AFTERnaut: A Multi-Modal Interactive Interface for Embodied Exploration of Neural Synthesis Latent Spaces

Evan O'Donnell
Department of Computing
Goldsmiths, University of London
London, UK SE14 6NW
eodon002@gold.ac.uk

Atau Tanaka
Department of Computing
Goldsmiths, University of London
London, UK SE14 6NW
a.tanaka@gold.ac.uk

Abstract

We present a multimodal interactive interface for exploration and embodied mapping of complex neural synthesis latent spaces. Users explore new sounds via manual adjustment of latent parameters, LFO activation of the latent space, and embodied navigation of latent dimensions via multimodal sensor inputs. They save, apply, and create variations on these sounds through a preset system, three-part dataset structure, and interactive machine learning workflow including both classification and regression. These affordances are collected in a single easy-to-navigate GUI. We evaluated this system through personal practice and a user study via qualitative methods, presenting the interface and allowing exploration to gather feedback. Most users felt they could identify interesting sounds through the system and build embodied interactions which felt intuitive or useful, helping them develop more intentional relationships with the latent spaces. This supports bespoke and embodied applications of neural audio synthesis models by non-specialist musicians.

1 INTRODUCTION

Although generative music AI has advanced rapidly, there is a significant research gap on how musicians can use these models in individual practice. Creative manipulation of neural audio synthesis (NAS) latent spaces is a promising and increasingly common approach to this. However, interpreting these latent spaces can be intimidating for musicians, and "little research has been done to characterize the latent spaces of audio models, and to investigate their affordances in practical musical scenarios" (Privato et al., 2024). One solution is to develop interactive systems which help musicians to interpret or "grasp" these latent spaces and apply them intentionally in their work.

Musical practice is strongly embodied (Leman, 2008), and often relies on tacit and implicit forms of knowledge (Bowman, 1982; Mitchell et al., 2026). As creative practitioners, we are therefore especially curious about systems which let musicians explore latent parameters with their ears and their bodies, rather than through labels and tags. Recent research demonstrates the feasibility of embodied exploration and mapping of NAS latent spaces (Nabi et al., 2024). Combining this strategy with ear-led approaches to latent space exploration in a multi-step workflow could allow musicians to apply existing musical preferences and embodied knowledge in developing intuitive, bespoke uses of these models.

In this study, we ask:

- How can we create interfaces which assist musicians in working with the latent spaces of NAS models?
- Can personalised, embodied relationships support novel creative interactions with NAS?
- Together, can these user interfaces and embodied interactions help artists "come to grips" with NAS latent spaces and identify results they find interesting?

To investigate these questions, we built an interface combining latent space exploration and embodied interaction workflows, testing this in our own practices and with others. We specifically worked with AFTER models (Demerlé et al., 2024), which offer both especially complex latent representations and the ability to work with both structural and timbral musical characteristics through latent space navigation.

We present:

- An interface combining multiple tools for purposeful and productive exploration of NAS latent spaces, and for developing bespoke embodied relationships with these.
- A multi-step workflow, including body and ear-driven exploration, preset and dataset building, and machine learning on gesture-sound associations.
- Documented insights from evaluations of this system within personal practice and in user studies.

2 BACKGROUND

The rapid growth of generative AI offers new pathways for music creation, but has also raised concerns over the creative role of individual musicians. Text-prompt-based music generation in particular has drawn criticism for homogenising musical output (Bryan-Kinns et al., 2025). There is growing discussion on ways musicians can harness these models for individualised purposes (Nika et al., 2025), including approaches which see AI more as a material or medium to work with than a tool for reproducing a targeted outcome, sometimes described as divergent uses of AI (Broad et al., 2021; Chemla-Romeu-Santos and Esling, 2022).

Variational Autoencoders (VAE), and particularly RAVE (Caillon and Esling, 2021), are a common model architecture in these exploratory musical applications of AI, falling

under the broader umbrella of neural audio synthesis (NAS) techniques. RAVE models are trained on a large corpus of audio files, distributing representations of musical information within a multidimensional latent space. They generate audio by encoding inputs as latent vectors in this distributed space and then decoding them. IRCAM's *nn~* objects (Caillon and Chemla—Romeu-Santos, 2026) offer a means of directly manipulating this latent space through a reduced dimensional representation in creative coding environments such as Max and Pure Data. However, these exposed latent spaces are often difficult to interpret through traditional musical terminology or interaction techniques.

The recently introduced AFTER architecture builds on RAVE by using adversarial criterion to differentiate structural and timbral information during the training process, which condition a latent diffusion model for audio generation (Demerlé et al., 2024). This results in three latent spaces exposable via *nn~*, offering a broader range of musical parameters to experiment with and suggesting rich sonic affordances. However, this structure also increases the difficulty of navigating latent parameters in an intentional way without an interface to aid in their interpretation.

