions (Tchemeube et al., 2023) that take into account musicians' previous experience and musical training, which is a critical aspect in situating new musical affordances of AI tools into music-making practices (Louie et al., 2020).

### 6.2.2 Bodily and cognitive effort

We identified that musicians' choice of techniques tends to be constrained by the perceived effort of sound-producing gestures. This effort here is twofold: bodily and cognitive. Bodily effort (Vertegaal and Ungvary, 1996) in our context refers to the physical effort perceived by one's hand, wrist, and arm when performing a gesture on the stylus. Cognitive effort in our context refers to whether the musician is attending to the movement (Bang and Fdili Alaoui, 2023) to maintain virtuosic and subtle performance of a gesture or a stencil pattern.

In terms of bodily effort, as quoted in Section 5.3.1, P3 commented that hatching lines in the top-right to bottom-left diagonal feel "natural", whereas in the top-left to bottom-right diagonal feel "disrupted" and "in the wrong direction." For right-handed participants, moving a pen in the top-right to bottom-left diagonal requires an abduction or adduction (rotating the hand toward or away from the side of the thumb) movement, whereas the top-left to bottom-right requires flexing multiple fingers. It is confirmed in research on hand tool development (Takayama et al., 2015) that the former technique requires fewer joints to perform. Therefore, right-handed participants hatching lines in such a direction results in a movement that requires less effort. A key observation of how this bodily effort affects participants' choice of techniques is that P2 and P3's sketching logs, notes, and scores have prevalent records of lines tilted toward the canvas's bottom-left to top-right direction.

Bodily effort can be largely affected by the size of the instrument, and previous research has typically referred to the body as the entire full-body experience (Mice and McPherson, 2021). Similar findings from Mice and McPherson (2022) have confirmed that performers typically prioritized the comfortability of the body over the sound when deploying performance techniques. This echoes the bodily affordance framework (de Vignemont, 2015) that the body also limits movement navigation. In this respect, the technique development in the latent space is constrained by the body through effort perceived by the musicians. However, since the size of our instrument's interface is relatively small, we limit our

claim specifically to hand-related effort and open the discussion on latent space navigation with different-size interfaces to future work.

In terms of cognitive effort, the attention required to virtuosically perform a gesture has been mentioned repetitively by participants. We noticed from participants' scores that, in our study, two types of drawing gestures emerged. The first type uses a clear intention of maintaining the shape of lines. Three participants [P7, P8, and P11] commented that they either intentionally focused on tracing the shape or consciously maintained a stable motion, and therefore paid high conscious effort to attend to the movement instead of the sound. The second type of gestures uses hurried and casual movement, such as P10's scores on both terrains, P9's score on Terrain 1, and P3's casual drawings on the bottom-left corner of their score on Terrain 2. They tended to enjoy the gestural aspect of this type of technique, which can require "less thinking about the movement" (paraphrased from P3).

This echoes findings in movement-based musical instruments (Bang and Fdili Alaoui, 2023) on the delineation between attending to the embodied movements and attending to the sound. It has been brought up as the virtuosity and subtlety aspect of DMI design (Jordà, 2004), in which instruments that allow for attention to detail and the development of craft have proven to be more musically interesting. The ability to sculpt the fine details of techniques can be missed in a typical techno-centered development (Bryan-Kinns et al., 2025). However, observing existing musical practices with AI-enhanced musical instruments (Tahiroğlu et al., 2021; Françoise et al., 2022), the affordance of being able to continuously engage with the virtuosity can be a demanding aspect in designing for DMIs with audio latent space.

