

Week 5 Meeting

2357351G - MSci Half Project

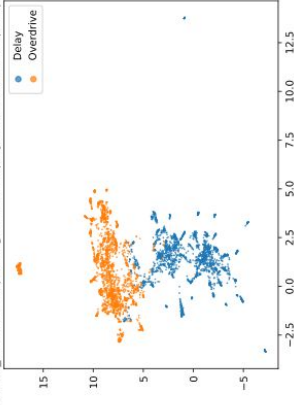
What I've done this week

- 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.

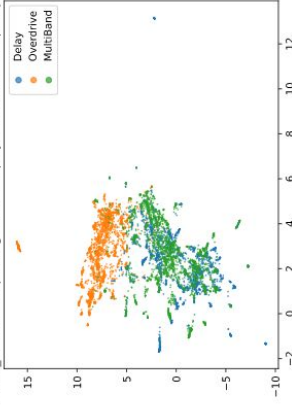
UMAP Projections

MusDB18 (Music) Dataset

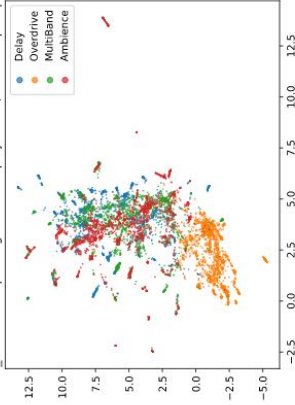
musdb18_24000 raw spectrogram UMAP projection (2500 samples per DAF)



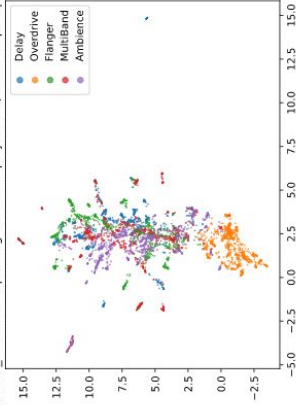
musdb18_24000 raw spectrogram UMAP projection (2500 samples per DAF)



musdb18_24000 raw spectrogram UMAP projection (2000 samples per DAF)

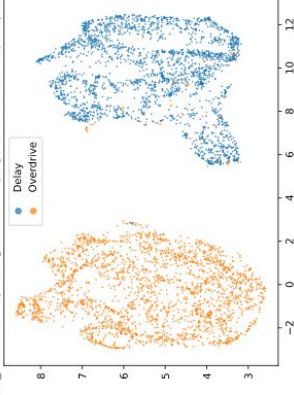


musdb18_24000 raw spectrogram UMAP projection (1000 samples per DAF)

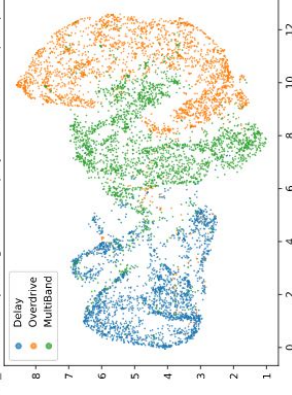


VCTK (Speech) Dataset

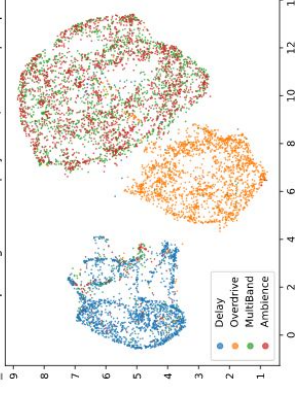
vctk_24000 raw spectrogram UMAP projection (2500 samples per DAF)



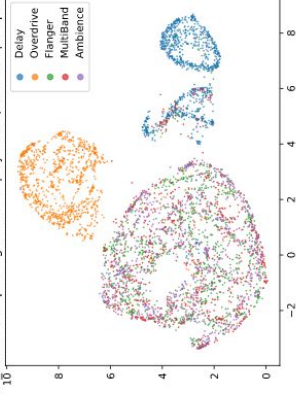
vctk_24000 raw spectrogram UMAP projection (2500 samples per DAF)



vctk_24000 raw spectrogram UMAP projection (2000 samples per DAF)