Several recent projects explore novel means of navigating the latent spaces created by VAE models, including a sound walking approach by Scurto and Postel (2023), Semilla AI by Valenzuela (2023), and Stacco by Privato et al. (2024). Researchers have also developed approaches to latent space mapping via interactive machine learning (IML), in which users construct bespoke datasets and use these to define interactions. Vigliensoni and Fiebrink (2023) used a regression-based approach to map human performance parameters to high-dimensional latent spaces, while Zheng et al. (2024) introduced a toolkit for visualising latent distributions and defining trajectories through these via a screen-based interface.

However, despite this diversity of approaches, relatively few projects address embodied interaction with latent spaces, and it remains an open question how musicians can best harness their existing tacit and implicit musical knowledge when working with generative AI. These are important challenges, since music is a deeply embodied phenomenon, with well-established links between movement and multiple characteristics of musical structure and expression (Godøy, 2022; Iyer, 2002). The process of learning music is also strongly rooted in both auditory and embodied experiences (Mitchell et al., 2026), which suggests that interactions focused on these aspects of musical knowledge could improve latent space learnability and navigability. While such embodied and implicit musical knowledge is notoriously difficult to quantify, several decades of research and performance practice with wearable gesture sensors, including author 2's work with electromyography (muscle signal), offer a range of useful strategies for experimentation (Sonami, Laetitia, nd; Tanaka, 2014). Practice-based research methods are also an established means of documenting embodied, tacit, and intuitive knowledge, which can usefully inform bespoke interaction designs (Bulley and Sahin, 2021; Kirby, 2023).

Nabi et al. recently experimented with some of these approaches, applying movement sensors, interactive machine learning, and practice research methods to build embodied relationships with RAVE latent spaces through collaboration with a dancer, and demonstrating successful embodied latent mappings through live performance. The researchers published several Max patches to assist other artists in embodied latent space exploration and mapping tasks (Nabi et al., 2024). Author 2 also recently premiered a percussion piece using electromyography data from live performers to navigate the latent dimensions of several RAVE models (Tanaka, 2026). This prior work establishes the utility of sensor-based interaction for building and exploring embodied relationships with NAS latent spaces. However, more research is needed on workflows and interfaces that enable a broader group of musicians to harness these techniques and apply NAS latent spaces creatively. There is also a need for more user studies and qualitative data on how musicians interact with and experience neural audio synthesis, particularly in the context of embodied interaction. Finally, few existing studies use multimodal sensor data in this context, or apply embodied inputs to work with the more complex structures of models such as AFTER.

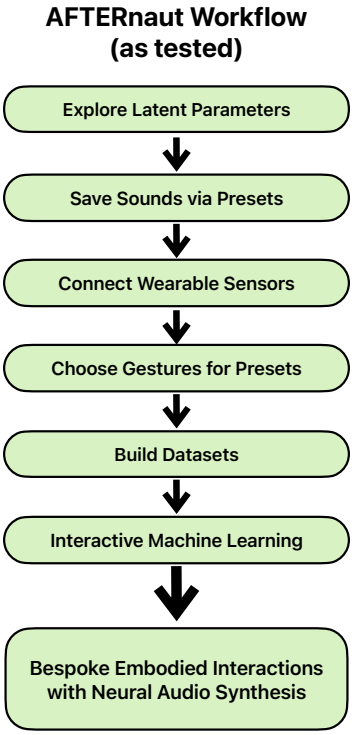

Figure 1: Chart displaying the primary workflow used during interface testing

3 METHODOLOGY

We designed a workflow which supports a multi-step latent space exploration and mapping process, where users can first investigate the latent space, then identify settings they like and define embodied associations with these. Users have access to multiple tools in one place, offering enough flexibility for highly personalised and longer-term creative applications of these models. By building and implementing this interface, we aimed to gain design feedback and practice-based insights towards systems which support individualised, embodied relationships with generative music models. The project was titled "AFTERnaut" as a play on the idea of deep (latent) space exploration: a tool for exploring unknown and "ungraspable" territory to encounter new knowledge.

Our development process was experimental and iterative, drawing on practice research methodologies (Bulley and Sahin, 2021) to design a system which supported intuitive, embodied, and tacit ways of knowing. This included a remutualised, movement-led design process (Cook, 2004; Kirby, 2023), in which our interactions with the evolving system further informed the design as part of a productive feedback loop. Our design was informed by prior interactive machine learning research (Fiebrink et al., 2009) and sound tracing methods (Godøy et al., 2016), and work using these to define specific gestural "anchor points" for interpolation in improvised performance (Tanaka et al., 2019). We applied these techniques in the context of divergent and exploratory approaches to generative AI described by Broad et al. (2021), Chemla-Romeu-Santos and Esling (2022), and others, focusing on the use of AFTER for novel synthesis rather than style transfer. Evaluation was twofold: through autoethnography (Rapp, 2018) and a small user study via one-on-one sessions with seven participants, offering insights into how the system functioned in practice. We offer a detailed account of both evaluation phases in section 5.