## 6.3 Reflections and limitations of the method

Our study with musicians was purposefully open-ended, and we encouraged creative and non-rigorous ways of using documentary notes. This openness offers insight into participants' diverse ways of exploration and sense-making. However, as with other works that used diverse and open-ended prompts in HCI (Sturdee et al., 2024), we found that generating insights was a challenging process due to the subjectivity of data and individual nuances (Sturdee, 2025). Our attempt to reconcile this subjectivity with criticality and validity was by first referencing our narrations to the literature on tablet and pen interfaces (see Section 4.4), and then focusing on individual analysis within each participant's journey. This approach helped us to observe progressions within participants' individual experiences with the instrument, instead of converging all participants' experiences into a unified "design brief." We suggest future works on designing interfaces for latent space navigation to support users' subjectivity instead of constraining activities (Dix, 2007).

Combining the documentary notes with the scores offered insights into how techniques participants explored were reflected in the final scores they created. In contrast to the soma trajectory tool (Tennent et al., 2021) that focuses on events and progressions through the experience, or broader methods that do not limit the format of documentations (Sturdee et al., 2024), we see our note



template as a way for participants to “stamp collect” their findings in a systematic way through their experience. This approach is similar to the material exploration (Paymal and Fdili Alaoui, 2023) in a broader design context, in which subjects explore and make sense of materials and turn them into artifacts. This is an approach that focuses on how individual participants use their explored techniques as materials and tailor them to scores.

Our insights and findings are limited to our neural synthesis model’s architecture and training data. We propose future studies could explore a plurality of models, data, and designs for generalizable insights across a broader range of latent spaces. In addition, a stylus and tablet interface itself constrains gestural affordances to be two-dimensional and centered around the fingers, hand, wrist, and arm. While this serves the purpose of being minimal, this approach already closes down the possibility for more advanced gestures such as a complex finger flex as in Visi et al. (2024). We encourage future work to consider a broader space for gestures, such as movements involving the full body as in Françoise et al. (2022), or micro-techniques such as muscle contraction as in Erdem and Jensenius (2020).

Our observations described in the previous sections focus on accounting for individual musicians’ journeys and experiences, and are qualitative due to the small sample size. Trends and voices from the participants could have been substantiated in a larger quantitative study. Therefore, here we summarize potential factors and directions for future quantitative studies. First, in Section 6.1 we described how previous musical training may impact the initial set of affordances participants chose to explore. For instance, training on physical musical instruments, quantitative metrics of ability to engage with music such as the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014), or the ability to image movements, especially using kinesthetic imagery such as the Movement Imagery Questionnaire 3 (Williams et al., 2012), could have better measured musicians’ skills and capacities. Activity measures such as interaction logs and activity maps can effectively assess participants’ engagement with the material (Bryan-Kinns, 2013; Nacenta et al., 2010). Measures such as the entropy of the drawing strokes (Daniele et al., 2021) have also been used as dependent variables to assess the content of participants’ interaction. In addition, we used two research probes (latent terrains) that offer distinct spectral complexities. We do not claim the two terrains as baseline and variation, instead we see them as two prompts with equal capacities in musical expressions. Indeed, moments of excitement highlighted by participants were reported on both terrains, and we did not observe a strong preference. This echoes Privato et al. (2024)’s finding that perception of algorithmic adaptation on a musical instrument typically depends on the musicians’ interests rather than on the choices of the designer. We suggest using more obvious aspects of the latent space or the interface, such as the training data, display of terrain visualization, or display of the drawing history, as independent variables.

Finally, we acknowledge that visual interpretations of the drawing strokes can play an important role in discovering affordance (Turvey, 1992) in our study. As described in T6 in Section 5.3.6, P1, P5, and P7 expected alignment between the semantic aspect of their drawings and the auditory aspect of the resulting sound. This shape and sound association has

been identified in previous literature (Löbbers et al., 2023). We interpret participants’ expectations of visual-auditory alignment as inspirations that come from the capabilities of AI Generated Content (AIGC) in other creative domains such as text-to-sound (Agostinelli et al., 2023) and image-to-sound translation (Zheng et al., 2024a). This is out of the scope of this study but has the potential as future work.

