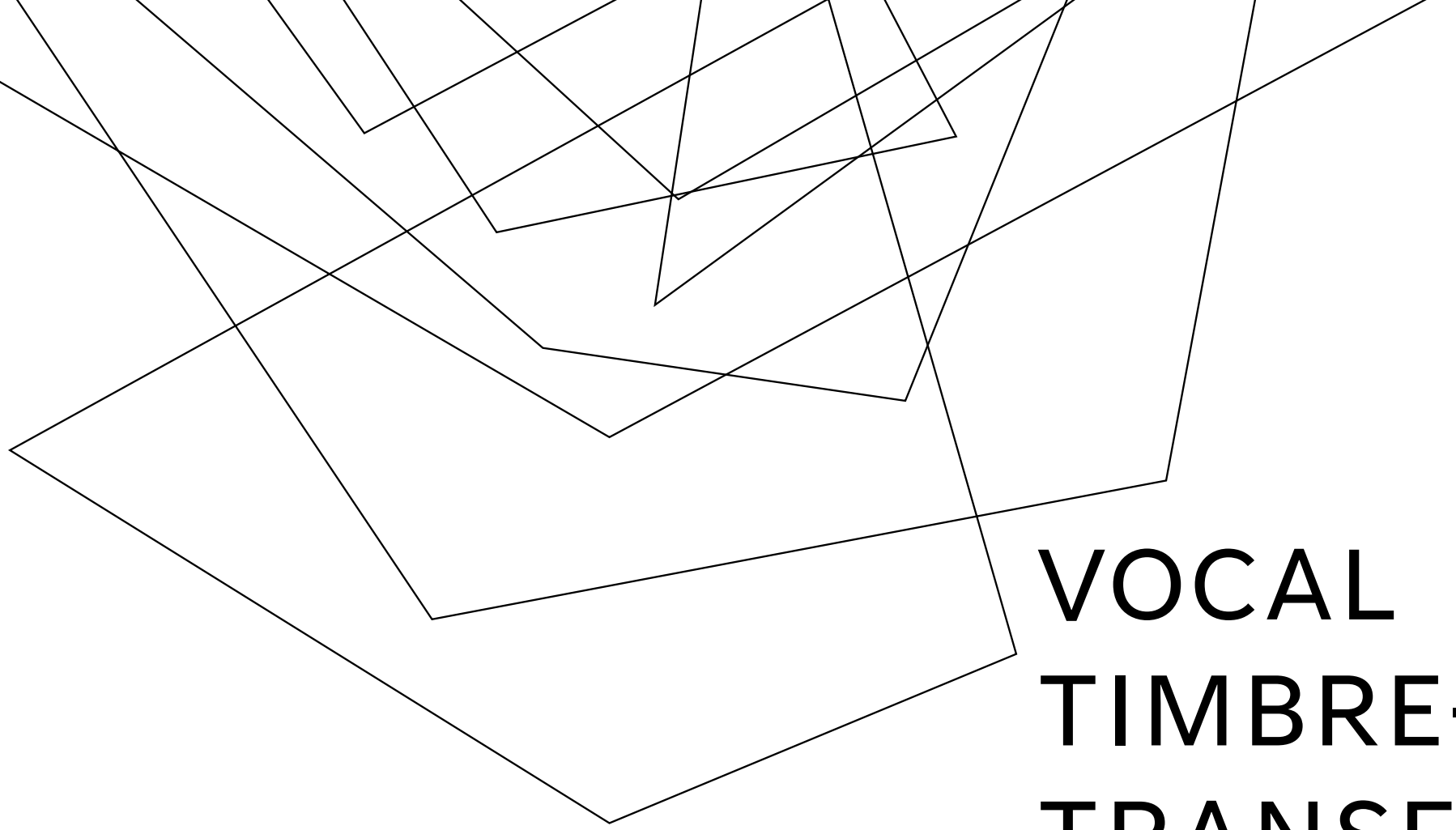


VOCAL TIMBRE-TRANSFER

21M.080 Final Presentation

by Thelonious Cooper

