

# degr d t n

NICHOLAS SHAHEED, Stanford University, USA

Additional Key Words and Phrases: neural networks, feedback loops, process music, interactive

## ACM Reference Format:

Nicholas Shaheed. 2024. degr d t n. 1, 1 (September 2024), 2 pages.

## 1 PROGRAM NOTES

Slow, repetitive, contemplative, and noisy. Repeatedly-applied neural reconstructions fail slowly and imperceptibly as feature space is explored.

Visuals designed by Emilio Ocelotl.

## 2 PROJECT DESCRIPTION

This audiovisual performance uses the 'I Am Sitting in a (Latent) Room' improvisation system [3]. Inspired by Alvin Lucier's 'I Am Sitting in a Room' and the general process of degrading sound by repeatedly passing it through an acoustic medium, it is a system that allows the improvisers to interact with the process of degradation in real time. The visuals reflect this sense of decay, with videos of nature quickly interspliced, processed, and faded.

Using a bespoke variational autoencoder (VAE) model, a 25-second audio clip is repeatedly encoded and decoded through two parallel instances of the model. On top of this process, the performers take on the role of improviser. By manipulating the model's latent embeddings of the audio in real-time, they explore the latent space (or "room") of the model over the course of the performance.

Designed for 1-3 performers, this work uses a mix of physical controllers and live coding. It consists of a core audio rendering server, and a set of clients (corresponding to each performer) sending commands to the server via OSC. The server was written in Chuck, and the clients a mix of Chuck and SuperCollider. The visuals were designed by Emilio Ocelotl using TouchDesigner.

The model uses the Realtime Audio Variational autoEncoder (RAVE) architecture[1]. With a custom UGen made by the author to perform real-time inference with RAVE models inside Chuck, based off of the *nn Max/MSP* object[2].

This system embodies an irony from the insertion of human interactivity into a fully automatic process. With a large body of current AI research working toward more and more advanced fully automated systems with perhaps no real regard for how much automation is what we'd want in the first place, and of course Lucier's work is a canonical example of process music. However, even with both of these pushing towards a fully automatic solution the interactivity is still there, for better or for worse.

---

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

### 3 PERFORMANCE NOTES

The performance of this work will require stereo output, and video projection of 1080p. There will be between 1-3 performers on separate laptops, and enough power outlets, chairs, and table space to accommodate the players. Only one computer will require audio output and only one will require video output. Laptops, audio interfaces, and controllers will be provided by the performers.

### 4 MEDIA LINKS

- Video: <https://vimeo.com/943876017>

### ETHICAL STANDARDS

There are many new and unanswered ethical and aesthetic questions regarding the use of AI, including the use of generative models in artistic fields. There remains more questions than answers regarding authorship, displacement of labor, aesthetic and cultural implications, and the role of advanced computational machinery in human creative endeavors. One virtue this paper prioritizes is working with AI at an individual human scale: building a holistic system from its software to its model, to its performance as an artistic endeavor. The model used in *degr d t n* was trained on a single, admittedly quite expensive, consumer-grade GPU. It's dataset was composed and recorded by a single person, and ultimately resulted in three artists coming together to perform music in the premiere of *degr d t n*. In contrast to many tools built with generative AI that seek to supplant or remove humans from creative efforts, I Am Sitting in a (Latent) Room offers a system that is channeling and facilitating human creativity, connection, and art making at every stage.

### REFERENCES

- [1] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011* (2021).
- [2] Antoine Caillon and Philippe Esling. 2022. Streamable Neural Audio Synthesis With Non-Causal Convolutions. *arXiv preprint arXiv:2204.07064* (2022).
- [3] Nicholas Shaheed and Ge Wang. 2024. I Am Sitting in a (Latent) Room. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.