

Latent Jamming in Reykjavik Sunburn: Electronic Music Improvisation Employing Generative Neural Audio Synthesis

Martin Heinze¹
Independent Artist
aimc2026@martstil.de

Abstract

Latent Jamming is an improvisation practice with neural audio synthesizers that adopts concepts from algorithmic composition and generative music. As an example, I present *Reykjavik Sunburn*, the most recent setup in an ongoing practical research effort dedicated to exploring musical qualities in working with generative neural nets for audio, conceived both as hybrid instruments and as (semi-)autonomous actors. In *Reykjavik Sunburn*, I act inside the latent space of four self-trained models, steering mood, density, and rhythm by generating, modulating, and organizing synthetic signal data streams that resemble latent embeddings. By doing so, I replace deterministic composition with guided exploration: tweak, listen, stabilize, vary.

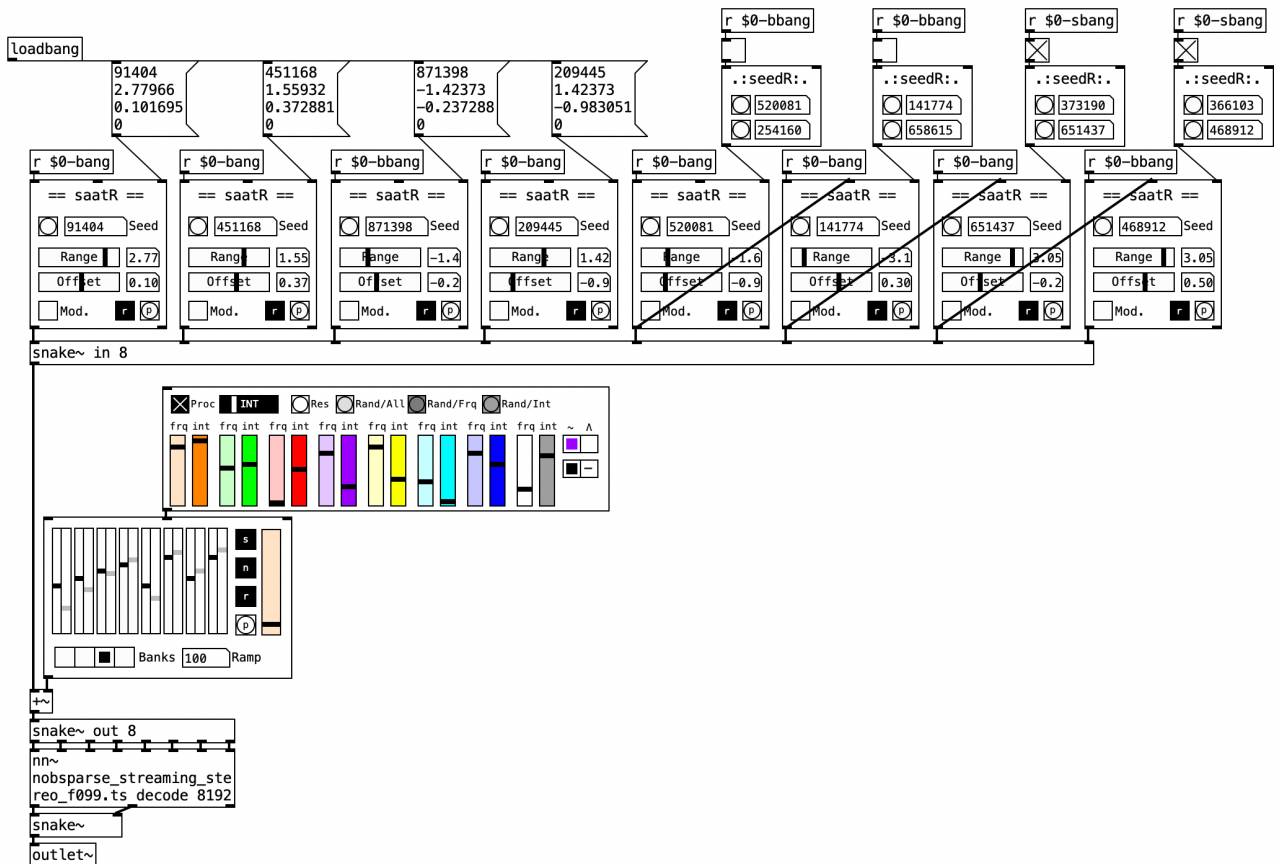


Figure 1: Exemplary interface for *Latent Jamming in Reykjavik Sunburn*

¹ <https://martstil.de>

1 CONTEXT AND CREATIVE POSITIONING

1.1 Training custom neural audio models

Latent Jamming positions itself in context with artistic practices in the "AI-Music" paradigm that explicitly include dataset curation and custom model training as part of a creative decision-making process (Jourdan et al., 2026; Tahiroğlu et al., 2024).