## 6.4 Summary

Participants’ creations and demonstrations of the musical scores enabled us to observe their perceived gestural affordances, which were shaped by their musical skills and capacities to engage with the latent space. These insights also suggested four ways of using gestures in latent space navigation: *using gestures to activate sonic materials*, *using gestures to define the trajectory of sonic objects*, *using gestures’ coordinates as an XY-pad*, and *using properties of gestures to control attributes of sound*. The narratives of participants’ explorations and demonstrations revealed two roles of subjective gestural and auditory perceptions in tailoring performing techniques in audio latent space: (i) they help one rely on *sensor alignment* to reconcile unknowns in the latent space, and (ii) they constrain one’s choice of techniques by *bodily and cognitive effort*.

## 7 Conclusion

This article explored how musicians perceive gestural affordances when navigating audio latent space and develop performance techniques for musical expression. We designed a DMI with a stylus and a tablet interface, embedded with latent spaces of a neural audio synthesis model, as a research probe to invite musicians to actively test out open-ended gestures with the stylus, and tasked them to create musical scores for the instrument. We contributed findings from an embodied music cognition perspective of how subjective perceptions of sound-producing gestures affect musicians’ technique development in latent space navigation. We also suggested four ways of using gestures in audio latent spaces discovered by participants in our workshop, aiming to elucidate new opportunities for gestural interface design for audio latent space navigation, and complement the literature on new musical affordances of latent spaces.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

## Ethics statement

The studies involving humans were approved by the Queen Mary University of London Research Ethics Committee (Reference

number: QMERC20.565.DSEEC24.068). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

SJZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. AXS: Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. NB-K: Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Shuoyang Jasper Zheng is a research student at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported by UK Research and Innovation [grant number EP/S022694/1].

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomp.2025.1575202/full#supplementary-material>

## References

- Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., et al. (2023). MusicLM: generating music from text. *arXiv:2301.11325*.
- Anne Cochrane, K., Mah, K., Ståhl, A., Núñez-Pacheco, C., Balaam, M., Ahmadpour, N., et al. (2022). "Body maps: a generative tool for soma-based design," in *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '22 (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3490149.3502262
- Avila, J. M., Tsaknaki, V., Karpashevich, P., Windlin, C., Valenti, N., Höök, K., et al. (2020). "Soma design for NIME," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, eds. R. Michon, and F. Schroeder (Birmingham, UK: Birmingham City University), 489–494.
- Banar, B., and Colton, S. (2022). "Connecting audio and graphic score using self-supervised representation learning" a case study with Gyorgy Ligeti's *Artikulation*," in *Proceedings of the Thirteenth International Conference on Computational Creativity* (Bozen-Bolzano, Italy: The Association for Computational Creativity).
- Bang, T. G., and Fdili Alaoui, S. (2023). "Suspended circles: soma designing a musical instrument," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23 (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3544548.3581488
- Benford, S., Giannachi, G., Koleva, B., and Rodden, T. (2009). "From interaction to trajectories: designing coherent journeys through user experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09 (New York, NY, USA: Association for Computing Machinery), 709–718. doi: 10.1145/1518701.1518812
- Benford, S., Greenhalgh, C., Crabtree, A., Flintham, M., Walker, B., Marshall, J., et al. (2013). Performance-led research in the wild. *ACM Trans. Comput.-Hum. Interact.* 20, 1–22. doi: 10.1145/2491500.2491502
- Bertissolo, G. (2019). "Composing understandings: music, motion, gesture and embodied cognition," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, eds. M. Queiroz, and A. X. Sedó (Porto Alegre, Brazil: UFRGS), 361–364.
- Bryan-Kinns, N. (2013). Mutual engagement and collocation with shared representations. *Int. J. Hum. Comput. Stud.* 71, 76–90. doi: 10.1016/j.ijhcs.2012.02.004
- Bryan-Kinns, N. (2024). Reflections on explainable AI for the Arts (XAIxArts). *Interactions* 31, 43–47. doi: 10.1145/3636457
- Bryan-Kinns, N., and Reed, C. N. (2023). "A guide to evaluating the experience of media and arts technology," in *Creating Digitally: Shifting Boundaries: Arts and Technologies—Contemporary Applications and Concepts*, ed. A. L. Brooks (Cham: Springer International Publishing), 267–300. doi: 10.1007/978-3-031-31360-8\_10
- Bryan-Kinns, N., Zhang, B., Zhao, S., and Banar, B. (2024). Exploring variational auto-encoder architectures, configurations, and datasets for generative music explainable AI. *Mach. Intell. Res.* 21, 29–45. doi: 10.1007/s11633-023-1457-1
- Bryan-Kinns, N., Zheng, S. J., Castro, F., Lewis, M., Chang, J.-R., Vigliensoni, G., et al. (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 (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3706599.3716227
- Caillon, A., and Esling, P. (2021). RAVE: a variational autoencoder for fast and high-quality neural audio synthesis. *arXiv:2111.05011*.
- Caillon, A., and Esling, P. (2022). Streamable neural audio synthesis with non-causal convolutions. *arXiv:2204.07064*.
- Casey, S. (2018). What do drawing and painting really mean? The phenomenology of image and gesture. *J. Visual Art Pract.* 17, 238–240. doi: 10.1080/14702029.2017.1366693
- Chris, C., and Michael, G. (2004). Network time delay and ensemble accuracy: effects of latency, asymmetry. *J. Audio Eng. Soc.* 117:6208.

