

# Reconstructing Vintage Electone Textures with RAVE

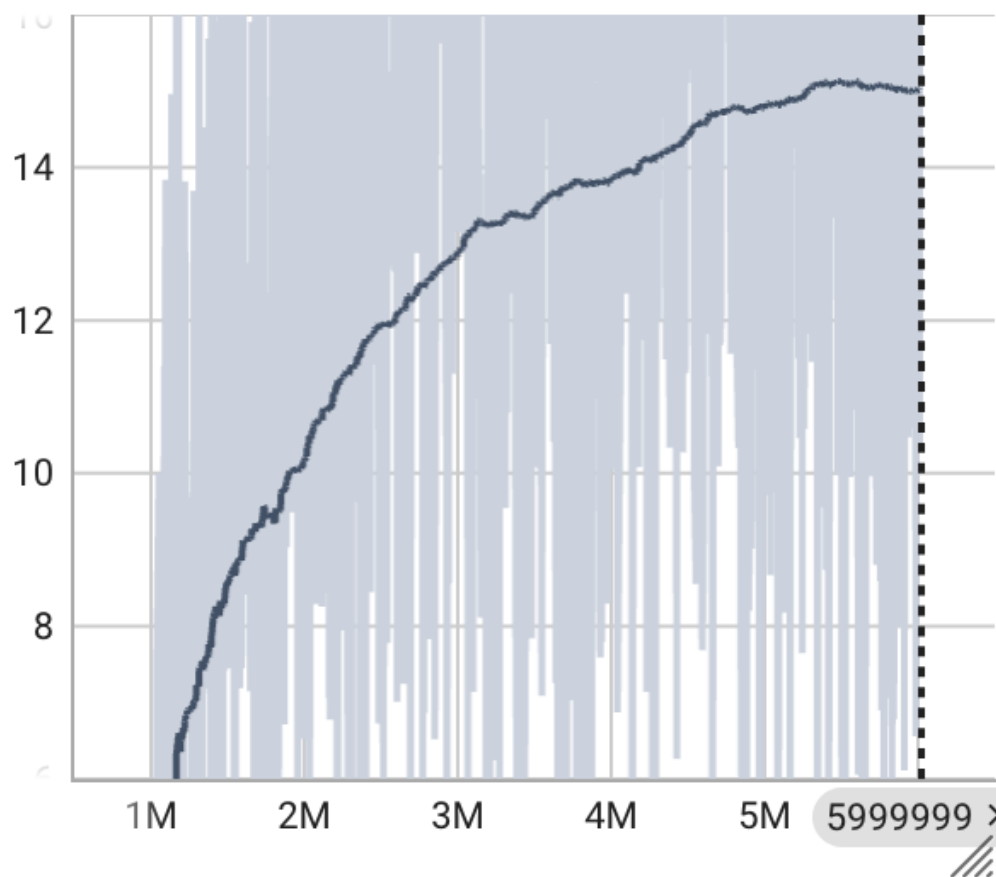
## Materials

- Hardware: Google Colab GPU A100
- Baseline Dataset: MusicNet
- Custom Dataset: *1970s Japanese Electone Music Album from Shigeo Sekito* (~90 minutes)
- Model: RAVE V2
- Sampling Rate: 44.1kHz
- Check Config.gin in attached for hyperparameters

## Training Process

- The training process followed the official IRCAM RAVE tutorial: [Training RAVE models on custom data – IRCAM Forum](#)
- Below are TensorBoard screenshots showing various training metrics:

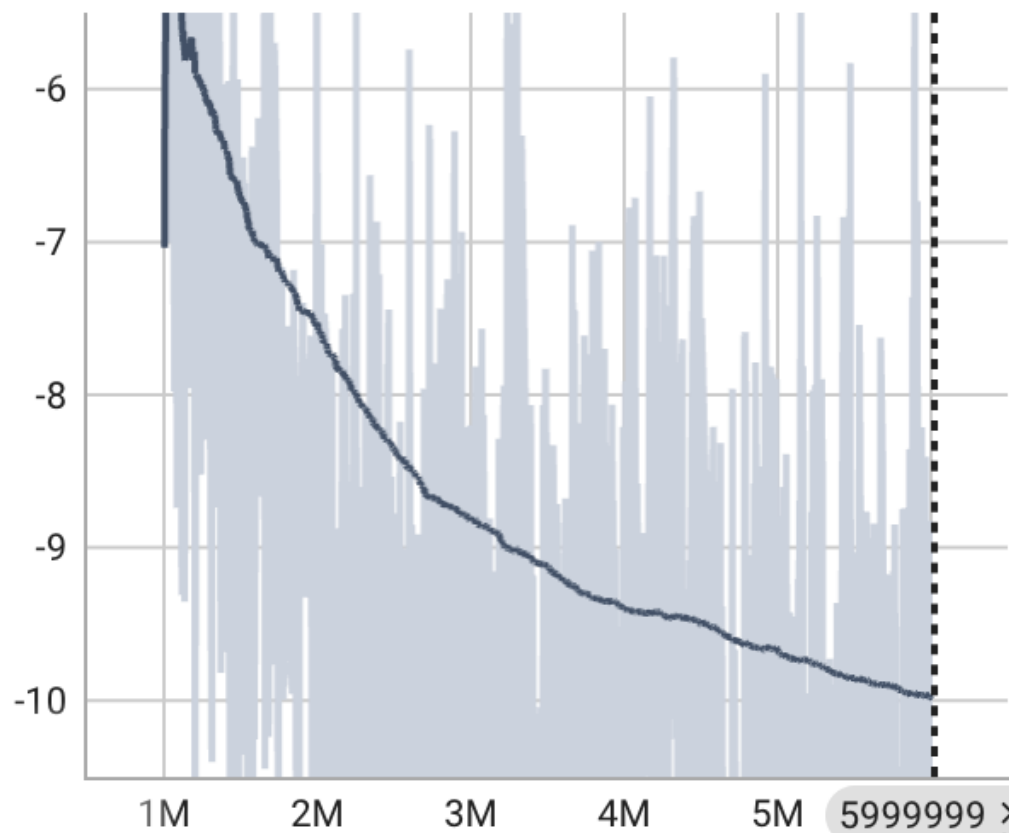
pred\_real



| Run ↑ | Smoothed | Value   | Step      | Relativ |
|-------|----------|---------|-----------|---------|
| .     | 15.0141  | 10.3976 | 5,999,999 | 6.142 c |



