Week 5 Meeting

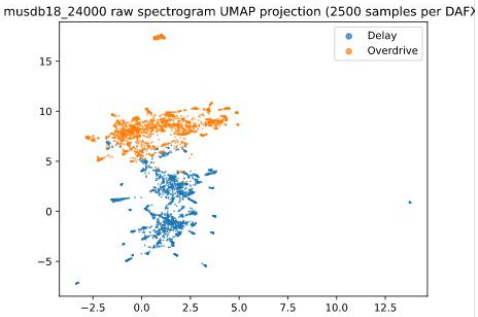
2357351G - MSci Half Project

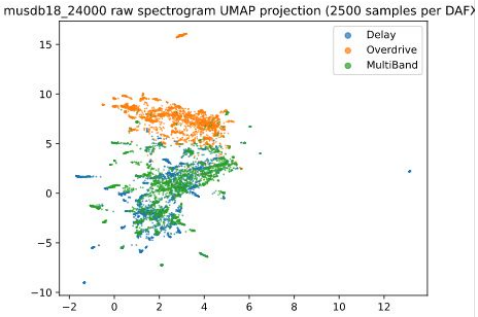
What I've done this week

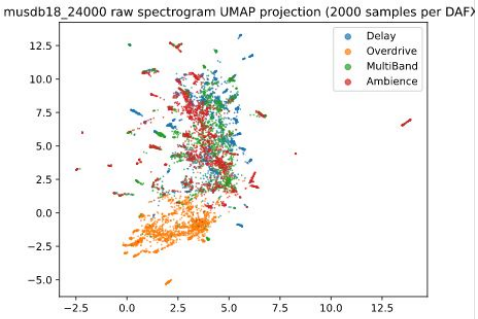
- Included a new, simpler dataset: the VTCK speech dataset to help with debugging.
- Plotted UMAP projections of spectrograms directly.
- Finished implementation of Spectrogram VAE.
- I did some more investigation into the poor Spectrogram VAE reconstruction using the MNIST dataset with my VAE.
    - Main issue seems to be weighing the KL loss too highly.
- Continued working on final report between tasks.

