Eco-Sonic Interfaces for Embodied AI Sound Exploration

Sabina Hyoju Ahn
University of California, Santa Barbara
Santa Barbara, USA
sabina_ahn@ucsb.edu

Ryan Millett
University of California, Santa Barbara
Santa Barbara, USA
rmillett@ucsb.edu

Seyeon Park
Artist (collective_nshp)
Seoul, South Korea
seyparc@gmail.com

Figure 1: Neural Tides premiered at Project Area, Sónar+D, Barcelona, Spain, in 2024. Photo credit: Sónar Festival.

Abstract

Neural Tides is a neural network-based granular synthesizer that examines plastiglomerates—hybrid formations of plastic and organic material in marine environments. The system maps sound grains from oceanic field recordings to a navigable latent space using autoencoders and clustering techniques, controlled via hand gestures and touch. This interface physically connects performers with sonic representations of anthropogenic material transformations in coastal environments.

Keywords

machine learning, granular synthesis, NIME, soundscapes, plastiglomerate, neural audio synthesis

1 Introduction

Neural Tides transforms environmental phenomena of the Anthropocene into an interactive sonic experience. The project integrates machine learning systems into a physically musical instrument, where performers interact with environmental soundscapes through gesture and touch. The instrument enables embodied interaction with environmental soundscapes, using neural audio mapping to convert marine debris recordings into spatialized, real-time synthesis. It invites performers to physically engage with the entanglement of synthetic and organic materials characteristic of the Anthropocene.

The project evolved from the collaborative soundscape piece *Wind Traces* [3], which involved field recording and audio analysis on isolated islands in South Korea. These environments, heavily impacted by marine debris, offered sonic and material insight into plastiglomerates—geological hybrids of plastic fused with sand, shells, and other organic matter [5, 12]. These artifacts of the Anthropocene inspired the development of a performance system capable of expressing material entanglement sonically rather than visually.

Neural Tides does not generate sound autonomously. Instead, performers navigate a latent audio space organized by spectral features of environmental recordings. The custom-built, sensor-driven interface allows performers to manipulate sonic textures in real time, mirroring the physical processes of breakdown and recombination seen in plastiglomerates.

This work contributes to AI-based sound synthesis, physical computing, and ecological media art by embedding environmental data in a tactile, auditory instrument. The physical interface

NIME '25, June 24–27, 2025, Canberra, Australia

makes Anthropocene transformations directly accessible through embodied user interaction.

Figure 2: Plastiglomerate objects found in Hakrim-do. (Photograph by Seyeon Park)

2 From Visual Phenomena to Sonic Experience: Plastiglomerates in Coastal Areas

The conceptual and sonic material of *Neural Tides* originates from field recordings collected on two South Korean islands in 2022. These locations were chosen for their distinct acoustic ecologies shaped by varying degrees of human and environmental interaction.

On *Hakrim-do*, a sparsely populated island, the recordings captured subtle environmental processes: the movement of waves, wind patterns, and the sonic residue of synthetic materials gradually breaking down. In contrast, *Ulleung-do*'s busier coastline featured a more dynamic and layered soundscape, combining industrial activity—like ship engines and construction sounds—with natural elements. These contrasting acoustic signatures provided the foundation for the instrument's exploration of synthetic-organic hybridization [4].

Approximately nine hours of field recordings were captured, of which four hours were selected for their spectral richness and compositional diversity. Departing from the compositional focus of previous works such as *Wind Traces* [3], Neural Tides engages raw, unfiltered field recordings, transforming them into performable material. The plastiglomerate—a hybrid of synthetic and organic debris—serves as both sonic subject and conceptual metaphor for this recombination (see Figure 2).

Unlike traditional ecological soundscape pieces [8, 15] that document environmental conditions, *Neural Tides* actively transforms field recordings. Just as these hybrid materials physically merge synthetic and natural elements, the instrument recombines and reshapes corresponding sonic elements.

3 Sound Analysis: AI-Driven Sonic Mapping

This section details the integration of audio analysis and AI components into the physical instrument. We transformed spectral features of environmental recordings into parameters for tactile and gestural control, creating direct connections between plastiglomerate-affected environments and performance actions.