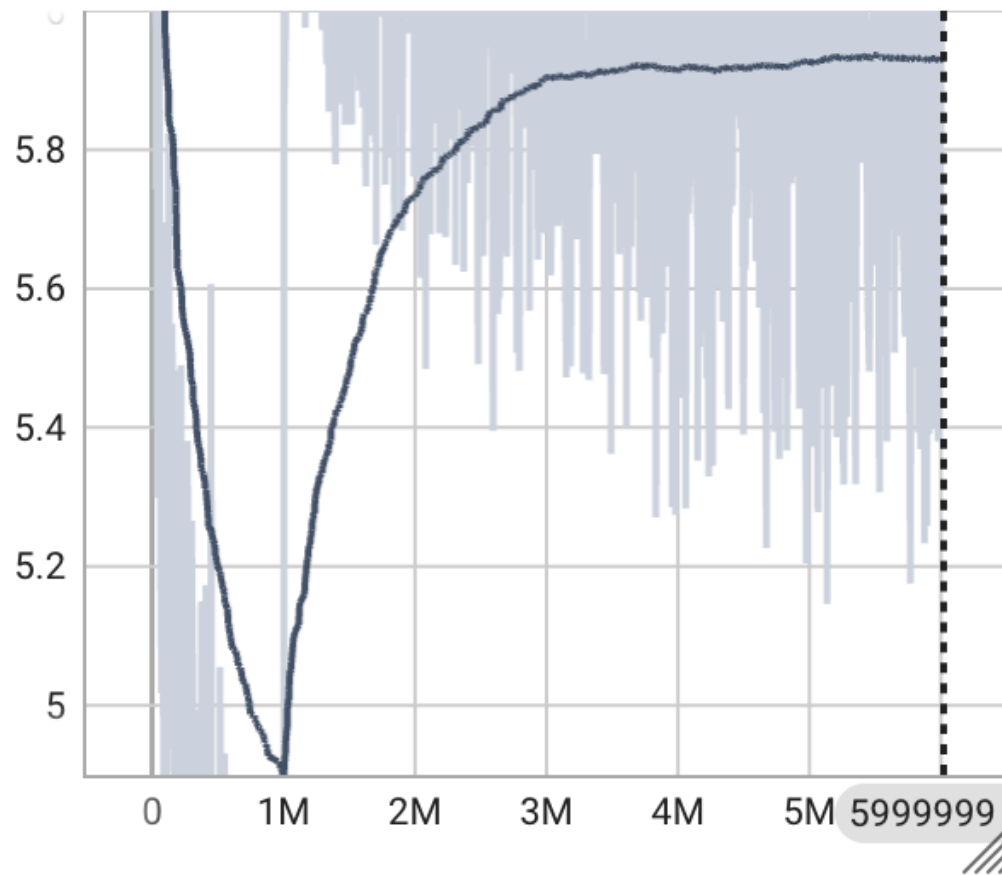
pred\_fake



| Run ↑ | Smoothed | Value    | Step      | Relative |
|-------|----------|----------|-----------|----------|
| .     | -9.9816  | -11.0183 | 5,999,999 | 6.1426   |



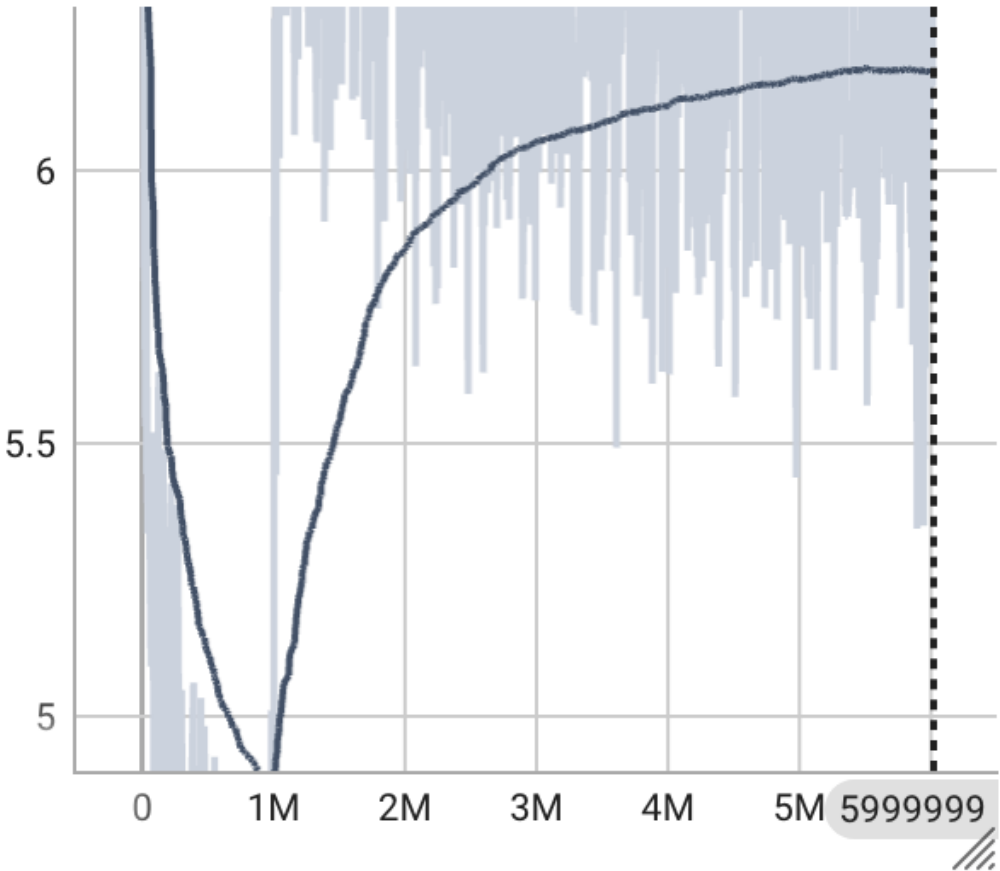
multiband\_spectral\_dista...