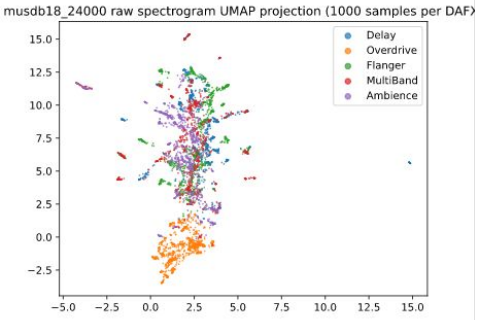
UMAP Projections

MusDB18 (Music) Dataset

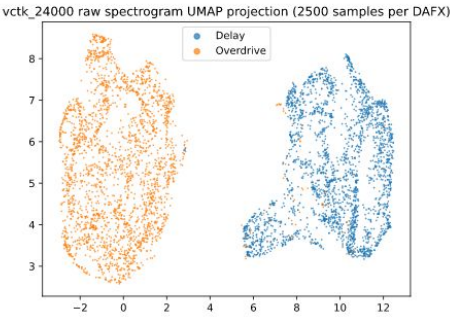


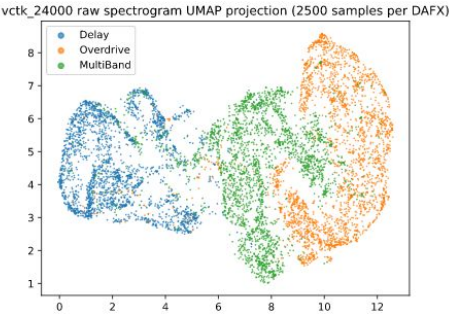


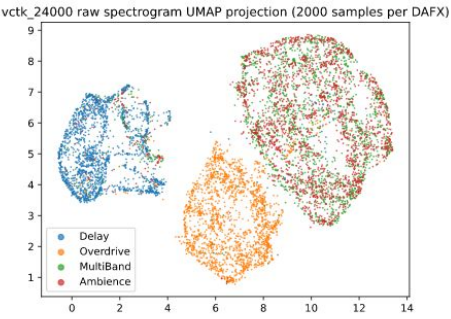


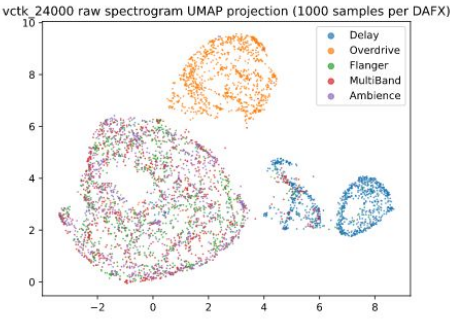


VCTK (Speech) Dataset



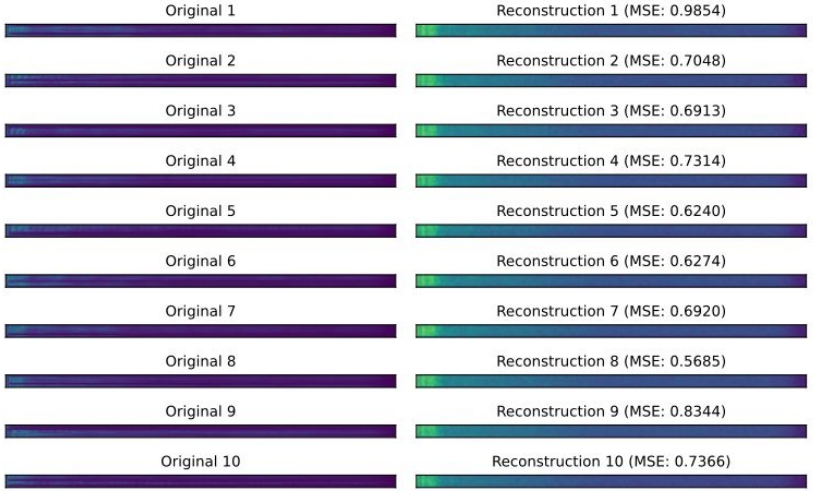




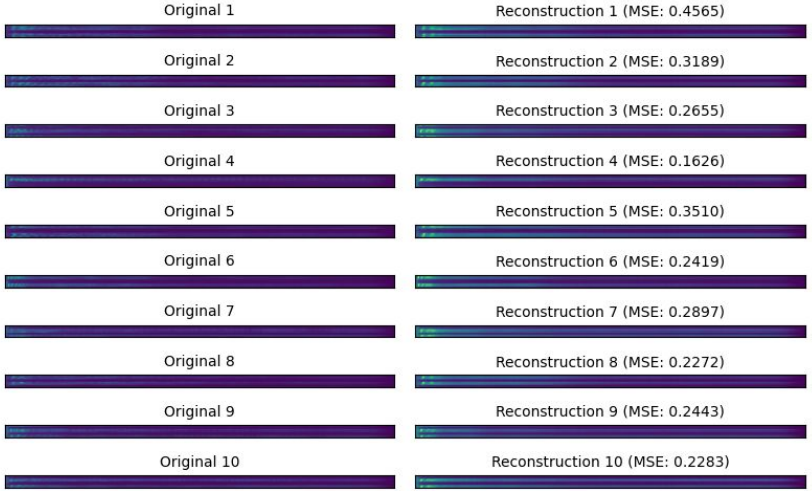


Spectrogram Reconstruction

β = 1 (200 epochs)