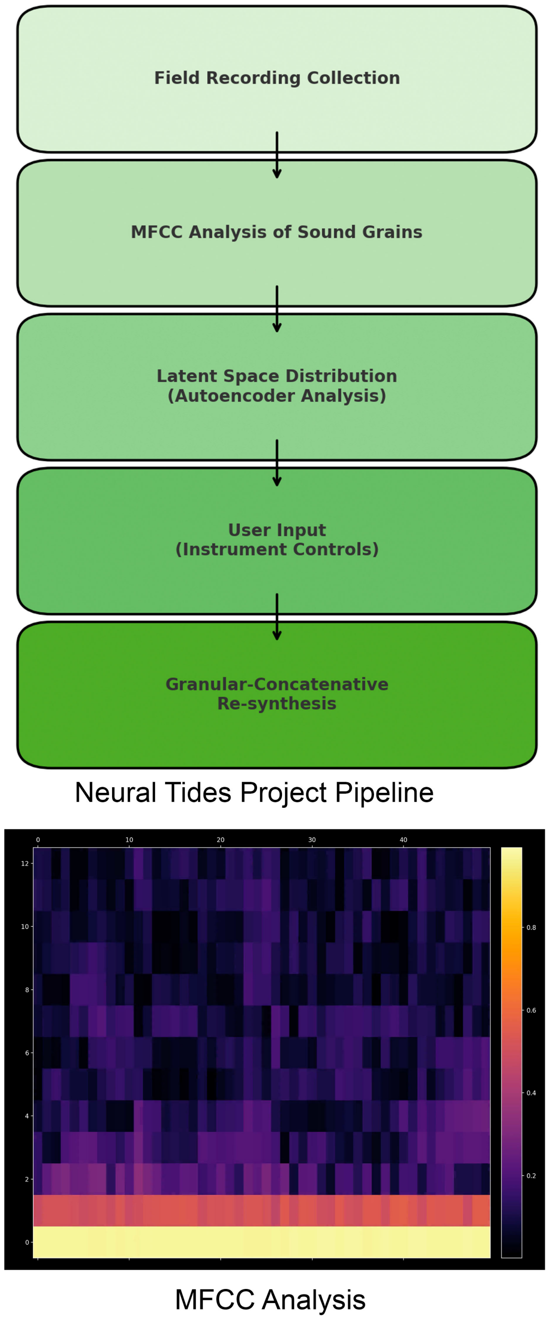

Figure 3: Neural Tides Project pipeline (top) and MFCC Analysis (bottom): MFCC coefficient matrix where the x-axis represents time frames and the y-axis shows the 13 mel-frequency cepstral coefficients. Higher intensity values (brighter colors) indicate stronger presence of specific spectral components.

3.1 MFCC Analysis and Granular-Concatenative Re-synthesis in Latent Space

The sound analysis process begins with segmenting field recordings into short audio grains, ranging from 10 to 50 milliseconds. These grains isolate microtemporal features that allow for highly flexible recombination and detailed control during synthesis.

Each grain is analyzed using 13-dimensional Mel-Frequency Cepstral Coefficients (MFCCs), a compact spectral representation grounded in human auditory perception [9]. MFCCs were selected for their ability to capture the subtle timbral variations found in coastal environments affected by plastiglomerates.

The resulting MFCC vectors are embedded into a two-dimensional latent space using an autoencoder neural network [6]. This transformation preserves perceptual relationships between grains, producing a sound topology where spatial proximity corresponds to timbral similarity. To delineate regions of the latent space, k-means clustering is applied to groups of grains into coherent sonic zones, while a KD-tree enables efficient nearest-neighbor searches during performance [14].

This latent space visualization is rendered in real time using `three.js` [1], providing performers with a dynamic visual interface for exploring sound materials. Navigation through the space becomes a compositional act by combining and reshaping fragments. Poetically, this reflects the entanglement of anthropogenic and organic sources.

3.2 Gesture-to-Sound Mapping and Software Implementation

Performers navigate this space through a combination of Leap Motion and Trill touch sensors [2]. The Leap Motion tracks hand gestures—position, orientation, and velocity—producing a 47-dimensional feature vector per frame (see Figure 4). A supervised multilayer perceptron (MLP) interprets these features—compared to ground truth gestures in the training set—and outputs an 8-dimensional control vector. Based on Fiebrink and Cook's approach [7], the MLP creates smooth transitions between gesture states, giving performers precise, non-linear control over sound selection and modulation.