| Run ↑ | Smoothed | Value  | Step      | Relativ |
|-------|----------|--------|-----------|---------|
| .     | 5.9305   | 6.0564 | 5,999,999 | 6.928 c |



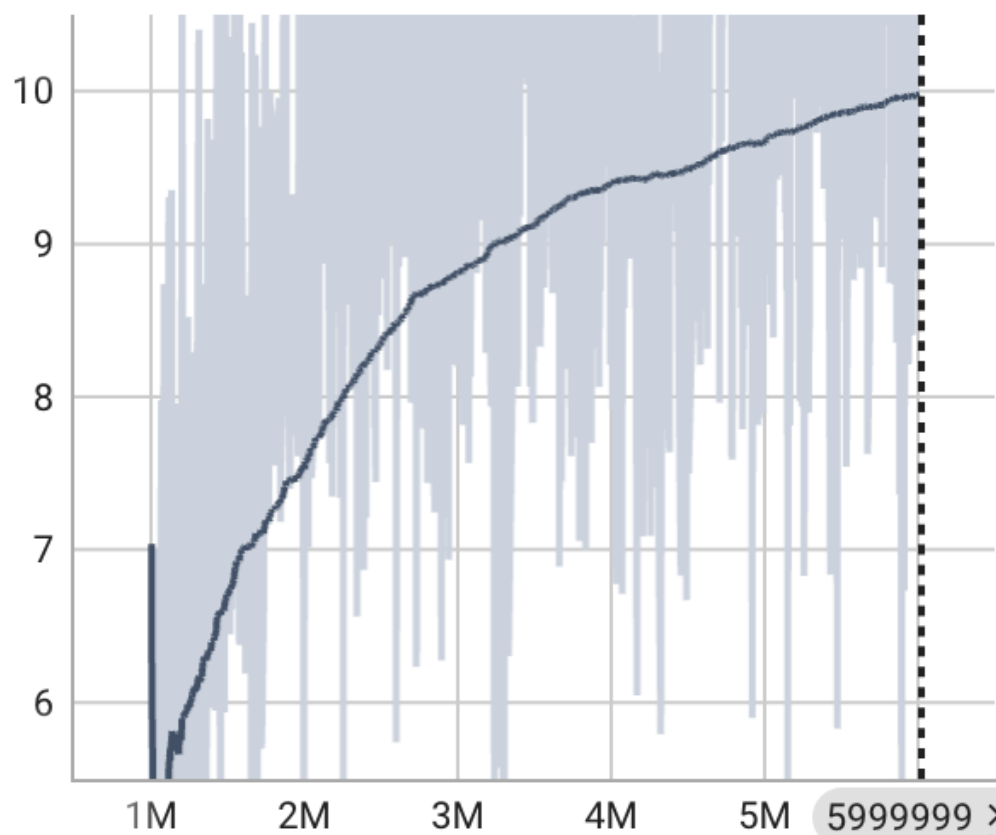
fullband\_spectral\_distance



| Run ↑ | Smoothed | Value  | Step      | Relativ |
|-------|----------|--------|-----------|---------|
| .     | 6.1819   | 6.1536 | 5,999,999 | 6.928 c |