- Clarke, E. F. (2005). *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195151947.001.0001
- Clarke, V., and Braun, V. (2017). Thematic analysis. *J. Posit. Psychol.* 12, 297–298. doi: 10.1080/17439760.2016.1262613
- Dalgleish, M. (2014). “Reconsidering process: bringing thoughtfulness to the design of digital musical instruments for disabled users,” in *International Conference on Live Interfaces (ICLI)*.
- Daniele, A., Di Bernardi Luft, C., and Bryan-Kinns, N. (2021). “What is human? A Turing test for artistic creativity,” in *Artificial Intelligence in Music, Sound, Art and Design*, eds. J. Romero, T. Martins, and N. Rodríguez-Fernández (Cham: Springer International Publishing), 396–411. doi: 10.1007/978-3-030-72914-1\_26
- de Vignemont, F. (2015). “Bodily affordances and bodily experiences,” in *Perceptual and Emotional Embodiment: Foundations of Embodied Cognition*, eds. Y. Coello, and M. H. Fischer (London: Routledge/Taylor & Francis Group).
- Dix, A. (2007). “Designing for appropriation,” in *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK* (BCS Learning & Development).
- Eckhardt, J. (2022). *Going Out - Walking, Listening, Soundmaking*. Brussels: Umland editions.
- Erdem, C., and Jensenius, A. R. (2020). “RAW: exploring control structures for muscle-based interaction in collective improvisation,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, eds. R. Michon, and F. Schroeder (Birmingham, UK: Birmingham City University), 477–482.
- Fitzpatrick, G. (2003). *The Locales Framework: Understanding and Designing for Wicked Problems*. New York: Springer Science & Business Media. doi: 10.1007/978-94-017-0363-5
- Ford, C., Noel-Hirst, A., Cardinale, S., Loth, J., Sarmiento, P., Wilson, E., et al. (2024). “Reflection across AI-based music composition,” in *Proceedings of the 16th Conference on Creativity & Cognition, C&C '24* (New York, NY, USA: Association for Computing Machinery), 398–412. doi: 10.1145/3635636.3656185
- Françoise, J., Fdili Alaoui, S., and Candau, Y. (2022). “CO/DA: live-coding movement-sound interactions for dance improvisation,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3491102.3501916
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception: Classic Edition*. New York: Psychology Press. doi: 10.4324/9781315740218
- Godøy, R. I. (2006). Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. *Organ. Sound* 11, 149–157. doi: 10.1017/S1355771806001439
- Godøy, R. I. (2009). “Gestural affordances of musical sound,” in *Musical Gestures* (Routledge). doi: 10.4324/9780203863411
- Godøy, R. I. (2018). “Sonic object cognition,” in *Springer Handbook of Systematic Musicology*, ed. R. Bader (Berlin, Heidelberg: Springer), 761–777. doi: 10.1007/978-3-662-55004-5\_35
- Godøy, R. I., and Leman, M. (2009). *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. New York: MIT Press.
- Hinckley, K., Pahud, M., Benko, H., Irani, P., Guimbretière, F., Gavrilu, M., et al. (2014). “Sensing techniques for tablet+stylus interaction,” in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, UIST '14* (New York, NY, USA: Association for Computing Machinery), 605–614. doi: 10.1145/2642918.2647379
- Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., et al. (2003). “Technology probes: inspiring design for and with families,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03* (New York, NY, USA: Association for Computing Machinery), 17–24. doi: 10.1145/642611.642616
- Huzaifah, M., and Wyse, L. (2021). “Deep generative models for musical audio synthesis,” in *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*, ed. E. R. Miranda (Cham: Springer International Publishing), 639–678. doi: 10.1007/978-3-030-72116-9\_22
- Jacobs, J., Gogia, S., Mundefinedch, R., and Brandt, J. R. (2017). “Supporting expressive procedural art creation through direct manipulation,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (New York, NY, USA: Association for Computing Machinery), 6330–6341. doi: 10.1145/3025453.3025927
- Jensenius, A. R. (2022). *Sound Actions: Conceptualizing Musical Instruments*. London: The MIT Press. doi: 10.7551/mitpress/14220.001.0001
- Jordá, S. (2004). Instruments and players: Some thoughts on digital lutherie. *J. New Music Res.* 33, 321–341. doi: 10.1080/0929821042000317886
- Kamath, P., Morreale, F., Bagaskara, P. L., Wei, Y., and Nanayakkara, S. (2024). “Sound designer-generative AI interactions: towards designing creative support tools for professional sound designers,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI '24* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3613904.3642040
- Kingma, D. P., and Welling, M. (2013). “Auto-encoding variational Bayes,” in *Proceedings of the International Conference on Learning Representations*.
- Kuzmin, I., Ma, Z., and Masu, R. (2024). “Locality and digital musical instruments design: a user study,” in *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures, AM '24* (New York, NY, USA: Association for Computing Machinery), 468–478. doi: 10.1145/3678299.3678347
- Kvifte, T. (2008). On the description of mapping structures. *J. New Music Res.* 37, 353–362. doi: 10.1080/09298210902731394
- Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. New York: The MIT Press. doi: 10.7551/mitpress/7476.001.0001
- Leman, M. (2012). “Musical gestures and embodied cognition,” in *Journées d'Informatique Musicale, Mons, Belgium*.
- Lepri, G., Privato, N., and Magnusson, T. (2024). “Embodied sketching for neural synthesis,” in *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures, AM '24* (New York, NY, USA: Association for Computing Machinery), 549–551. doi: 10.1145/3678299.3678358
- Löbbers, S., Thorpe, L., and Fazekas, G. (2023). “SketchSynth: cross-modal control of sound synthesis,” in *Artificial Intelligence in Music, Sound, Art and Design*, eds. C. Johnson, N. Rodríguez-Fernández, and S. M. Rebelo (Cham: Springer Nature Switzerland), 164–179. doi: 10.1007/978-3-031-29956-8\_11
- Lohan, F. (2012). *Sketching Domestic and Wild Cats: Pen and Pencil Techniques*. Dover Art Instruction Series. Courier Corporation.
- López-Cano, R. (2006). “What kind of affordances are musical affordances? A semiotic approach,” in *Lascolto musicale: condotte, pratiche, grammatiche* (Bologna), 23–25.