4 SYSTEM ARCHITECTURE

We took a modular approach to system design, placing several existing technologies in dialogue with each other through a single easy-to-navigate GUI. In addition to AFTER, these included a latent space exploration workflow from IRCAM (ACIDS, 2026), machine learning tools from the Fluid Corpus Manipulation library (Tremblay et al., 2021), and systems for capturing movement and muscle data from wearable sensors (Lough et al., 2018; Di Maggio et al., 2023). For this initial study, we chose to use a single AFTER model as we found we could get a significant amount of variation out of it, and because this made it easier to compare and contrast experiences of different users. We specifically chose the percussion.ts model due to our interest in the temporal or structural variations offered by AFTER (ACIDS, 2026). Our interface offers several modules, discussed below.

4.1 Manual Exploration

This section of the interface builds on IRCAM's *after.audio.mega testor* demonstration patch (ACIDS, 2026), featuring adjustable sliders and dials to either bias (add to) or scale (multiply) eight of the first eight exposed dimensions of both timbral and structural latent spaces. This also includes a bank of eight oscillators which can be used to continuously modulate these. Along with the Max "playlist" audio player and sample audio files included by IRCAM, we added expanded input options, including an external source, a drag-and-drop file browser, and an additional adjustable frequency oscillator sending either sine waves or square wave-generated pulses into the structural space of the model. We also added a two-dimensional "diffusion bias" interface, similar in appearance to IRCAM's Max4Live implementation (ACIDS, 2026), but instead allowing users to also skew select dimensions of model's final diffusion layer. Each of these parameters are exposed to the Max *pattr* system, allowing users to save and retrieve settings as presets, as discussed in section 4.3.

4.2 Gestural Inputs

This module receives multimodal sensor data inputs which can be used in several different ways within the patch. We specifically chose Inertial Measurement Units (IMU) which transmit descriptors of motion and relative position, and electromyography (EMG) sensors which quantify muscle tension and effort. For IMU, we used a glove-based Mugic unit (Lough et al., 2018) sending data via OSC over WiFi, making accelerometer (x y z), gyroscope (x y z), and quaternion (w x y z) measurements available for routing in the Max patch. For EMG input, we used the EAVI board (Di Donato, Balandino et al., 2019) sending up to four separate muscle measurements over MIDI-BLE to the BBDMI library in Max (Di Maggio et al., 2023), with a Bayesian filter available for data smoothing. While the EMG sensors we used can be placed anywhere, we generally used the forearm and bicep. Both types of sensor inputs receive additional smoothing via a running average over a 50ms window. Via a Max *crosspatch* object, any combination of sensor values can be routed to any of eight exposed latent dimensions in either the timbral or structural space. Users can also take snapshots of incoming sensor values for dataset construction and machine learning.

Figure 2: IMU and EMG sensors used for gesture data capture, including the Mugic unit and glove, the EAVI board, and attachable muscle sensors

4.3 Presets and Datasets

To support a multi-step exploration and mapping process, the interface includes both preset and dataset construction workflows. Through the Max *pattr* system, users can save combinations of settings which produce sounds they like and recall them using a numeric label. These labels can also be included in a three-part dataset structure constructed from FluCoMa *labelset* and *dataset* objects (Tremblay et al., 2021) along with Max *buffer* objects. These datasets consist of:

- Preset numbers as classification labels
- Snapshots of sensor data via an eight-point input vector (based on routings from the crosspatch object)
- Snapshots of latent biaser and scaler readings from both structure and timbre spaces as a 32-point vector.

Data points in all three data streams share common identifiers. Once captured, users can view these datasets using a Max *dict* object table, and both presets and datasets can be saved on the local hard drive as JSON files and recalled for future use. This system enables users to document

and revisit findings from their latent space explorations, while simultaneously capturing useful datasets for machine learning.

4.4 Interactive Machine Learning

The interface offers both classification and regression workflows for gesture mapping via machine learning, via FluCoMa Multi-Layer Perceptron (MLP) objects (Tremblay et al., 2021). Classification training uses parts one and two of the dataset, associating preset labels with sensor readings so that the trained model will output predicted presets for new sensor data inputs. Regression training pairs parts two and three of the dataset, linking sensor readings with latent variables. Trained regression models continuously output interpolated latent biaser and slider settings based on new sensor readings. Like presets and datasets, these models can be saved to disk as JSON files and recalled for future sessions.