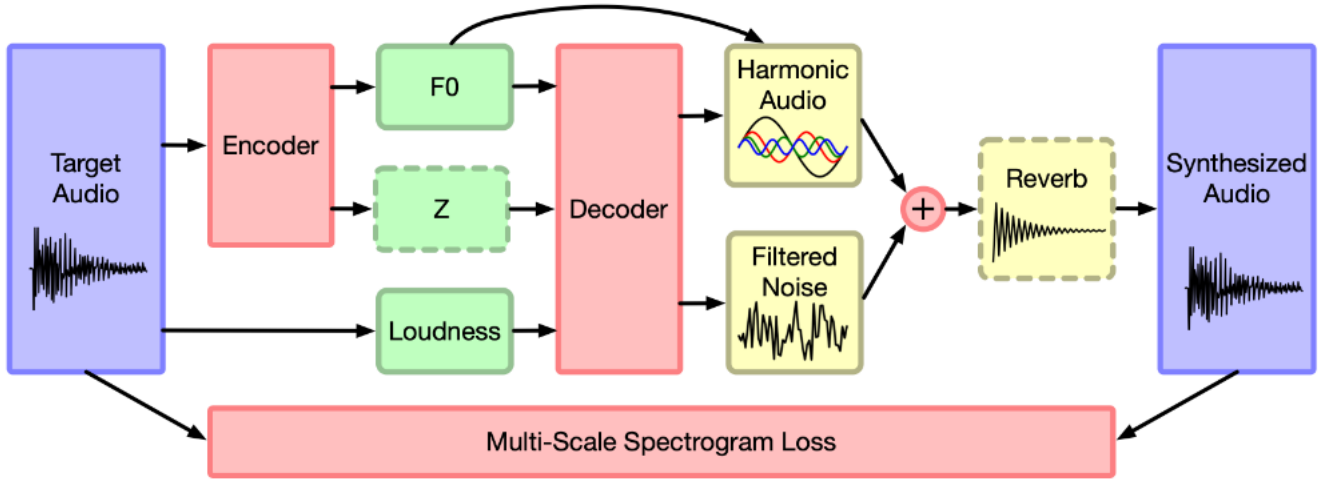
ML MODEL: AUTOENCODER

Encoder ->

Latent Variables ->

Decoder ->

Result