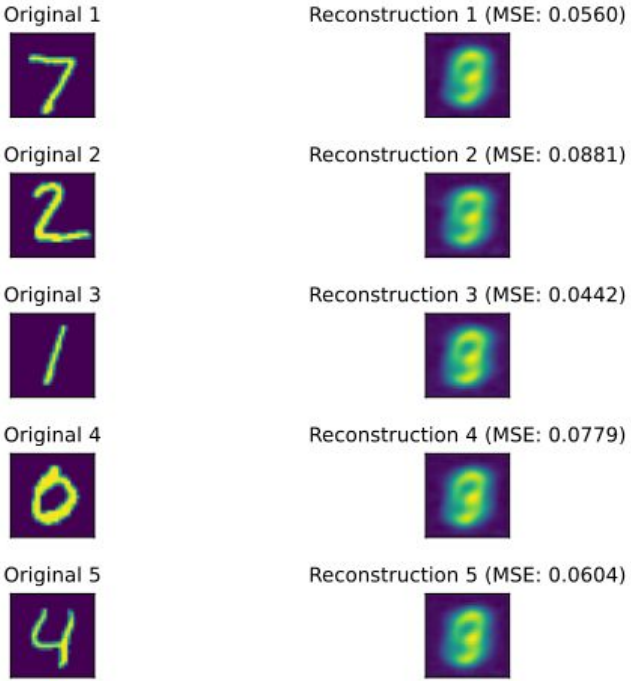


β = 0.001 (100 epochs)

