

Latent Sampling: a PLAUD Performance

Błażej Kotowski*

Music Technology Group
Universitat Pompeu Fabra
Barcelona
blazej.kotowski@upf.edu

Abstract

This performance presents a live set built around PLAUD, a DDSP architecture which encodes corpus timbres into a variational latent space and learns its temporal character via transformer-based autoregressive priors. The performance is built around the use of multiple specialized models trained on small, curated corpora. Trajectory sampling is a central interaction mode that translates familiar audio sampling operations into the latent domain: control paths from manual modulation or the prior can be looped, sliced, reversed, and replayed at variable speeds. Differently from audio sampling, due to the temporal context, the synthesised sound depends not only on the punctual playback position, but also on the broader playback conditions, including the preceding control trajectory and playback speed. Bending operations further intervene via component limiting and waveshaping, producing out-of-distribution behaviour. The performance attempts to expand the gestural repertoire of audio sampling by leveraging the material affordances of the neural synthesis.

1 CONTEXT

When I prepare a performance with PLAUD, the first act is not opening up the software but selecting recordings. Thirty minutes of a modular jam I recorded years ago, metal impact sounds from a field recording session, a selection of tracks I selected for the last mix I made. These collections are small and personal, and each of them becomes a different instrument, with its own timbral vocabulary, temporal tendencies, and reactive character. In this sense, the practice begins with the corpus.

In small-data creative ML, training data functions less as ground truth than as a vehicle for artistic intention. Vigliensoni and Fiebrink (2024) frame data curation as an expressive act that enables sustained creative practice with machine learning. Scurto et al. (2021) identify small data and shallow models as conditions for meaningful material engagement, proposing that creative ML gains specificity from the constraints of limited, personally assembled collections. A model trained on my own materials learns a narrow, specific sonic territory—one shaped not only by my style but, importantly, the architecture that interpreted it.

The relationship between performer and corpus recalls sampling, the practice of selecting, transforming, and re-contextualising existing recordings. Cutler (1994) argues that sound recording made all recorded music available as compositional material. A PLAUD model trained on a personal corpus begins from this condition: the recordings function as raw material, and the model’s output remains inseparable from them. Yet the model does not preserve these materials as recognisable fragments. Through the architecture of the network, the corpus is dissolved into distributions, tendencies, and constraints. In this sense, the system behaves less like a sampler than like the kinds of generative systems described by Fell (2022), where the behaviour of the apparatus cannot be reduced to the intentions or gestures of the performer alone. What persists from the

corpus is not the recording itself but a conditioned field of possible sonic behaviours. The latent space becomes a kind of navigable environment, analogous to what Schafer (1993) terms a soundscape, whose internal structure determines what may emerge within it. Navigating this space is therefore not selecting from a collection of sounds, but moving through a territory shaped by the recordings from which the model was trained. In my performance, this navigation is enacted through a set of interaction techniques that translate familiar audio sampling operations into the latent domain.

Recent accounts frame latent spaces as domains to be explored through strategies derived from sampling practice (Noel-Hirst et al., 2025). PLAUD’s trajectory sampling module builds on this framing. The module records latent control trajectories—whether manually produced or generated by the prior—into a buffer, where they can be looped, sliced, reversed, and played back at variable speeds. The interaction vocabulary is borrowed from audio sampling: loop points, playback direction, grid quantisation. But the material being sampled is latent control signal, and what plays back is path-dependent: the synthesis chain responds differently to the same control input depending on the temporal context leading up to it. The same trajectory through a different model, or at a different speed, yields different sonic results, because of this context-dependence.

This is where a question of agency surfaces. Alongside the performer’s intentions, the model introduces some agency through temporal patterns and timbral tendencies it has generalised from the training data. Performing with PLAUD is a negotiation between these two: I set a playback position, but the GRU has its own momentum; I determine a trajectory, but the decoder produces something the recorded trajectory alone does not determine. In this sense, the system’s behavior is shaped by the material properties of the architecture itself, which introduce tendencies that are not

*Website of the author: <https://blazejkotowski.com>

reducible to either performer input or training data alone (Fell, 2022).¹



Figure 1: Performing with PLAUD in a live club setting. The setup includes a laptop running Ableton Live with the PLAUD Max for Live device, alongside an external Push controller.