Preselection and categorization of music and audio data is considered a first creative act in the process, where material with a particular sonic character (e.g. sparse, dense, energetic, calm), attributed to a particular music genre, or such from a particular working phase or dedicated output selection (e.g. an album), is first separated into various data subsets and augmented in preparation for training.

Using architectures RAVE (Caillon et al., 2021), vschaos2 (Chemla—Romeu-Santos, 2020), or AFTER (Demerlé et al., 2024), among others, I train (variational) autoencoders on these curated selections with the aim of reproducing certain characteristics but even more importantly generating sounds that have not been part of the original dataset.

Observing models in action, the verdict is often that behavior and output can vary significantly, even when initial training conditions are similar or identical. This has contributed to a growing understanding of models possessing their own creative agency, while humans employing them in their work must share theirs (Xambó et al., 2024; Vear et al., 2023). Therefore, the models can be considered hybrids of instruments that embody the artist's own aesthetics and sound-machines that partly act autonomously.

1.2. Latent space as entry point

In a composition-performance continuum, *Latent Jamming* follows an approach that embraces this notion of both model-as-instrument and (semi-)autonomous actor. Human-machine interaction and mutual influence are established and negotiated inside the models' latent space, where low-level representations of the original domain data are accessible as latent embeddings or "encodings" for creative practitioners to engage with (Tahiroğlu et al., 2024; Yee-King, 2022)

In the past years, practitioners have found various ways of acting from within latent space, all with the intent of establishing a control layer (often including an interface) that aims for reproducibility and meaningful action-reaction patterns with regards to the models' output. Usually, these techniques are titled in association with the terminology of space, e.g. "navigation" (Tahiroğlu et al., 2024), "mapping" (Zheng et al., 2026), or

"exploration" (e.g. Liu et al., 2026; Tatar et al., 2023). Other approaches in latent space interaction make use of solutions like dimensionality reduction (e.g. Horta et al., 2025; Tahiroğlu et al., 2021) or aligning encodings of non-audio-domain control data along with latent embeddings of audio data (Tahiroğlu et al., 2026; Zheng et al., 2024).

Since in practice, each model requires statistical or empirical observation post-training as stated above, any composition-performance setup and its control interface can at most be a boilerplate template for techniques that have generally proven useful in similar cases, while putting it into action often resembles learning an instrument from scratch.

1.3. Learning an instrument in real-time

Latent Jamming builds on this instrument metaphor by using autoencoders' decoder as a neural audio synthesizer. "Tuning" the instrument happens as an iterative process of generating arbitrary synthetic latent data, observing the decoded outcome in the audio domain followed by changing or stabilizing the initial input until an aesthetically convincing, yet more or less unstructured result has surfaced.

While this first step mainly works as initialization and exploration of the models' general capabilities, a second step aims to structure the output by applying action patterns, modulation sequences, or simple repetition to synthetic data. In this context, the composition-performance practice becomes relevant, and the influence of human and artificial agents in a co-creative process is negotiated.

For *Latent Jamming*, I use Pure Data (PD) where real-time neural audio synthesis is made possible through nn~.² This object exposes the number of latent dimensions a trained model has been exported with and allows observing and modulating latent trajectories created through its encoder, or injecting similar information into its decoder.

In PD, I programmed a set of custom abstractions that are tailored for nn~ compatible model types³ (Figure 1), serving as both a control interface for "playing" neural audio synthesizers as instruments and define the range within which the models can generate sound autonomously.

These abstractions can spawn and control signal streams that mimic latent embeddings and their trajectories while allowing to preserve constellations that have led to convincing results in the music-making process. Technically, they operate within a spectrum of random sampling from pre-defined distribution types to manual selection of data points both within a specified range.

² https://github.com/acids-ircam/nn_tilde

³ <https://codeberg.org/martstilde/PD-latent-jamming> or <https://github.com/devstermarts/PD-latent-jamming>

2 REYKJAVÍK SUNBURN

Reykjavík Sunburn is the most recent in a series of *Latent Jamming* setups built on the aforementioned considerations and techniques.⁴ Four different, self-trained neural audio synthesizers are used: two RAVE and two vschaos2 models:

- **Black Latents**: a RAVE V2 model trained on the *Black Plastics* series.⁵ The dataset includes 3 hours of drum- and percussion-heavy electronic music. The resulting model generates mainly percussive output with rough textures and high grittiness. In *Reykjavík Sunburn*, this model is used as a leading asset to generate the rhythmic baseline and general percussive structure.
- **Nobspare**: a RAVE V2 model trained on a hybrid dataset of Tech House and sonically sparse Drum & Bass (about 4 hours of audio material). The model's characteristics are clear, sterile, and lightweight sounds, harmonic textures, and an isolated but dominant low end. Depending on the process development during the improvisation session, this model serves as a secondary texture generator but can also replace **Black Latents**' role in the framework.
- **VSC2 Nobspare**: this vschaos2 model has been trained on the same dataset as the **Nobspare** RAVE model. In *Reykjavík Sunburn*, it is used to generate interchanging pads and noise textures for transitions or simply to enrich the composition-performance with a harmonic layer.
- **VSC2 Martha2023**: being the only model trained on voice data (1,5 hours), this vschaos2 model adds a layer of rhythmic, pseudo-vocal sound on top of the otherwise "instrumental" generations of the three other models.