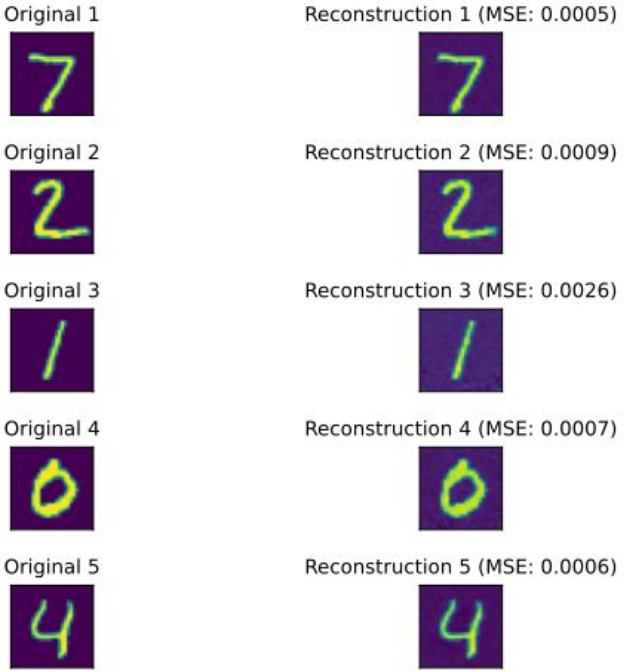


MNIST Reconstruction

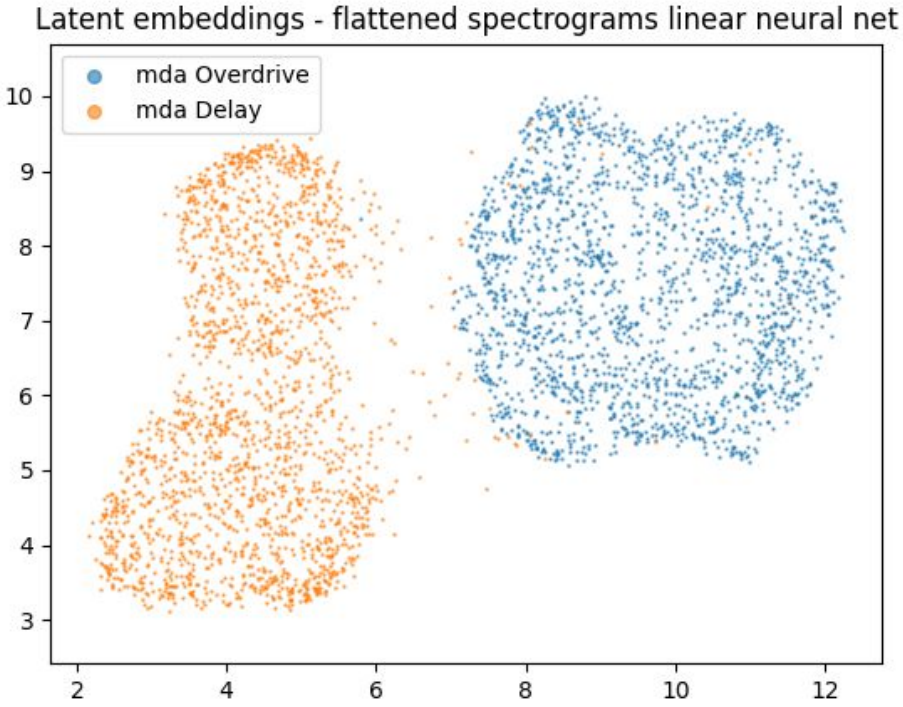
β = 1



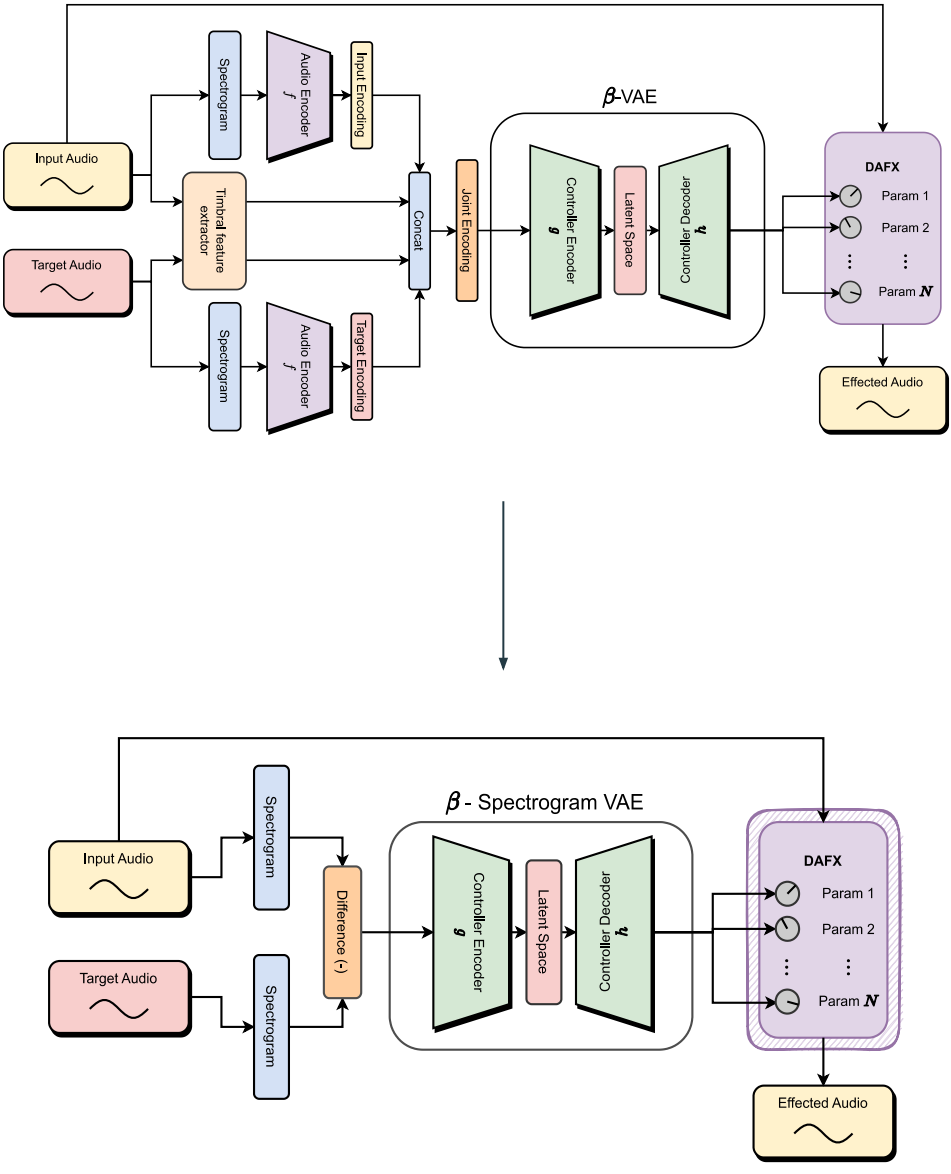
β = 0



Latent spectrogram embeddings - LinearNN



Potential simplification of network



Questions

- Is there anything obviously wrong in my implementation of the VAE loss? Or is it purely a case of performing hyperparameter tuning on the β value?

Plan for next week

- Revisit Spectrogram VAE architecture to improve reconstruction loss.
- Train on variety of DAFX and both datasets.
- Visualisations of reconstructions and latent embeddings.

Where I am in schedule

- Still working on creating sensible embeddings of audio which reflect the difference in effects.
- Working on final report during model training etc.