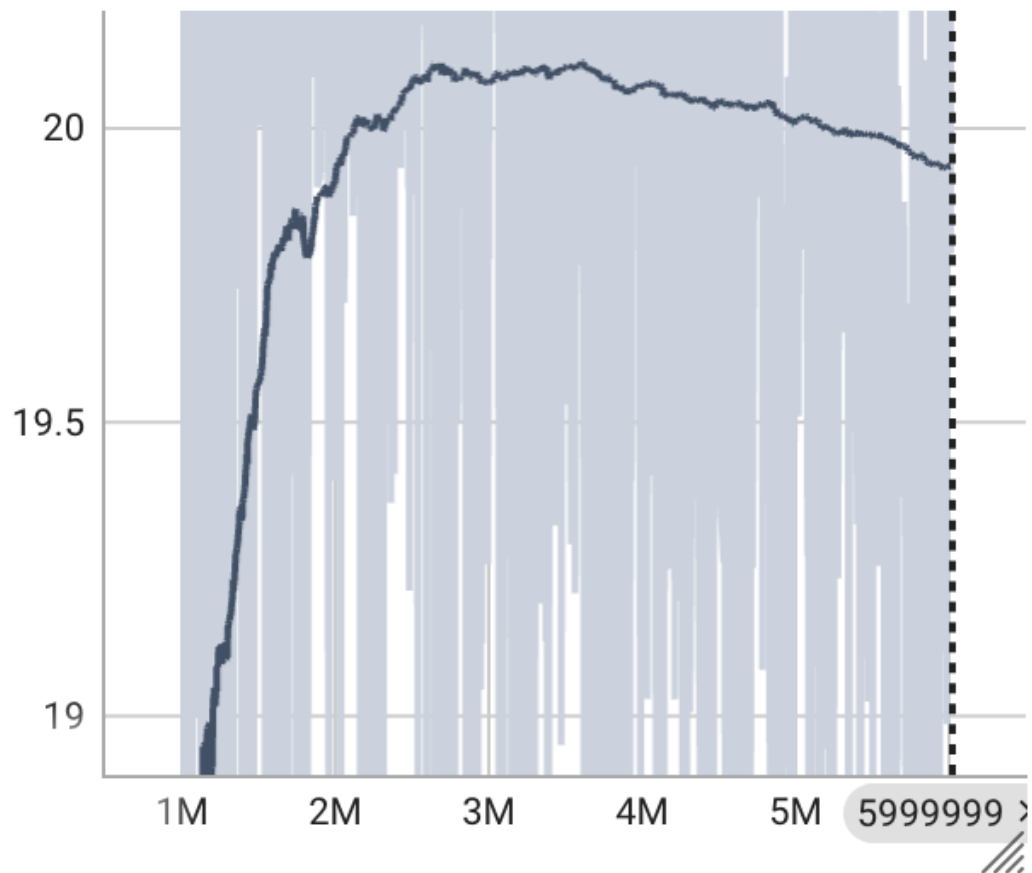
adversarial



| Run ↑ | Smoothed | Value   | Step      | Relative |
|-------|----------|---------|-----------|----------|
| .     | 9.9816   | 11.0183 | 5,999,999 | 6.142 c  |



feature\_matching



| Run ↑ | Smoothed | Value   | Step      | Relativ |
|-------|----------|---------|-----------|---------|
| .     | 19.936   | 19.7873 | 5,999,999 | 6.142 c |



## Training Process and Observations

- At the beginning, I experimented with several pre-trained RAVE models available from IRCAM and Hugging Face, generating various beats and textures for reference and comparison.

- Before training with my own dataset, I tested the workflow using the **MusicNet** dataset and successfully replicated the results presented in IRCAM's official documentation.
- After confirming the baseline performance, I proceeded to train **RAVE V2** with my custom *1970s Japanese Electone Music Album* dataset (~90 minutes in total).

## Training Behavior

- Initially, the model **did not converge**, and there was **no meaningful output** observed in the `audio_val` tab.
- After enabling the **"noise"** configuration in the hyperparameters, the model **gradually began to converge**.
  - I assume this improvement is due to *1970s Japanese Electone Music Album* recordings contain a variety of **textural and noisy components**.
  - For datasets with cleaner or more tonal material, disabling the "noise" option might lead the model to behave more like a **timbre transfer model** rather than a texture generator.
- During training, the **"prediction\_fake"** and **"prediction\_real"** metrics did not behave exactly as shown in the tutorial (they did not diverge early in Phase 2 before converging later).
  - This could be related to the **small dataset size** or **limited timbral diversity**.

## Results

- Despite the irregular metric curves, the trained model successfully produced **interesting rhythmic and glitch-like audio textures** when integrated with **Max/MSP** and **Ableton Live**.
- Both the **trained checkpoint model** and `.ts` are attached for reference and further evaluation.

## Observations

- The smaller dataset likely affected model stability and convergence behavior.

- Nevertheless, qualitative results (audio output) remained musically valuable, especially for experimental electronic textures.
- Future work could involve:
  - Training a Prior Model.
  - Expanding the dataset with additional *Electone* recordings.
  - Comparing results using RAVE V3 or a different discriminator.