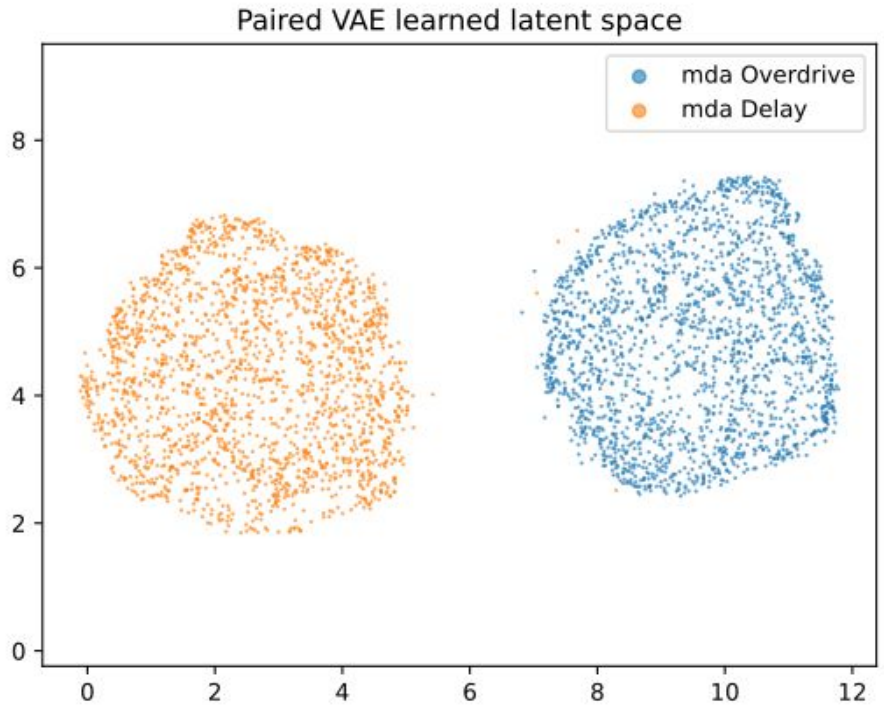
Week 6 Meeting

2357351G - MSci Half Project

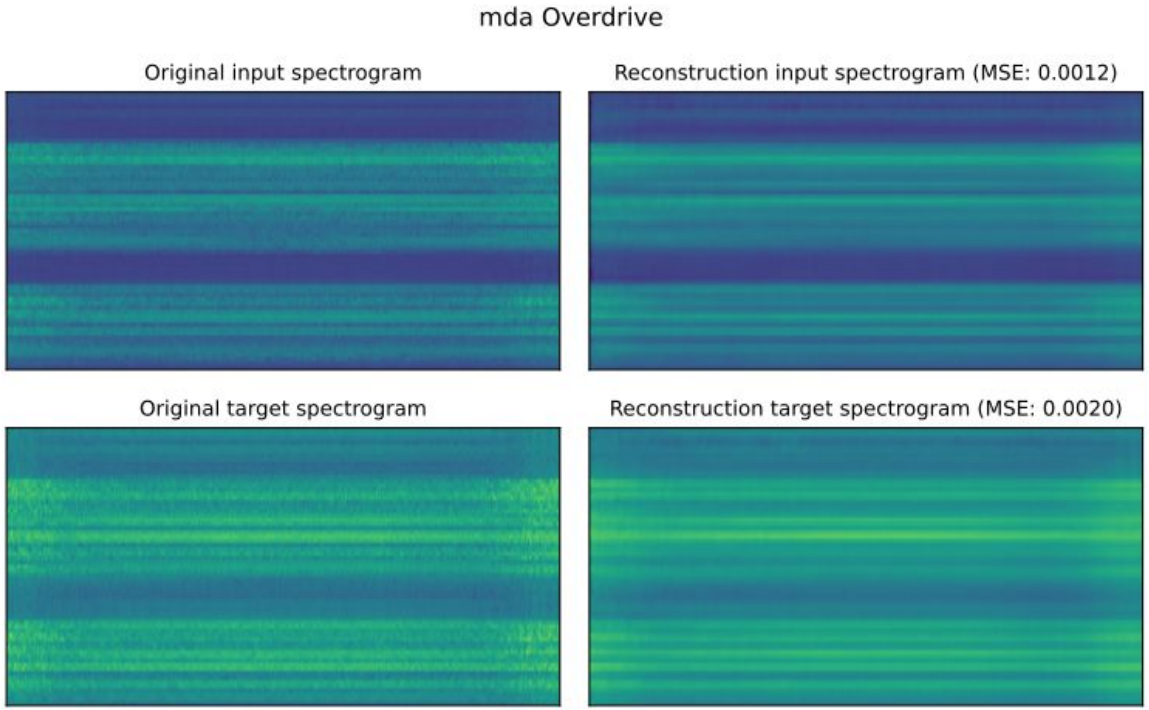
What I've done this week

- Updated Spectrogram-VAE model to take joint input-target audio pairs, stacked across channels.
- Trained Joint-Spectrogram VAE on VTCK dataset with Delay and Overdrive DAFX (fixed settings for each DAFX).
- Created visualisations for trained model including interpolation of latent space.
- Started implementation of simple end-to-end system, without bottleneck.
    - Ran short training session, had some issues with predicting correct settings with dummy values - will look more into this next week.