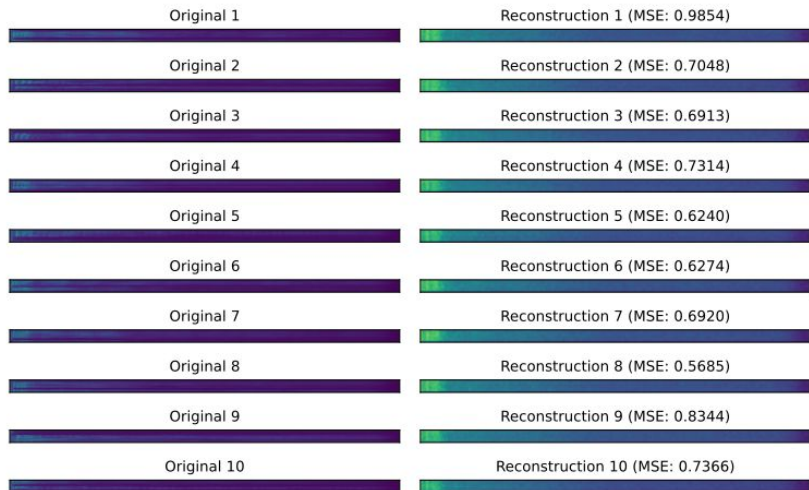


vctk_24000 raw spectrogram UMAP projection (1000 samples per DAF)

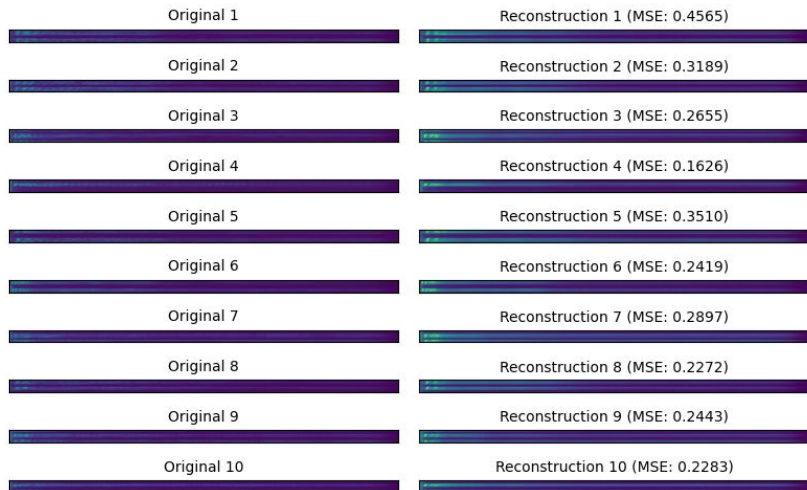


Spectrogram Reconstruction

$\beta = 1$ (200 epochs)



$\beta = 0.001$ (100 epochs)



MNIST Reconstruction

$\beta = 1$

Original 1



Reconstruction 1 (MSE: 0.0560)



Original 2



Reconstruction 2 (MSE: 0.0881)



Original 3



Reconstruction 3 (MSE: 0.0442)



Original 4



Reconstruction 4 (MSE: 0.0779)



Original 5



Reconstruction 5 (MSE: 0.0604)



$\beta = 0$

Original 1



Reconstruction 1 (MSE: 0.0005)



Original 2



Reconstruction 2 (MSE: 0.0009)



Original 3



Reconstruction 3 (MSE: 0.0026)



Original 4



Reconstruction 4 (MSE: 0.0007)