Together, these four models are responsible for all of the audio information created in *Reykjavík Sunburn*. No additional synthesizing techniques or sound generators are being used.

During the composition-performance, each model is individually initialized, or "tuned", then gradually integrated into a global arrangement structured by arbitrarily generated action patterns or manually determined sequences. This process is not deterministically but exploratorily driven - less "writing" a piece of music than "finding" it.

3 CONCLUSION

Reykjavík Sunburn exemplifies the creative practice of *Latent Jamming*, a music-making approach using neural audio synthesis and employing techniques of generative music and algorithmic composition. It demonstrates a shared human-machine-agency in an improvisational setting, embracing the spontaneity of four generative neural audio networks while catering to human intent by providing a control interface for latent space interaction.

REFERENCES

- Caillon, A., and Esling, P. (2021). RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. 10.48550/arXiv.2111.05011.
- Chemla—Romeu-Santos, A. (2020). Manifold representations of musical signals and generative spaces.
- Demerlé, N., Esling, P., Doras, G., and Genova, D. (2024). Combining audio control and style transfer using latent diffusion. ArXiv, abs/2408.00196.
- Horta Valenzuela, M., and Tomás, E. (2025). NEBULA: A PCA-Based Method to Explore RAVE-Encoded Audio Representations. In *Proceedings of the 22nd Sound and Music Computing Conference (SMC 2025)*, Graz, 49–56.
- Jourdan, T., Françoise, J., and Bevilacqua, F. (2026). Collective Craft: How Artists Collaborate to Train AI-based Audio Synthesis Model for Music. In *Proceedings of the 2026 Conference on Creativity and Cognition (C&C 2026)*, New York, 921–933.
- Liu, J., and Zheng, S. J. (2026). Exploring AI Audio Models in Soundwalking with Broader Audiences. In *Proceedings of the 2026 Conference on Creativity and Cognition (C&C 2026)*, New York, 1699–1703.
- Tahiroğlu, K., Kastemaa, M., and Koli, O. (2021). GANSpaceSynth: A Hybrid Generative Adversarial Network Architecture for Organising the Latent Space using a Dimensionality Reduction for Real-Time Audio Synthesis. In *Proceedings of the 2nd Conference on AI Music Creativity (AIMC 2021)*.
- Tahiroğlu, K., and Wyse, L. (2024). Latent Spaces as Platforms for Sonic Creativity. In *Proceedings of the 15th International Conference on Computational Creativity (ICCC 2024)*, Sweden.
- Tahiroğlu, K., Hokkanen, M., and Marta, A. (2026). Human-in-the-Loop: Crossmodal AI Alignment between Movement and Audio Latent Spaces for Expressive Sonification in Dance Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2026)*, London.
- Tatar, K., Cotton, K., and Bisig, D. (2023). Sound Design Strategies for Latent Audio Space Explorations Using Deep Learning Architectures. 10.48550/arXiv.2305.15571.
- Vear, C., Benford, S., Avila, J. M., and Moroz, S. (2023). Human-AI Musicking: A Framework for Designing AI for Music Co-creativity. In *Proceedings of the 4th AI Music Creativity*

⁴ <https://martstil.de/works/reykjavik-sunburn>

⁵ <https://soundcloud.com/martsman/sets/black-plastics-series>

Conference (AIMC 2023), Sussex.

- Xambó, A. and Roma, G. (2024). Human-machine agencies in live coding for music performance. *Journal of New Music Research*, 53, 33–46.
- Yee-King, M. (2022). Latent Spaces: A Creative Approach. In *The Language of Creative AI: Practices, Aesthetics and Structures*. Springer Series on Cultural Computing. Cham, 137–154.
- Zheng, S., Sedó, A., and Bryan-Kinns, N. (2024). A Mapping Strategy for Interacting with Latent Audio Synthesis Using Artistic Materials. ArXiv, abs/2407.04379.
- Zheng, S. J., Yoshida, K., García-Peguinho, N., Liu, J., Hearn, D., Xambó Sedó, A., and Bryan-Kinns, N. (2026). Latent Terrain: Adapting Neural Audio Autoencoders as Design Materials in NIME. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2026)*, London.