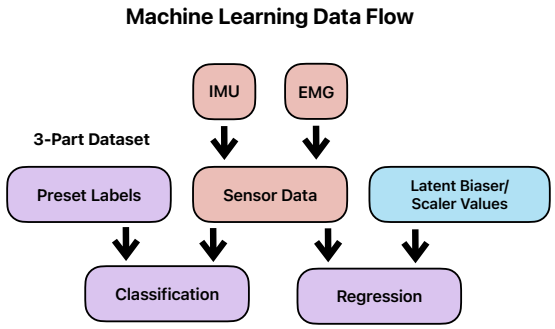

Figure 3: Three-part dataset structure and its application in interactive machine learning

4.5 GUI

Each of these modules is accessible via a single user interface in Max presentation mode, colour-coded by function. Tools for manual latent space exploration have a blue background. Primary inputs are arranged on the top right, with buttons to choose between sources. These are positioned above two banks of sliders (biasers) and dials (scalers) for both the structural and timbral latent spaces. Each bank of sliders and dials has a switch to choose between four input modes: manual adjustment, oscillator modulation, direct sensor mapping, or regression model predictions. The two-dimensional diffusion biaser is placed alongside these, along with a bank of modulators which can be set to either sine wave or square wave pulse and routed to the latent biasers and scalers. Sensor inputs are placed on the left side and have a red background. This area features toggles and switches to initialise IMU and EMG inputs and to smooth and calibrate EMG signals. There are also multisliders to visualise sensor input, and a *crosspatch* object for routing these signals. The center of the patch has a purple background and offers dataset and machine learning functions. The upper-center features interfaces for labeling and saving presets, and for building, managing, and viewing IML datasets. Below this are interfaces for training, running, loading, and saving both classification and regression models.

4.6 Workflow

Users activate the AFTER model in the upper left corner of the patch, then choose an audio input type. They can then adjust or modulate the biasers and scalers in each latent space to explore how these impact the resulting output. When users encounter sounds or patterns they like, they can save their settings as a preset by choosing a numerical label. To work with sensor data, they first attach and activate the sensors, then use the interface to receive, smooth, and calibrate the incoming data. They can then route this data to the latent dimensions of their choice and use these to adjust latent biasers and scalers in real time through movement or muscle tension. They can also build datasets by demonstrating gestural positions alongside preset labels or latent biaser and scaler settings and clicking a button. To train a model on these associations, they collect several examples, then click the training button several times until the displayed loss function is acceptable. They then click "run" and demonstrate their chosen gestural positions to obtain predictions. Classification models allow them to shift between selected presets using their movements, while regression models allow for smooth interpolation between selected latent biaser and scaler settings via gesture.

5 EVALUATION

During the development process the authors tested both the functionality and creative capacities of the system, taking notes on our observations and conducting formal testing in our own practice upon completion. The final version of the patch worked as intended, allowing us to explore the latent space, save presets and datasets, receive and route sensor data, and create reliably-performing gesture recognition models via classification and regression. User testing provided additional confirmation that the patch performed as designed and achieved our technical goals. Authoethnography and one-on-one sessions with individual participants gave us insights into how well AFTERnaut fulfilled its role as an exploratory or creative tool for building individualised and embodied interactions with neural audio synthesis.

5.1 Autoethnography

Author 1 conducted a 2-3 hour formal evaluation of the workflow through practice, using it to explore AFTER's latent space, identify outputs he found interesting, and apply them in personalised embodied mappings. He took detailed step-by-step notes on his choices and interactions with the interface, alongside subjective impressions and creative insights which emerged. He then reviewed these for common themes.

For this session, he worked with a continuous oscillator input to the structural latent space, generating variations in pattern and texture and mapping these to specific arm positions. Using a consistent signal input at a set frequency allowed for methodical exploration of the impacts of different combinations of latent biaser and scaler settings. He also applied a technique we encountered during development, using the additional oscillator bank to modulate the biasers and scalers of each latent space in multiples of the input frequency, creating a type of Euclidian sequencer by moving through latent dimensions at regular intervals.

Author 1 began by performing subtle manual adjustments of latent scalers and biasers with LFO pulse or wave inputs (2-8 Hz). He then switched to using the modulators on