- Louie, R., Coenen, A., Huang, C. Z., Terry, M., and Cai, C. J. (2020). “Novice-AI music co-creation via AI-steering tools for deep generative models,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20* (New York, NY, USA: Association for Computing Machinery), 1–13. doi: 10.1145/3313831.3376739
- Magnusson, T. (2009). Of epistemic tools: musical instruments as cognitive extensions. *Org. Sound* 14, 168–176. doi: 10.1017/S1355771809000272
- Magnusson, T. (2010). Designing constraints: composing and performing with digital musical systems. *Comput. Music J.* 34, 62–73. doi: 10.1162/COMJ\_a\_00026
- Marshall, J., Benford, S., Byrne, R., and Tennent, P. (2019). “Sensory alignment in immersive entertainment,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19* (New York, NY, USA: Association for Computing Machinery), 1–13. doi: 10.1145/3290605.3300930
- Mice, L., and McPherson, A. P. (2021). “Embodied cognition in performers of large acoustic instruments as a method of designing new large digital musical instruments,” in *Perception, Representations, Image, Sound, Music*, eds. R. Kronland-Martinet, S. Ystad, and M. Aramaki (Cham: Springer International Publishing), 577–590. doi: 10.1007/978-3-030-70210-6\_37
- Mice, L., and McPherson, A. P. (2022). “Super size me: interface size, identity and embodiment in digital musical instrument design,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3491102.3517626
- Mitsuhashi, Y. (1982). Audio signal synthesis by functions of two variables. *J. Audio Eng. Soc.* 30, 701–706.
- Morreale, F., and McPherson, A. (2017). “Design for longevity: ongoing use of instruments from NIME 2010–14,” in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Copenhagen, Denmark: Aalborg University Copenhagen), 192–197.
- Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). Measuring the facets of musicality: the Goldsmiths Musical Sophistication Index (Gold-MSI). *Pers. Individ. Dif.* 60:S35. doi: 10.1016/j.paid.2013.07.081
- Nacenta, M. A., Pinelle, D., Gutwin, C., and Mandryk, R. (2010). “Individual and group support in tabletop interaction techniques,” in *Tabletops - Horizontal Interactive Displays*, ed. C. Müller-Tomfelde (London: Springer London), 303–333. doi: 10.1007/978-1-84996-113-4\_13
- Nijs, L., Grinspun, N., and Fortuna, S. (2024). Developing musical creativity through movement: navigating the musical affordance landscape. *Creat. Res. J.* 37, 427–451. doi: 10.1080/10400419.2023.2299159
- Núñez-Pacheco, C. (2021). “Tangible body maps of felt-sensing experience,” in *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '21* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3430524.3442700
- Parkinson, A. (2013). “Embodied listening, affordances and performing with computers,” in *International Computer Music Conference Proceedings*, 162–168.
- Paymal, L., and Fdili Alaoui, S. (2023). “Physicalizing loops,” in *Proceedings of the 15th Conference on Creativity and Cognition, C&C '23* (New York, NY, USA: Association for Computing Machinery), 465–477. doi: 10.1145/3591196.3593365