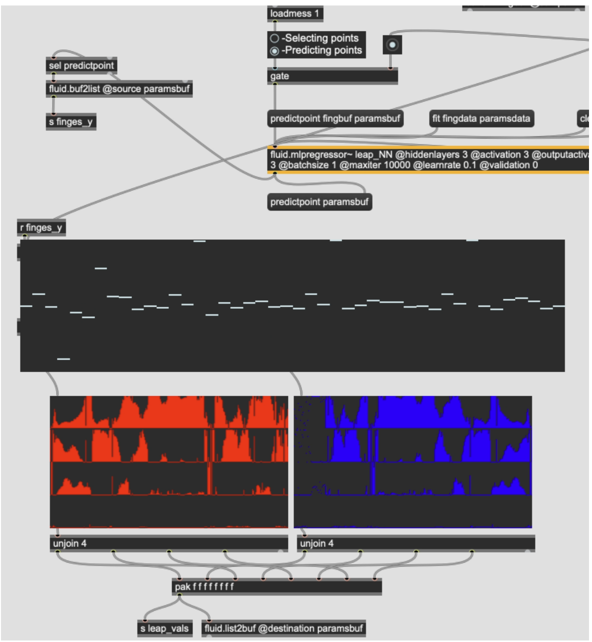

Figure 4: 47-dimensional feature vector representing gestural-tracking information routed to a neural network that maps hand gestures to granular synthesis parameters.

The Max/MSP environment handled real-time gesture processing and OSC routing to SuperCollider, where concatenative re-synthesis was executed using the FluCoMa toolkit [13]. The granular synthesis engine draws upon Curtis Roads' foundational work in granular synthesis and microsound [11], implementing per-grain control over pitch, density, window shape, and spatialization. This allowed the performer to engage with sound at a microscopic temporal resolution, enabling transformations that mirror the fragmentation and reformation processes observed in plastiglomerates.

The system architecture integrates hardware and software with low-latency communication. Rather than functioning as a black box, the machine learning components form an integral part of the performance feedback loop, directly connecting physical gestures to sonic transformations that mirror the material processes found in coastal environments.

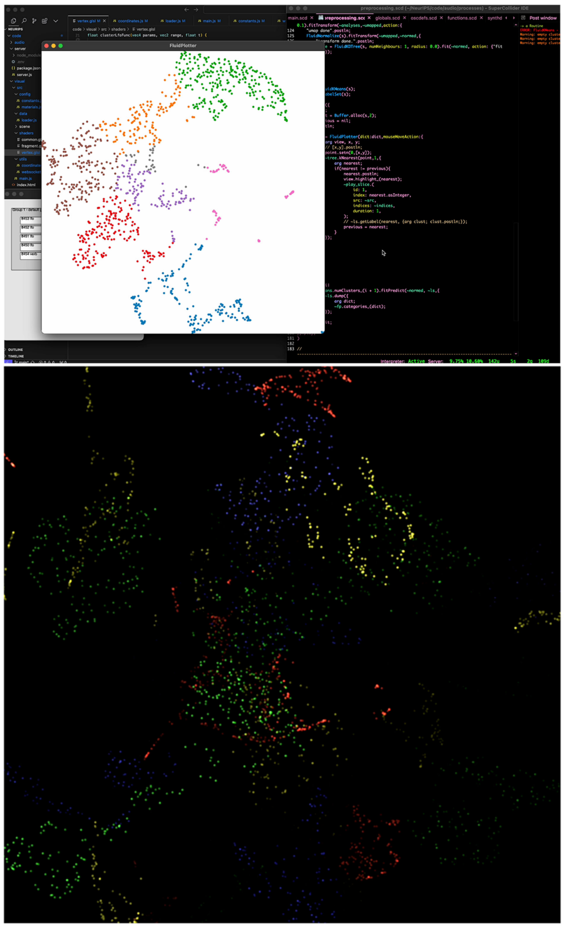

Figure 5: Two-dimensional projection of the learned latent space showing the distribution of sound grains. Colors indicate different sound categories.

4 Interaction Design: Connecting AI Systems with Instrument Design

The design of the physical interface for *Neural Tides* is foundational to its conceptual and functional aims. It embodies the project's core principle of translating ecological processes into sonic interaction by embedding AI into an expressive hardware system. This section details how custom hardware components, sensor integration, real-time software communication, and sustainable design converge to form a unified performance instrument.