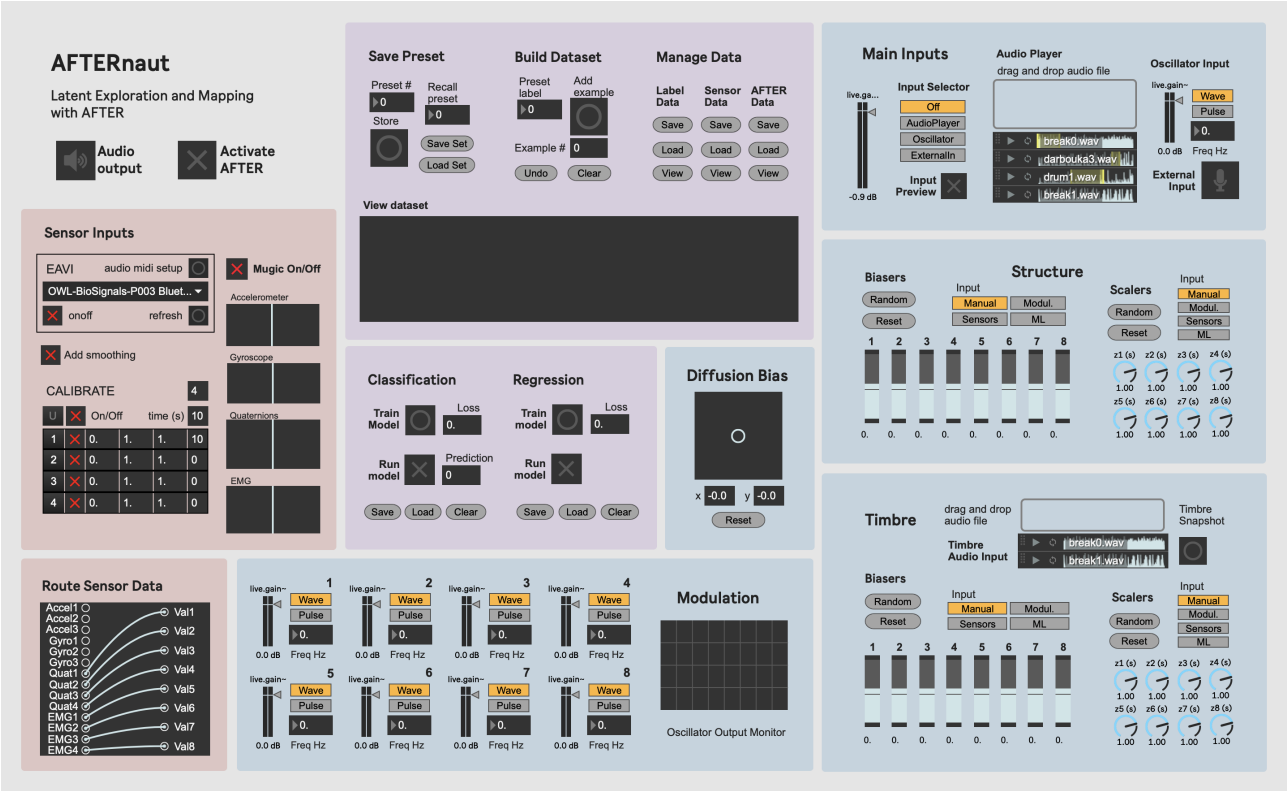

Figure 4: Screenshot of the user interface of the AFTERnaut Max patch

the latent space to generate compound rhythmic patterns, gradually testing higher input oscillator frequencies. Inputs of 16 or 32 Hz led to rapid, unpredictable pattern variations akin to high-BPM deconstructed club or breakcore. Higher input frequencies produced more textural results, with sine waves over 100 Hz eliciting consistent tones from the percussion model, and square wave pulses producing sounds akin to a percussion orchestra conducting free improvisation. Using modulators on the latent space with frequencies above 500 Hz obtained increasingly unpredictable and abstract results, sometimes moving rapidly between percussive sounds and pitched textural layers. Some areas of the latent space also produced much quieter textures at this frequency reminiscent of field recordings or streetscapes.

He concluded the test by returning to manual latent dimension adjustment and much slower inputs, using both 1 Hz and 0.25 Hz oscillator waves. While changing latent settings in the structural space didn't create specific temporal patterns, it appeared to impact the likelihood or density of rhythmic fills and ornaments, as well as timbral and dynamic variations, which was especially apparent at 1 Hz. This included adding subdivisions, flams, rolls, missed beats, and noise to the audio output. The slowest input frequencies had a breath-like effect, reaching silence at points in the cycle, then slowly pulling through the latent space in between. Disrupting these cycles via latent parameter changes produced variations ranging from peaceful and meditative to harsh, glitchy and alien.

Throughout this process, author 1 saved several dozen presets, ranging from fast rhythmic patterns to noise and ambient textures. Upon completing his explorations of the latent space, he returned to review these presets, choosing a group he wished to use for embodied mapping. He then attached sensors to his hand and arm, and experimented with different gestures while listening to the presets. He chose positions to associate with several, including variations in arm position, tilt, extension, and tension. He captured data for these through the interface, trained several classification and regression models, and tested them in live improvisation. While the audio latency of the current version of AFTER (2-3 seconds) posed challenges in assessing the outcome of faster movements, he was able to successfully navigate between his pre-chosen parameters through slower gestural shifts. He felt that this gave him an intuitive tool for directing changes in sonic outputs using his body, building a meaningful physical relationship with novel patterns and textures he had uncovered by exploring the latent space. Some results of this test are illustrated in the accompanying videos.