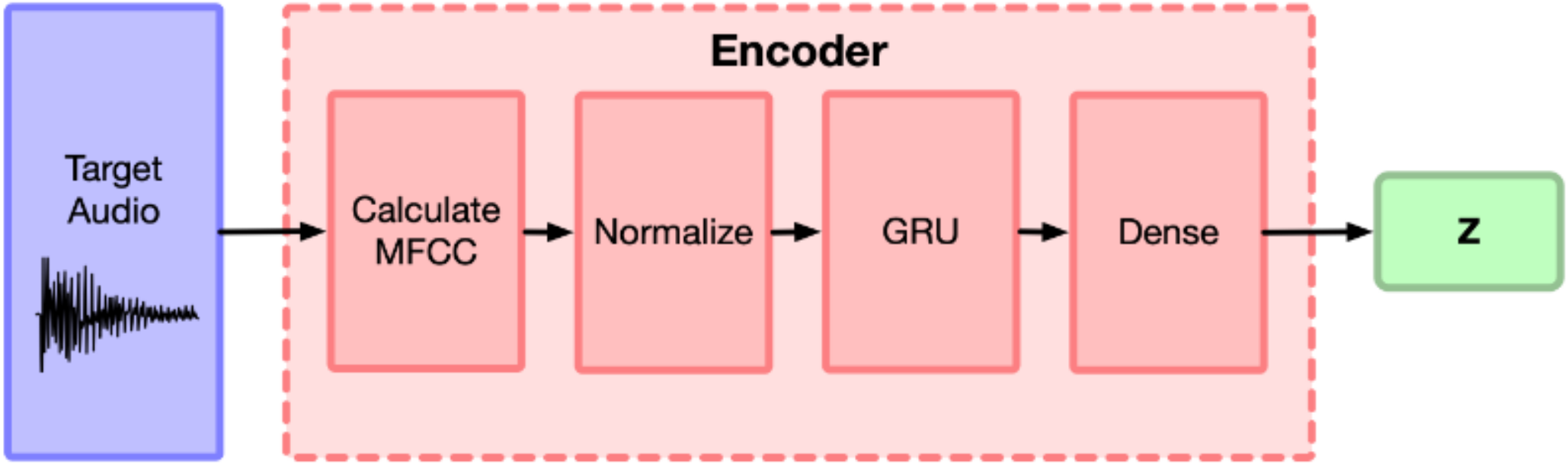
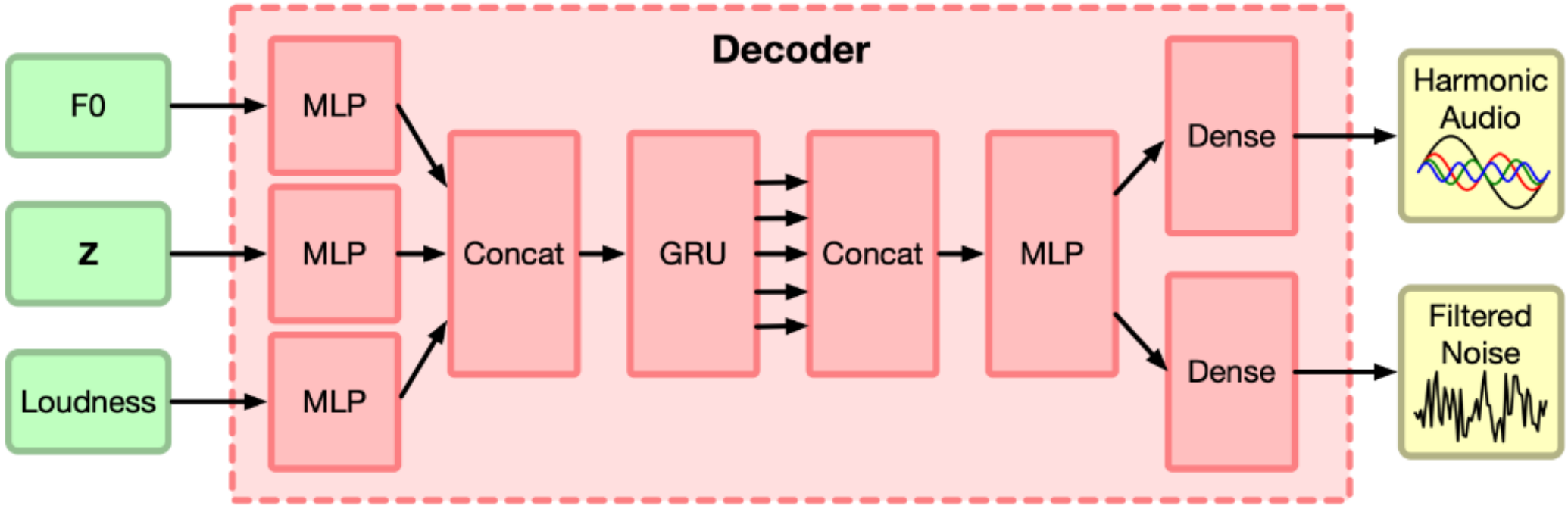
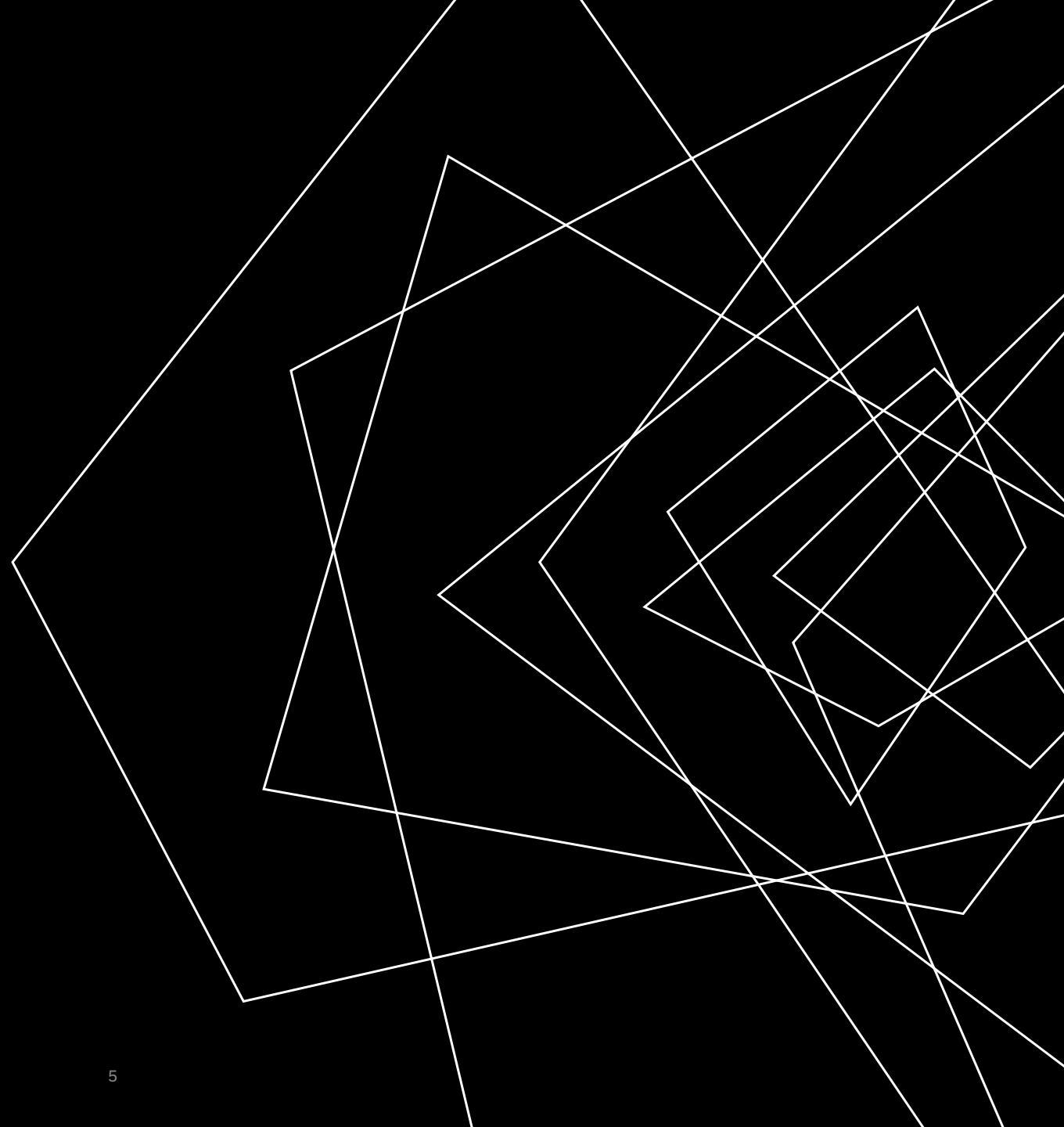