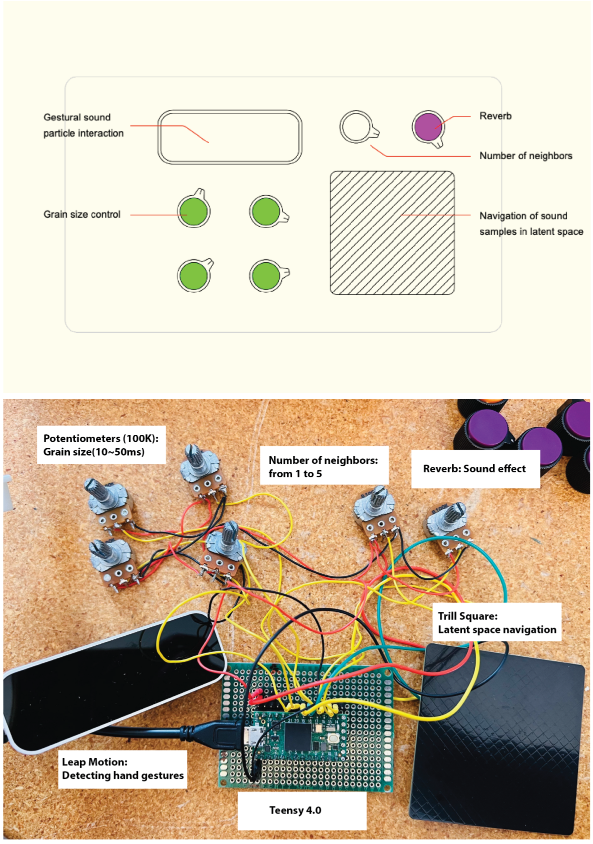

Figure 6: Neural Tides interface layout (top) and components (bottom)

4.1 Hardware and Software System Integration

The instrument centers around the Teensy 4.0 microcontroller, chosen for its processing speed and multimedia compatibility [10]. It manages input from custom sensors and controls, enabling expressive, multimodal interaction. Six potentiometers provide continuous control over synthesis parameters such as grain density, pitch, grain size, reverb, and modulation depth (see figure 6).

Touch interaction is facilitated by the Trill Square Sensor from Bela [2], a high-resolution capacitive surface that detects position, pressure, and multitouch gestures, allowing intuitive navigation of the latent sound space. In parallel, the Leap Motion controller captures 3D hand movement, offering mid-air control over parameters such as synthesis layer blending and spatialization (see figure 1).

These interfaces support simultaneous manipulation across multiple synthesis dimensions—for example, adjusting reverb and grain size with knobs while selecting timbral layers via hand gestures.

Sensor data is processed through a trained MLP within Max/MSP and routed to SuperCollider's granular synthesis engine. As detailed in Section 3.2, this system enables low-latency integration between gestural input and timbral modulation, forming a seamless bridge between physical control and sonic output.

Real-time visual feedback is rendered in `three.js` [1], reinforcing the spatial metaphor of the interface and aiding performers in navigating the latent space. Together, these components form a unified, embodied system where machine learning, gesture, and ecological concept merge in real-time performance.

4.2 Meaningful Embodiment and Sustainability

Neural Tides repositions machine learning as an embodied, performative process rather than a distant computational abstraction. The neural network is directly embedded into a tactile interface, enabling real-time sonic feedback in response to nuanced physical gestures. This tight coupling between gesture and sound expands the expressive potential of AI in live performance.

The instrument's material construction reinforces its conceptual grounding. Fabricated entirely from biodegradable algae-based filament, the enclosure reflects the project's ecological focus not only symbolically but physically. This alignment between form and theme ensures that sustainability is not merely referenced, but actively enacted.

Designed in Autodesk Fusion 360 and 3D-printed with algae bioplastic (see Figure 7), the enclosure integrates all sensing components within a unified, portable structure. The use of sustainable materials, coupled with embodied interaction, creates a coherent system where environmental data, performative control, and material ethics converge.

Figure 7: 3D-Printed Enclosure with Algae-Based Filament.

5 Discussion and Conclusion

Neural Tides explores how machine learning can be transformed from a disembodied computational tool into a tactile, performative instrument that enables embodied engagement with ecological phenomena. Rather than functioning as a black-box system, the neural architecture is materially and sonically present—translating the spectral traces of marine debris into an interactive sound experience.

By embedding autoencoded field recordings into a latent space and allowing performers to navigate it via gestural and touch input, the instrument invites real-time exploration of environmental fragmentation and fusion. These interactions mirror the physical characteristics of plastiglomerates, echoing how synthetic and organic elements merge and mutate in the Anthropocene.