5.2 User Studies

We conducted one-on-one testing sessions with a total of seven individual participants. We recruited subjects from the extended community of music and computational arts students and researchers in London, UK. All had some type of creative practice and either musical experience, coding experience, or both. One was a trained dancer. Half had only cursory knowledge of music AI, the other half had detailed knowledge of the field. All participants were informed of the aims, methods, and funding of the project and gave written, informed consent for their participation. Data was collected and stored in compliance with GDPR regulations [*].

Testing sessions lasted 60 to 90 minutes. We explained the overall structure of the system to participants, then progressively introduced its features. The protocol comprised:

[*] https://gdpr-info.eu/

- Testing the system with audio player input, inviting users to make adjustments to individual biasers and scalers in structural and timbral latent spaces
- Using the oscillator input as an LFO (2 Hz pulse), to explore how latent biasers and scalers changed the sound
- Trying different oscillator frequencies, and introducing modulators and preset features. We invited them to experiment (free play) for 3 minutes, identifying and saving at least three presets based on sounds they found
- Adding sensor inputs with the Mugic glove IMU, and EAVI EMG electrodes on their forearm or a combination of forearm and bicep
- Free play with direct mapping of sensor inputs to latent biasers and scalers, combining these inputs with other settings
- IML dataset building, asking subjects to choose previously stored presets and arm positions to associate with each. Participants recorded at least three examples of each position for each preset class.
- Training both classification and regression models on their data, then free play to improvise with each.

We documented these sessions via notes, audio recording, and video, and concluded with an outgoing interview with each participant about their experiences and opinions about the system. We then analysed the results for emerging themes and specific insights.

5.2.1 Observations

All participants (P1-7) approached the project with curiosity, but with different strategies. Some took a methodical approach to testing each parameter to understand the system, while others took a "free play" approach, adjusting several parameters at a time and seeking unusual sounds. Most gravitated to the input oscillator rather than the audio player input once both had been demonstrated, with some stating they found the oscillator more interesting. The majority had an enthusiastic reaction to the sounds they were able to produce and were drawn into experimenting with the system for longer than expected. In most cases we needed to interrupt the session in order to stay to time. One described building several "sound worlds" (P2), one conducted an extended movement improvisation (P1), and several were drawn into looping and adjusting generated patterns for an extended period, suggesting strong creative engagement.

Each person took their own approach to exploring, in some cases identifying techniques we hadn't anticipated. P7 used a high frequency oscillator input (over 1k Hz) to the latent spaces and subdivided the output using the modulators. They also set either biasers and scalers to a preset they liked while using the sensor inputs to vary the parameters of the other, letting them improvise within different global settings. P6 identified several areas of the latent space which produced near-silence and used very slow LFOs to gradually vary it (less than 0.1 Hz), building a set of subtle, ambient soundscapes. P2 conducted deeper exploration of the impact of different sensor data routing options on latent biasers, and achieved movement-sound relationships with strong visual correlations. They also experimented with using the patch 'no input' (Mudd and Brown, 2023), scrubbing through the latent space with sensor values and sliders.

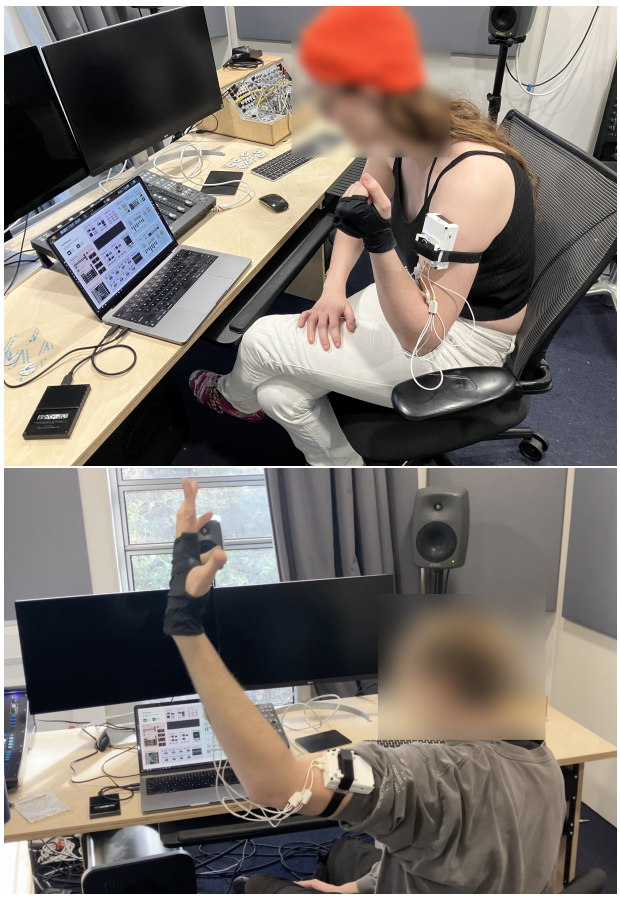

Figure 5: Participants testing the patch with wearable sensors

5.2.2 Interviews

In concluding interviews, all participants stated they felt they could get interesting sounds from the system. Some compared it to playing with a modular synthesiser for the first time, in that "you don't really know what the controls do, and you just tweak them until you have a sound you like" (P4). Many felt that the preset system was helpful in giving them a foothold. One user expressed skepticism over their own role in creating the sounds versus the model. For many, it was initially "difficult to figure out which dimension, which axes, created which change in the sound" (P2). However, most were drawn to exploring until they found elements they could work with, and the impacts of different latent scalers and biasers became clearer with practice.