2 METHODS

PLAUD is a neural synthesizer and Max for Live instrument built on NoiseBandNet (Barahona-Ríos and Collins, 2024), a DDSP architecture replacing harmonic oscillators with a filterbank of narrow noise bands. The synthesis model predicts per-frame amplitudes driven by a variational latent space. Training combines multi-scale spectral reconstruction with adversarial losses, and a latent smoothing strategy pushes the decoder’s GRU to internalise short-term temporal structure. An optional transformer-based autoregressive prior learns longer-range temporal patterns from the corpus. The architecture is detailed in a companion paper submitted to this conference.²

The performance setup consists of a laptop running Ableton Live with the PLAUD Max for Live device, controlled via Push controller. Several models are prepared in advance, each trained on a distinct personal corpus. The set follows an open score: models provide stylistic anchors, while the temporal evolution is improvised. The performance revolves around the mangling of the synthesis control trajectories. These are generated through manual modulation, as well as by the prior network, and dynamically modified through a number of sampling operations.³

Two bending operations intervene in the synthesis chain. Component limiting reduces active noise bands, producing spectral discontinuities and clicks. Waveshaping substitutes noise bands with frequency-matched sinusoids shaped toward square waves.

AUTHOR DECLARATIONS

Claude has been used for language editing to improve grammar and clarity. All the content, including the re-

¹To me, this condition is not unique to neural synthesis. Using field recordings as a raw material involves surrendering the agency partially to the environment within which the recording was taken. The neural synthesis makes this shared agency only more explicit.

²Companion video demonstrating the features of PLAUD is available at <https://plaudaimc2026.github.io/>.

³A recording of a representative performance is available at <https://youtu.be/q5MtkY-EWZA?si=ND0UTjQcBp20VlgB&t=1198>.

search, interpretations, and conclusions was developed by the author.

REFERENCES

- Barahona-Ríos, A. and Collins, T. (2024). Noisebandnet: Controllable time-varying neural synthesis of sound effects using filterbanks. *IEEE ACM Trans. Audio Speech Lang. Process.*, 32:1573–1585.
- Cutler, C. (1994). Plunderphonia. *Musicworks*, (60).
- Fell, M. (2022). *Structure and synthesis: the anatomy of practice*. MIT Press.
- Noel-Hirst, A., Saitis, C., and Bryan-Kinns, N. (2025). Sampling the latent space: Exploring the creative potential of generative AI through the lens of sample-based music making. In *Proceedings of the 6th Conference on AI Music Creativity (AIMC 2025)*, Brussels.
- Schafer, R. M. (1993). *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster.
- Scurto, H., Caramiaux, B., and Bevilacqua, F. (2021). Prototyping machine learning through diffractive art practice. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference, DIS ’21*, page 2013–2025, New York, NY, USA. Association for Computing Machinery.
- Vigliensoni, G. and Fiebrink, R. (2024). Data- and interaction-driven approaches for sustained musical practices with machine learning. *Journal of New Music Research*, 53(1-2):19–32.

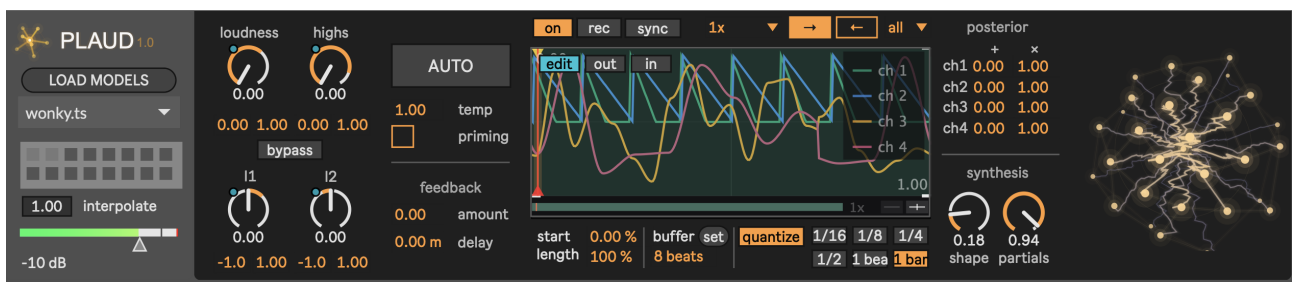


Figure 2: The PLAUD Max for Live interface used in performance. From left: model loading, presets and interpolation; loudness, highs, and latent control knobs; autoregressive prior controls; trajectory sampling module; posterior trajectory modifiers and synthesis bending parameters; bending visualizer.