- Privato, N., Shepardson, V., Lepri, G., and Magnusson, T. (2024). "Stacco: exploring the embodied perception of latent representations in neural synthesis," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 424–431.
- Reybrouck, M. (2005). Body, mind and music: musical semantics between experiential cognition and cognitive economy. *Trans. Revista transcultural de música* 9, 1–55.
- Rietveld, E., and Kiverstein, J. (2014). A rich landscape of affordances. *Ecol. Psychol.* 26, 325–352. doi: 10.1080/10407413.2014.958035
- Rodger, M., Stapleton, P., van Walstijn, M., Ortiz, M., and Pardue, L. S. (2020). "What makes a good musical instrument? A matter of processes, ecologies and specificities," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, eds. R. Michon, and F. Schroeder (Birmingham, UK: Birmingham City University), 405–410.
- Roma, G. (2023). Agent-based music live coding: sonic adventures in 2D. *Organ. Sound* 28, 231–240. doi: 10.1017/S1355771823000274
- Roma, G., Green, O., and Tremblay, P. A. (2019). "Adaptive mapping of sound collections for data-driven musical interfaces," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ed. A. Xambó Sedó (Porto Alegre, Brazil: UFRGS), 313–318.
- Saitis, C., Del Sette, B. M., Shier, J., Tian, H., Zheng, S., Skach, S., et al. (2024). "Timbre tools: ethnographic perspectives on timbre and sonic cultures in hackathon designs," in *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures*, AM '24 (New York, NY, USA: Association for Computing Machinery), 229–244. doi: 10.1145/3678299.3678322
- Scurto, H., and Postel, L. (2023). "Soundwalking deep latent spaces," in *Proceedings of the 23rd International Conference on New Interfaces for Musical Expression (NIME'23)* (Mexico).
- Shaheed, N., and Wang, G. (2024). "I am sitting in a (latent) room," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 333–338.
- Sparkes, A. C., and Smith, B. (2008). "Narrative constructionist inquiry," in *Handbook of Constructionist Research* (Guilford Press, United States), 295–314.
- Ståhl, A., Tsaknaki, V., and Balaam, M. (2021). Validity and rigour in soma design-sketching with the soma. *ACM Trans. Comput.-Hum. Interact.* 28, 1–36. doi: 10.1145/3470132
- Stapleton, P., Walstijn, M., v., and Mehes, S. (2018). "Co-tuning virtual-acoustic performance ecosystems: observations on the development of skill and style in the study of musician-instrument relationships," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, eds. T. M. Luke Dahl, Douglas Bowman (Blacksburg, Virginia, USA: Virginia Tech), 311–314.
- Sturdee, M. (2025). A step toward formalising visual data analysis practices in human computer interaction. *Interact. Comput.* 2025:iwae063. doi: 10.1093/iwc/iwae063
- Sturdee, M., Gen, H. U., and Wanick, V. (2024). "Diversifying knowledge production in HCI: exploring materiality and novel formats for scholarly expression," in *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '24 (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3623509.3634743
- Sturdee, M., Lewis, M., Strohmayer, A., Spiel, K., Koulidou, N., Alaoui, S. F., et al. (2021). "A plurality of practices: artistic narratives in HCI research," in *Proceedings of the 13th Conference on Creativity and Cognition, C&C '21* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3450741.3466771
- Tahiroğlu, K., Kastemaa, M., and Koli, O. (2021). "AI-terity 2.0: an autonomous NIME featuring GANSpaceSynth deep learning model," in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Shanghai, China).
- Tahiroğlu, K., Magnusson, T., Parkinson, A., Garrelfs, I., and Tanaka, A. (2020). Digital musical instruments as probes: how computation changes the mode-of-being of musical instruments. *Organ. Sound* 25, 64–74. doi: 10.1017/S1355771819000475
- Tahiroğlu, K., and Wyse, L. (2024). "Latent spaces as platforms for sonic creativity," in *Proceedings of the 15th International Conference on Computational Creativity, Sweden*.