Most participants were able to find embodied interactions which felt intuitive or useful to them, and felt the use of wearable sensors enhanced the experience of working with the patch. While P2 stated that direct sensor interaction with latent dimensions was helpful in exploration, the majority expressed a preference for machine learning mapping of sensor data to pre-discovered settings. Many felt this made it easier to perform with the system intentionally and reliably, and some therefore perceived this embodied mapping as the patch's primary purpose. P1 stated that "the embodied parameter really made me want to think of how it could be put into a performance somehow". However, it was still difficult for most to establish direct or precise control with the sensors, with several recognising that they would need more practice.

Most felt the interface was missing very little, with P3 stating "The patch was really cool, it had everything I would have wanted to play around with the model". The most interesting part of the interface for participants was the process of trying to explore and understand the latent space of the model, and working with the sensors, both of which were novel for many. Critiques included a desire to try the system with a different AFTER model, and the length of time it took to get sounds they wanted. Some also suggested small tweaks to the system, including the ability to scale the range of latent sliders in the GUI, or to selectively freeze some of them.

6 DISCUSSION

In our evaluations, AFTERnaut successfully helped both ourselves and study participants to explore the latent space and define individualised and embodied interactions with neural audio synthesis. Although participants still encountered some challenges with the interpretability of the latent space, all were able to identify sounds they found interesting and recall them to build gestural relationships with the sonic outputs of the system. Users achieved a wide range of sounds and mappings from a single percussion model through the interface and demonstrated a high level of creative engagement. Participants' reports of feeling more connected to the model's outputs via sensor-based IML support our longer-term goal of making the latent space more "graspable" or relatable via embodied interaction. In our personal testing, we found the system highly useful for purposeful and productive exploration of the latent space. Although the model's latency placed some constraints on improvisation, we also found the IML workflow valuable in helping us relate to and perform with the sounds we discovered.

In the bigger picture, building and testing this interface offered useful insights into systems which prioritise listening and moving in divergent musical applications of NAS. While the system we present is in its first iteration, it represents a productive step towards multifaceted interfaces and workflows which enable a wider population of musicians to develop personalised, embodied interactions with neural audio synthesis through practice. In our future work, we hope to expand on the machine learning workflow in the patch, including hands-free capture of data examples and training strategies which enable a wider variety of gesture types, including time-series approaches. We also intend to experiment with a broader range of AFTER models, and to train our own models, to explore the resulting differences in latent space behaviour. Finally, as the system evolves we aim to conduct further user studies with a wider group of musicians. This system will be available open-source at the time of publication[†], and we welcome other users to fork it and add additional features that meet their creative needs.

7 CONCLUSION

In this paper we describe the design and evaluation of a multimodal interactive interface which facilitates exploration, mapping, and bespoke embodied interaction with neural audio synthesis latent spaces. Through individual evaluation and testing with lay users, we identified numerous creative affordances of this approach and demonstrated its potential in developing individualised sounds and embodied mappings. The system makes the abstract nature of latent space navigation more intuitive and embodied for musicians seeking new sounds, and contributes to the development of interactive systems which help a broader population of creatives explore and define bespoke uses of generative music models via practice-centered methods.

[†]https://github.com/evvvvod/AFTERnaut

AUTHOR DECLARATIONS

The authors are not aware of any conflicts of interest related to this paper. All study participants gave fully informed consent and their data has been protected in compliance with GDPR regulations. No participants were exposed to any harmful risks during this work, and all responses have been pseudonymised. The authors used publicly licensed generative AI models in this work with proper attribution of their source. All interactive machine learning training on participants' data was done under their supervision and with their consent, and all sensor data from participants was deleted from our servers immediately after the session. Otherwise, the authors only conducted ML training on their own data.

REFERENCES

ACIDS (2026). acids-ircam/AFTER.

Bowman, W. D. (1982). Polanyi and Instructional Method in Music. *Journal of Aesthetic Education*, 16(2):75.

Broad, T., Berns, S., Colton, S., and Grierson, M. (2021). Active divergence with generative deep learning – a survey and taxonomy. In *In: Proceedings of the 12th International Conference on Computational Creativity.*

Bryan-Kinns, N., Zheng, S., Castro, F., Lewis, M., Chang, J.-R., Vigliensoni, G., Broad, T., Clemens, M. P., and Wilson, E. (2025). XAIxArts Manifesto: Explainable AI for the Arts. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8, Yokohama Japan. ACM.