Across exhibitions at a music festival, AI conferences, the system has drawn interest from artists, musicians, and researchers alike. Audience feedback frequently highlighted the unique immediacy of the system—its ability to make invisible data perceptible through gesture, and to translate environmental complexity into intuitive sonic improvisation. Many noted that the system blurred boundaries between listening, performing, and reflecting.

Looking forward, the project will explore generative neural synthesis models that create new grains based on the learned sonic space, expanding the instrument from one of navigation to co-creation. Additionally, a transition to three-dimensional latent space interaction will support greater expressive resolution, especially when paired with enhanced spatial gesture tracking.

Neural Tides suggests a future for sonic instruments grounded not just in technological innovation, but in environmental and perceptual accountability. As sonic interfaces become more intelligent, they can also become more empathetic—serving as bridges between data and the body, landscape and sound, abstraction and action.

6 Ethical Standards

This research adheres to the ethical guidelines established by the NIME community. Field recordings were conducted with minimal environmental impact, and the instrument's design incorporates biodegradable materials to promote sustainability. No human participants were involved, and all AI models were trained on ethically sourced, non-sensitive data.

7 Acknowledgments

This work was supported by the Ministry of Culture, Sports and Tourism, the Arts Management Support Center, and the Arts Korea Lab. We thank the Media Arts and Technology Program at the University of California, Santa Barbara, as well as Curtis Roads, George Legardy, and JoAnn Kuchera-Morin for their support and contributions.

For video documentation of *Neural Tides*, visit https://youtu.be/GX4YHeGIreY?si=Z21Q8cUksUUqUcbI.

References

[1] [n. d.]. *three.js*. https://threejs.org/
[2] Bela. 2025. Trill Square Touch Sensor. https://bela.io/products/trill/
[3] collective_nshp. 2022. *Plastic Snow, Wind Traces*. https://a-nam-san-hot-pot.vev.site/plastic-snow-2023_kr/
[4] collective_nshp. 2022. *sound samples*. https://drive.google.com/drive/folders/1Y5ga4j99h6s7i1fhMfqzSoc5dL3dTAhJ?usp=sharing
[5] Patricia L. Corcoran, Charles J. Moore, and Kelly Jazvac. 2014. An Anthropogenic Marker Horizon in the Future Rock Record. *GSA Today* 24, 6 (2014), 4–8. https://doi.org/10.1130/GSAT-G198A.1
[6] Jesse Engel, Cinjon Resnick, Adam Roberts, Sander Dieleman, Mohammad Norouzi, Douglas Eck, and Karen Simonyan. 2017. Neural audio synthesis of musical notes with WaveNet autoencoders. In *Proceedings of the 34th International Conference on Machine Learning*. 1068–1077. https://proceedings.mlr.press/v70/engel17a.html
[7] Rebecca Fiebrink and Perry R. Cook. 2011. The Wekinator: A System for Real-time, Interactive Machine Learning in Music. In *Proceedings of The International Society for Music Information Retrieval Conference*. 425–430.
[8] Annea Lockwood. 1989. *A Sound Map of the Hudson River*. https://www.lovely.com/titles/cd2081.html Audio recording.
[9] Geoffroy Peeters, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. 2011. The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America* 130, 5 (2011), 2902–2916. https://doi.org/10.1121/1.3642604
[10] PJRC. 2025. Teensy 4.0 Pinout. https://www.pjrc.com/store/teensy40_pins.html
[11] Curtis Roads. 2001. *Microsound*. MIT Press, Cambridge, MA. https://mitpress.mit.edu/9780262681544/microsound/
[12] Kirsty Robertson. 2016. Plastiglomerate. *e-flux journal* 78 (2016). https://www.e-flux.com/journal/78/82878/plastiglomerate/
[13] Pierre Alexandre Tremblay, Owen Green, Gerard Roma, and Alex Harker. 2019. Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation. In *Proceedings of the International Computer Music Conference*. 238–243.
[14] Pierre Alexandre Tremblay, Gerard Roma, and Owen Green. 2021. Enabling programmatic data mining as musicking: The Fluid Corpus Manipulation Toolkit. *Computer Music Journal* 45, 2 (2021), 9–23. https://doi.org/10.1162/comj_a_00600
[15] Jana Winderen. 2022. *The Art of Listening: Under Water*. https://arts.columbia.edu/art-listening-under-water Immersive sound installation.