- Takayama, L., Merino, G., Merino, E., Garcia, L., Cunha, J., and Domenech, S. (2015). Hand tool project requirements: the case of banana cultivation and its physical demands (OWAS). *Product Manag. Dev.* 13, 119–130. doi: 10.4322/pmd.2015.012
- Tchemeube, R. B., Ens, J., Plut, C., Pasquier, P., Safi, M., Grabit, Y., et al. (2023). "Evaluating human-AI interaction via usability, user experience and acceptance measures for MMM-C: a creative AI system for music composition," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI '23* (Macao, P.R.China).
- Tennent, P., Höök, K., Benford, S., Tsaknaki, V., Stl, A., Dauden Roquet, C., et al. (2021). "Soma design and sensory misalignment," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3411764.3445482
- Tennent, P., Marshall, J., Tsaknaki, V., Windlin, C., Höök, K., and Alfaras, M. (2020). "Soma design and sensory misalignment," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20* (New York, NY, USA: Association for Computing Machinery), 1–12. doi: 10.1145/3313831.3376812
- Tesfaldet, M., Snelgrove, X., and Vazquez, D. (2019). "Fourier-CPPNs for image synthesis," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 3173–3176. doi: 10.1109/ICCVW.2019.00392
- Turvey, M. (1992). Affordances and prospective control: an outline of the ontology. *Ecol. Psychol.* 4, 173–187. doi: 10.1207/s15326969eco0403\_3
- Vertegaal, R., and Ungvary, T. (1996). "Towards a musician's cockpit: transducers feedback and musical function," in *1996 International Computer Music Conference, ICMC 1996* (Michigan Publishing), 308–311.
- Vigliensoni, G., and Fiebrink, R. (2023). "Steering latent audio models through interactive machine learning," in *Proceedings of the 14th International Conference on Computational Creativity* (Ontario, Canada).
- Visi, F., Schramm, R., Frödin, K., Unander-Scharin, A., and Östersjö, S. (2024). "Empirical analysis of gestural sonic objects combining qualitative and quantitative methods," in *Sonic Design*, ed. A. R. Jensenius (Cham: Springer Nature Switzerland), 115–137. doi: 10.1007/978-3-031-57892-2\_7
- Wan, Q., and Lu, Z. (2023). "Investigating semantically-enhanced exploration of GAN latent space via a digital mood board," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, CHI EA '23* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3544549.3585740
- Weijters, B., Cabooter, E., and Schillewaert, N. (2010). The effect of rating scale format on response styles: the number of response categories and response category labels. *Int. J. Res. Market.* 27, 236–247. doi: 10.1016/j.ijresmar.2010.02.004
- Williams, S. E., Cumming, J., Ntoumanis, N., Nordin-Bates, S. M., Ramsey, R., and Hall, C. (2012). Further validation and development of the movement imagery questionnaire. *J. Sport Exerc. Psychol.* 34, 621–646. doi: 10.1123/jsep.34.5.621
- Wilson, E., Schubert, D., Satomi, M., McLean, A., and Amaya Gozalez, J. F. (2023). "MosAIck: staging contemporary ai performance - connecting live coding, e-textiles and movement," in *Proceedings of the 7th International Conference on Live Coding* (Utrecht, Netherlands: Zenodo).
- Xambó Sedó, A. (2023). Discovering creative commons sounds in live coding. *Organ. Sound* 28, 276–289. doi: 10.1017/S1355771823000262
- Yee-King, M. (2022). "Latent spaces: a creative approach," in *The Language of Creative AI: Practices, Aesthetics and Structures*, eds. C. Vear, and F. Poltronieri (Cham: Springer International Publishing), 137–154. doi: 10.1007/978-3-031-10960-7\_8
- Zappi, V., and McPherson, A. (2018). "Dimensionality and appropriation in digital musical instrument design," in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Zenodo), 455–460.
- Zheng, S., Del Sette, B. M., Saitis, C., Xambó Sedó, A., and Bryan-Kinns, N. (2024a). "Building sketch-to-sound mapping with unsupervised feature extraction and interactive machine learning," in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Utrecht, Netherlands).
- Zheng, S., Xambó Sedó, A., and Bryan-Kinns, N. (2024b). "A mapping strategy for interacting with latent audio synthesis using artistic materials," in *Proceedings of the second international workshop on eXplainable AI for the Arts (XAIxArts 2)*.