Bulley, J. and Sahin, O. (2021). Practice Research - Report 1: What is practice research? and Report 2: How can practice research be shared? Technical report, British Library.

Caillon, A. and Chemla—Romeu-Santos, A. (2026). acids-ircam/nn_tilde.

Caillon, A. and Esling, P. (2021). Rave: A variational autoencoder for fast and high-quality neural audio synthesis.

Chemla-Romeu-Santos, A. and Esling, P. (2022). Creative divergent synthesis with generative models. In *In Proceedings of the 3rd AI Music Creativity Conference*.

Cook, P. R. (2004). Remutualizing the Musical Instrument: Co-Design of Synthesis Algorithms and Controllers. *Journal of New Music Research*, 33(3):315–320.

Demerlé, N., Esling, P., Doras, G., and Genova, D. (2024). Combining audio control and style transfer using latent diffusion. In *Proceedings of the 25th Int. Society for Music Information Retrieval Conference*. arXiv.

Di Donato, Balandino, Tanaka, Atau, Zbyszynski, M., and Klang, M. (2019). EAVI EMG board. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Di Maggio, F., Tanaka, A., Fierro, D., and Whitmarsh, S. (2023). An Interactive Modular System for Electrophysiological DMIs. In *Proceedings of the 18th International Audio Mostly Conference*, pages 79–84, Edinburgh United Kingdom. ACM.

Fiebrink, R., Trueman, D., and Cook, P. R. (2009). A Meta-Instrument for Interactive, On-the-fly Machine Learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 280–285.

Godøy, R. I. (2022). Thinking rhythm objects. *Frontiers in Psychology*, 13.

Godøy, R. I., Song, M., Nymoen, K., Haugen, M. R., and Jensenius, A. R. (2016). Exploring Sound-Motion Similarity in Musical Experience. *Journal of New Music Research*, 45(3):210–222.

Iyer, V. (2002). Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music. *Music Perception*, 19(3):387–414.

Kirby, J. (2023). Approaches for Working with the Body in the Design of Electronic Music Performance Systems. *Contemporary Music Review*, 42(3):304–318.

Leman, M. (2008). *Embodied music cognition and mediation technology*. MIT Press, Cambridge, Mass.

Lough, A., Micchelli, M., and Kimura, M. (2018). Gestural Envelopes: Aesthetic Considerations for Mapping Physical Gestures Using Wireless Motion Sensors. In *International Computer Music Conference Proceedings*, volume 2018. Michigan Publishing, University of Michigan Library.

Mitchell, H. F., Blom, D., and Long, P. (2026). 'We hear with our eyes!' Unlocking tacit knowledge about multisensory music performing. *International Journal of Music Education*, 44(1):174–187.

Mudd, T. and Brown, A. (2023). Contrasting approaches to the no-input mixer. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 402–408. Zenodo.

Nabi, S., Esling, P., Peeters, G., and Bevilacqua, F. (2024). Embodied exploration of deep latent spaces in interactive dance-music performance. In *Proceedings of the 9th International Conference on Movement and Computing*, pages 1–9, Utrecht Netherlands. ACM.

Nika, J., Schwarz, D., and Muller, A. (2025). Crafting Musical Agents: Interactive Generation as a Catalyst for Artistic Formalization. In *Proceedings of the 6th Conference on AI Music Creativity*. Zenodo.

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*, pages 424–431. Zenodo.

Rapp, A. (2018). Autoethnography in Human-Computer Interaction: Theory and Practice. In Filimowicz, M. and Tzankova, V., editors, *New Directions in Third Wave Human-Computer Interaction: Volume 2 - Methodologies*, pages 25–42. Springer International Publishing, Cham.

Scurto, H. and Postel, L. (2023). Soundwalking Deep Latent Spaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Zenodo.

Sonami, Laetitia (n.d.). lady's glove.

Tanaka, A. (2014). The Use of Electromyogram Signals (EMG) in Musical Performance: A Personal survey of two decades of practice.

Tanaka, A. (2026). Transfert for percussion trio, emg and rave.

Tanaka, A., Di Donato, B., Zbyszynski, M., and Roks, G. (2019). Designing Gestures for Continuous Sonic Interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Zenodo.

Tremblay, P. A., Roma, G., and Green, O. (2021). Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit. *Computer Music Journal*, 45(2):9–23.

Valenzuela, M. H. (2023). SEMILLA.AI STUDIO.

Vigliensoni, G. and Fiebrink, R. (2023). Steering latent audio models through interactive machine learning. In *Proceedings of the 14th International Conference on Computational Creativity*. Zenodo.

Zheng, S., Sette, B. M. D., Saitis, C., Xambó, A., and Bryan-Kinns, N. (2024). 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*, pages 591–597. Zenodo.