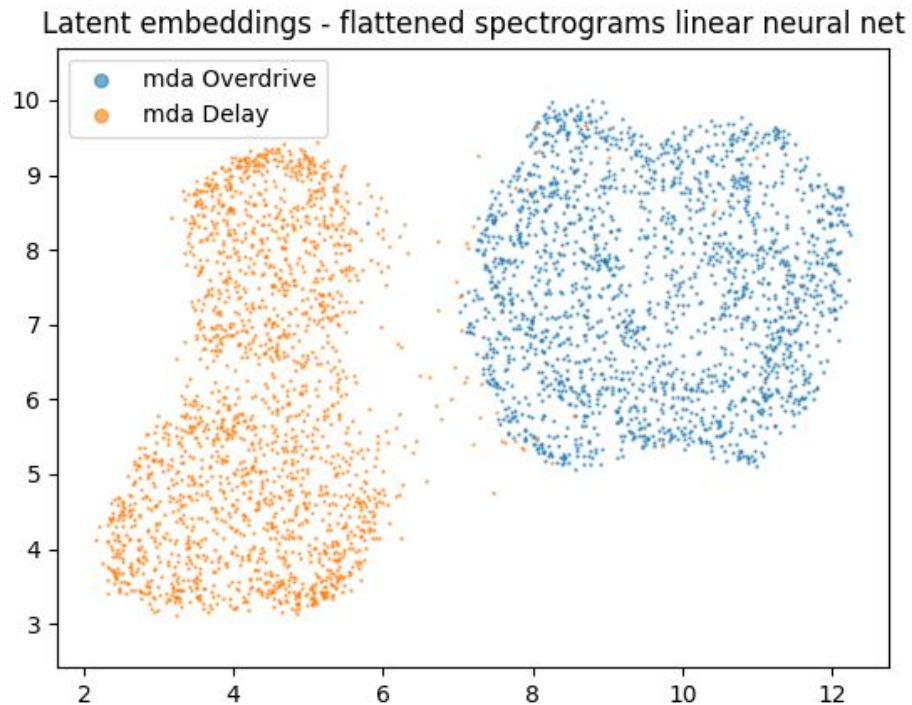
Original 5



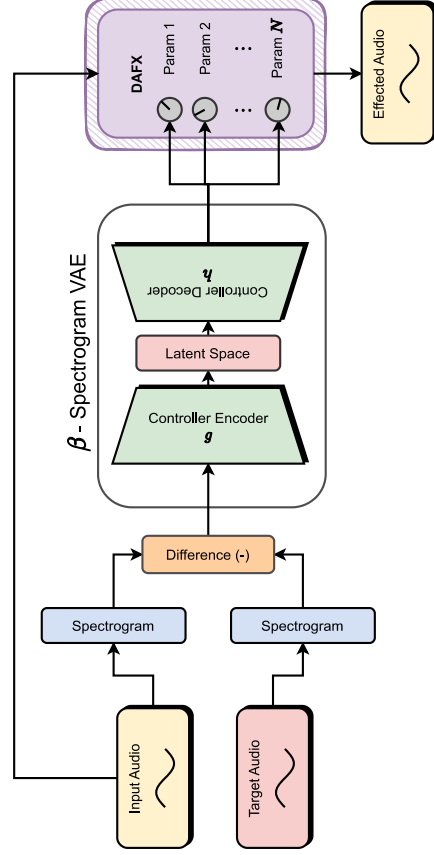
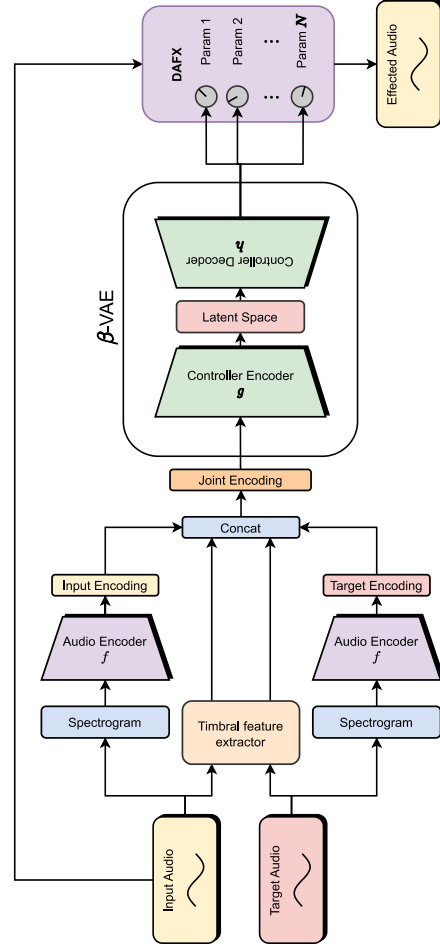
Reconstruction 5 (MSE: 0.0006)



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.