Qosmo

Emergent Rhythm — Real-time AI Generative DJ Set

Nao Tokui, Ryosuke Nakajima, Keito Takaishi (Qosmo/Neutone)

neutone

Overview

"Emergent Rhythm" is an audio-visual DJ performance using real-time AI audio generation. Artist/DJ Tokui manipulates multiple models on stage to spontaneously generate rhythms and melodies. He then combines and mixes the generated audio loops to create musical developments. The artist faces unprecedented challenges: Everything heard during this performance is purely AI-generated sound.

As the title suggests, we focus on the musical and visual "rhythms" and recurring patterns that emerge in the interaction between multiple AI models and the artist. The accompanying visuals feature not only the periodicity over time but also the common patterns across multiple scales ranging from the extreme large scale of the universe to the extreme small-scale of cell and atomic structures.

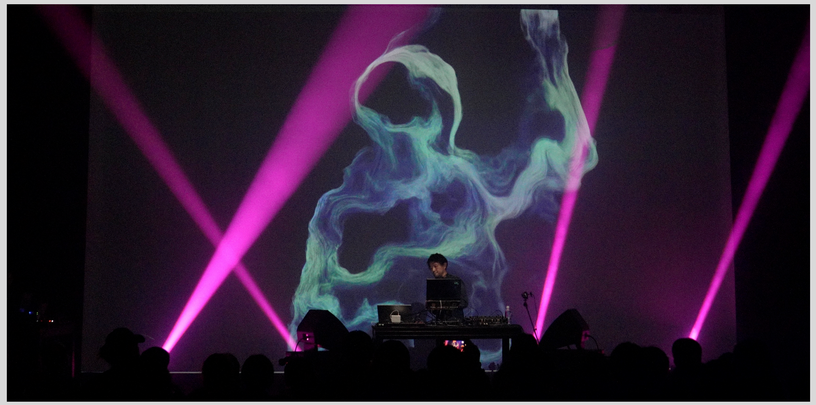
Aligning with the visual theme, we extracted loops from natural and man-made environmental sounds and used them as training data for audio generation. The artist also employs real-time timbre transfer that converts incoming audio into various singing voices, such as Buddhist chants. This highlights the diversity and commonality within the human cultural heritage.

From a DJ session, in which existing songs are selected and mixed, to a live performance that generates songs spontaneously and develops them in response to the audience's reactions: In this performance, the human DJ is expected to become an AJ, or "AI Jockey," rather than a "Disk Jockey," taming and riding the AI-generated audio stream in real-time. With the unique morphing sounds created by AI and the new degrees of freedom that AI allows, the AI Jockey will offer audiences a unique, even otherworldly sonic experience.

Performances

2022.12.8 MUTEK.JP (Shibuya Stream Hall, Tokyo)
2022.12.9 Craft Alive (Daikanyama UNIT, Tokyo)
2023.6.13 Sonar+D/+RAIN Film Festival (Barcelona, Spain)
2023.9.13 LVMH DATA AI Summit (Paris,France)
2023.10.21 Digital Art Zurich (Kunsthaus Zurich, Switzerland)

Video and more info



Technical Specifications

Audio Generation

We adapted the GAN (Generative Adversarial Networks) architecture for audio synthesis. **StyleGAN**[1] models trained on spectrograms of various sound materials generate spectrograms[2], and vocoder GAN models (**MelGAN**)[3] convert them into audio files. By leveraging GAN-based architecture, we can generate novel, constantly changing, morphing sounds similar to GAN-generated animated faces of people who don't exist. We created a custom music dataset to train those models while respecting the copyrights of other artists. It takes about 0.5 seconds to generate 4-second-long 2-bar loops in a batch; hence it's faster than real-time.

Controllability

We also implemented a method called **GANSpace**, proposed by Härkönen et al[4], to provide perceptual controls during the performance. GANSpace applies Principal Component Analysis (PCA) on the style vector of a trained StyleGAN model to find perceptually meaningful directions in the latent style space. Adding offsets according to these vectors allows the DJ to influence the audio generation in their desired direction. The core part of the audio generation system is open-sourced and available on GitHub[5].

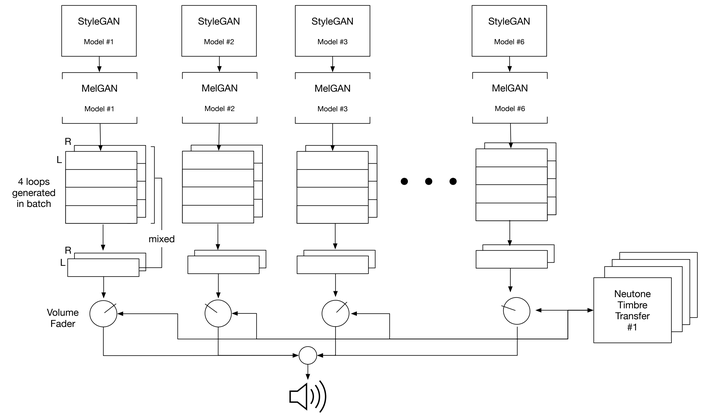
Timbre Transfer

In this performance, the DJ also utilized multiple neural timbre transfer models in the **RAVE** architecture[6] trained on various singing styles from diverse cultures, such as church choirs, and natural sounds like bird songs. These models are run on the Neutone plugin[7], a universal host for real-time AI audio models.

AI Visuals

We utilized an open-source pre-trained text-to-image generation model, **Stable Diffusion**, to generate over 1.0 million images. For seamless playback of these images in response to the DJ performance, we used custom-made VJ software written in TouchDesigner and openFrameworks.

System Diagram



References

1. Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2019. "Analyzing and Improving the Image Quality of StyleGAN."
2. Hung, Tun-Min, Bo-Yu Chen, Yen-Tung Yeh, and Yi-Hsuan Yang. 2021. "A Benchmarking Initiative for Audio-Domain Music Generation Using the Freesound Loop Dataset."
3. Kumar, Kundan, Rithesh Kumar, Thibault de Boissiere, Lucas Gestin, Wei Zhen Teoh, Jose Sotelo, Alexandre de Brebisson, Yoshua Bengio, and Aaron Courville. 2019. "MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis."
4. Härkönen, Erik, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. "GANSpace: Discovering Interpretable GAN Controls."
5. Tokui, Nao. 2021. "LoopGAN" https://github.com/naotokui/LoopGAN
6. Caillon, Antoine, and Philippe Esling. 2021. "RAVE: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis."
7. Neutone AI audio plugin and community. 2021. https://neutone.space/

contact@qosmo.jp