Figure 8: Diagram of the z-encoder.


Decoder
F0
z
Loudness
MLP
MLP
MLP
Concat
GRU
Concat
MLP
Dense
Dense
Harmonic Audio
Filtered Noise

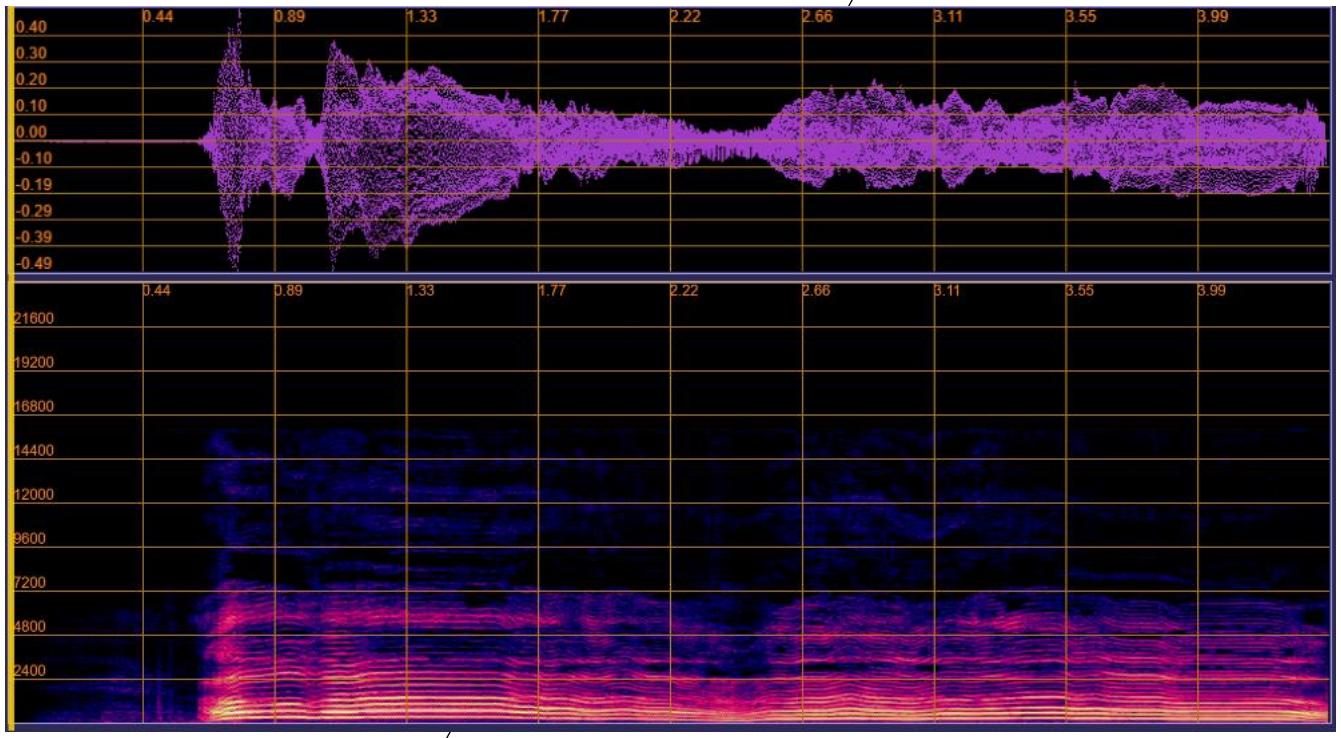
PROCESS

- Data Acquisition
- Data Preparation
- Training
- Results



DATA ACQUISITION:

Vocals were provided by friend and peer David Morgan, whose vocal abilities far surpass my own



DATA PREP:

Had to convert from .m4a to .wav

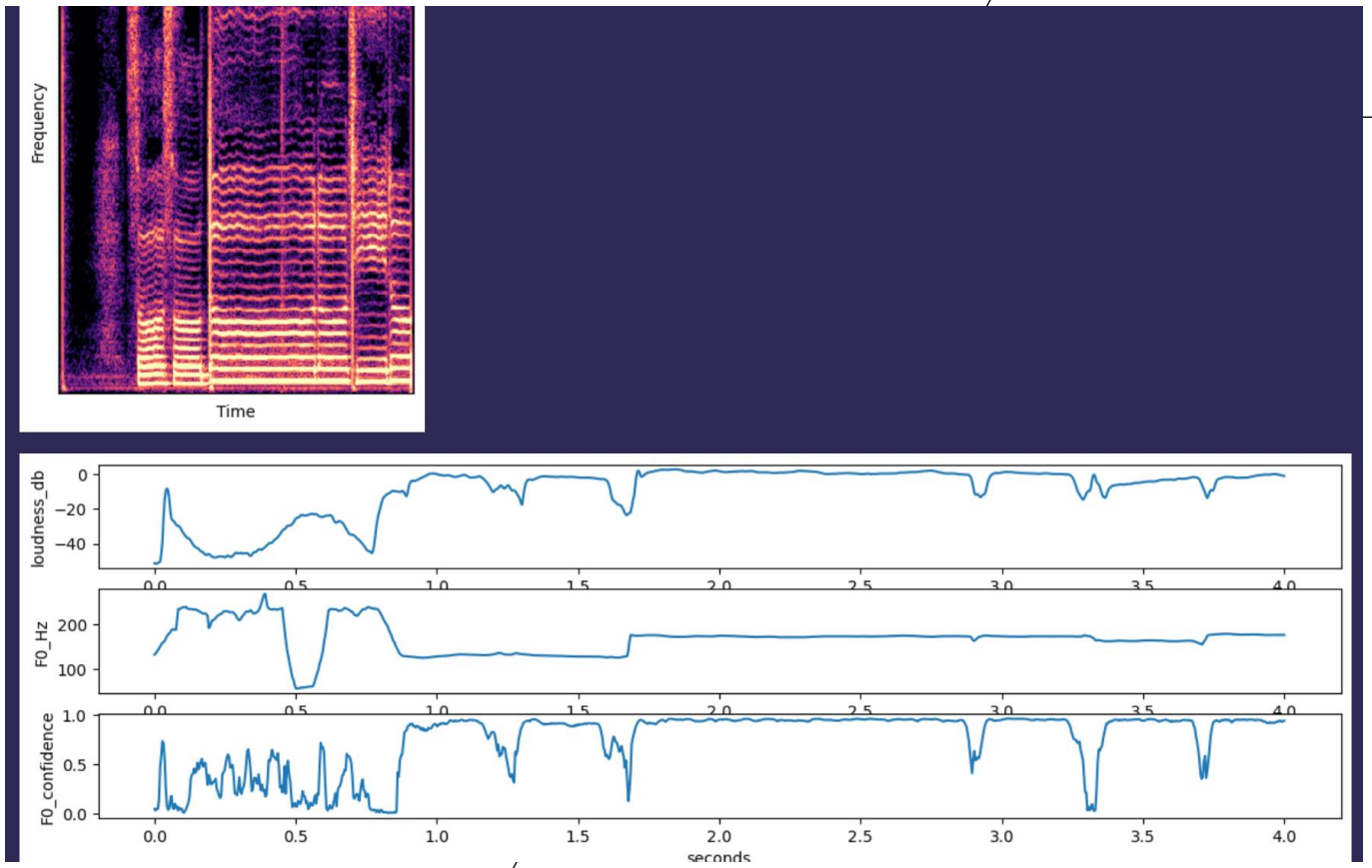
Normalize loudness

Convert into many 4s clips

Pre-compute Audio Features

- Loudness
- Fundamental Frequency
- F0_confidence

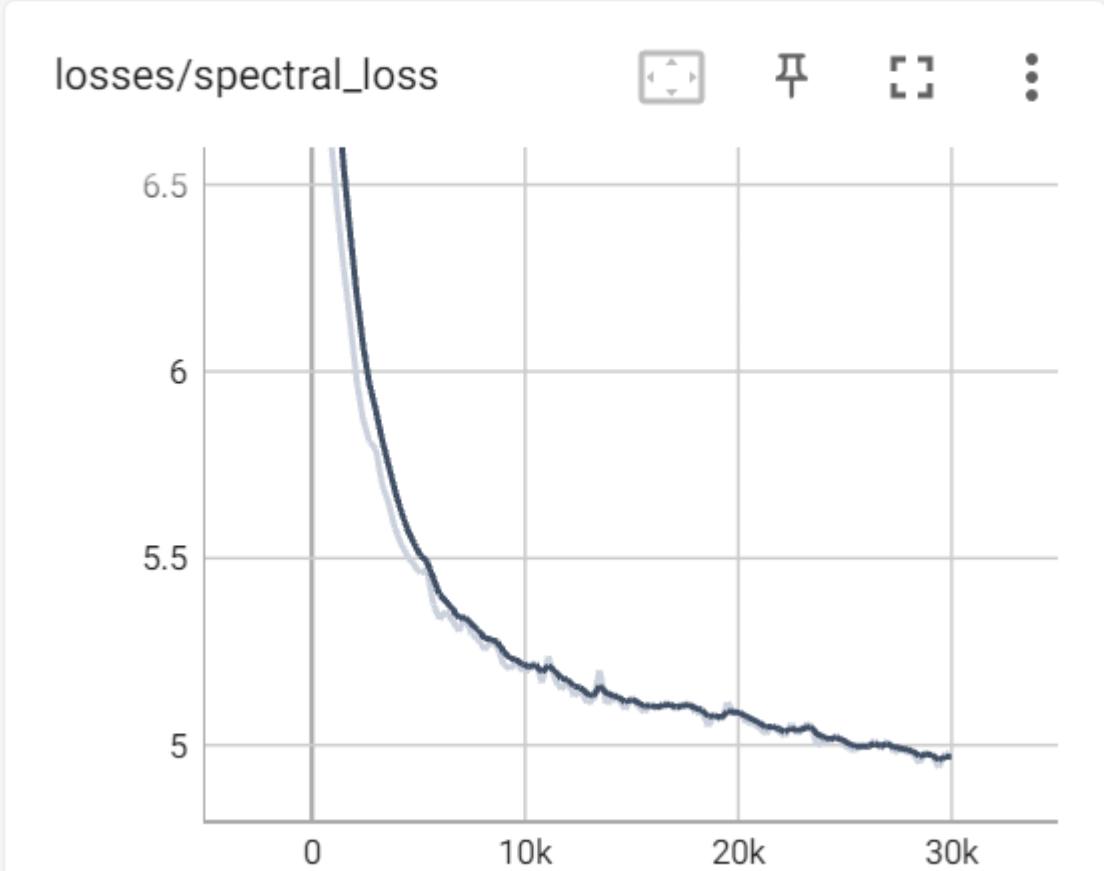
Convert into tensors

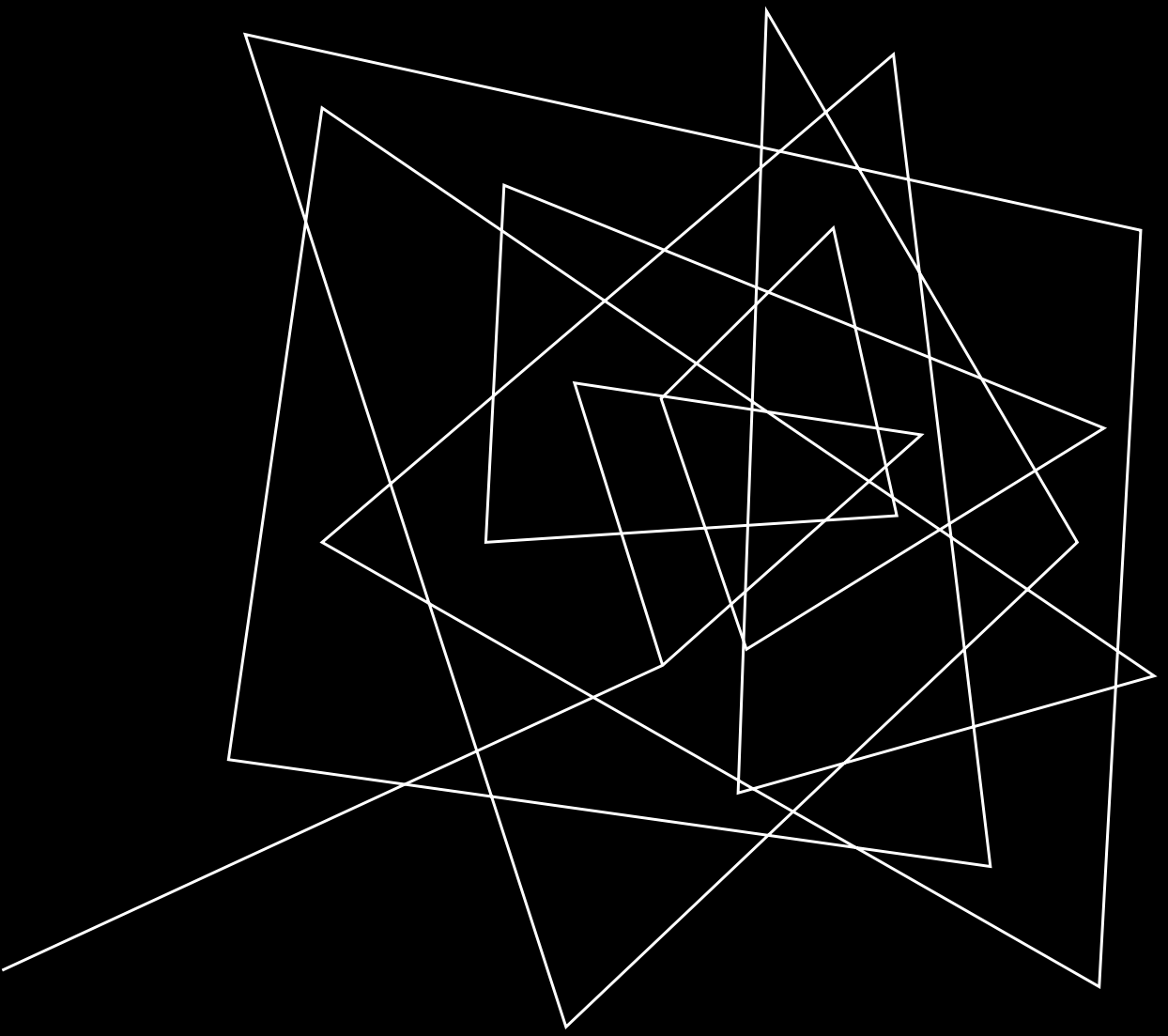


TRAINING:

Trained via traditional Stocastic Gradient Descent on a Spectral Loss Metric

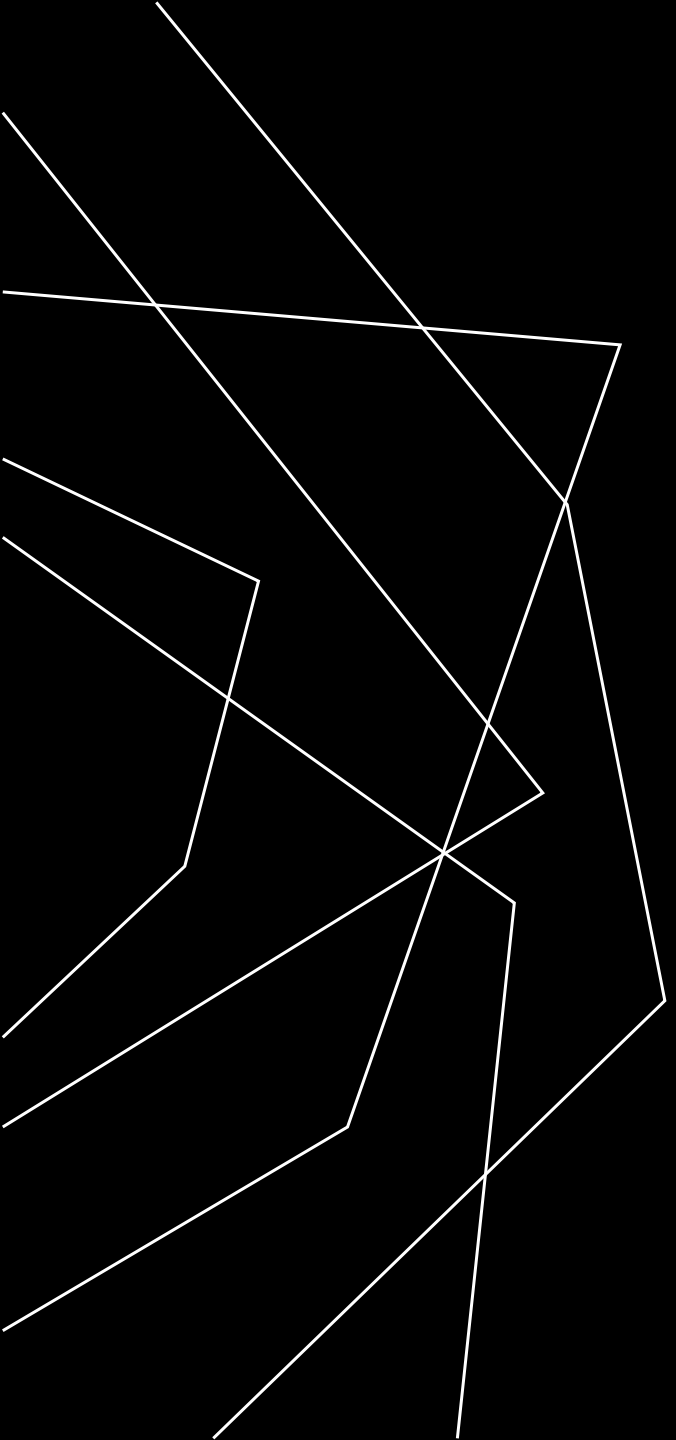
$$L_i = ||S_i - \hat{S}_i||_1 + \alpha|| \log S_i - \log \hat{S}_i||_1.$$





RESULTS

Ill Show some audio clips from vscode



THANK YOU

Thelonious Cooper

theloni@mit.edu

Theloniouscoop.dev