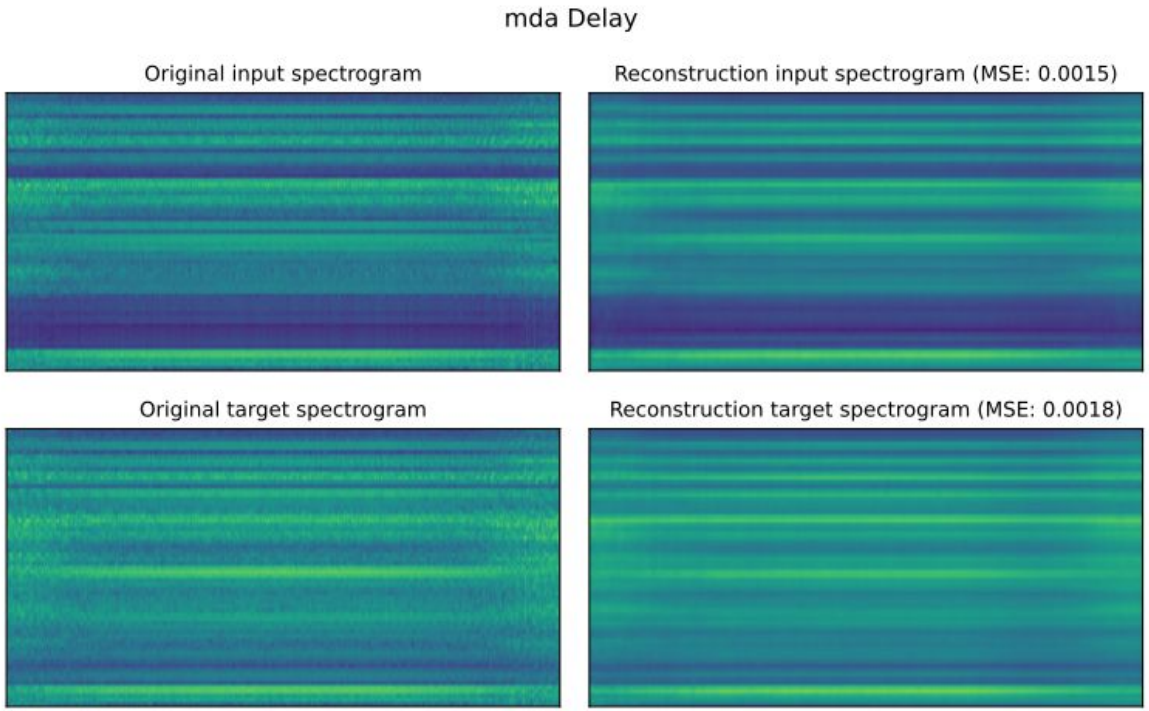
Projection of latent embeddings



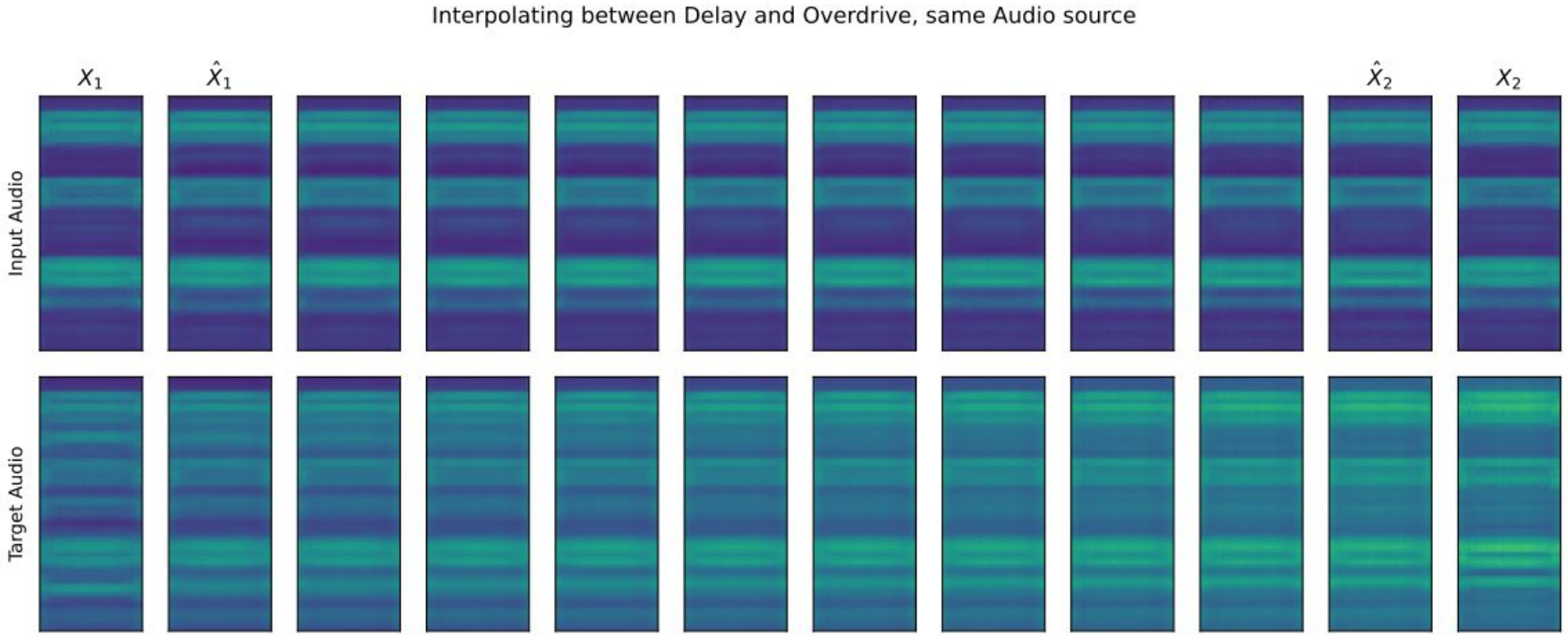
Spectrogram Reconstruction (Overdrive)



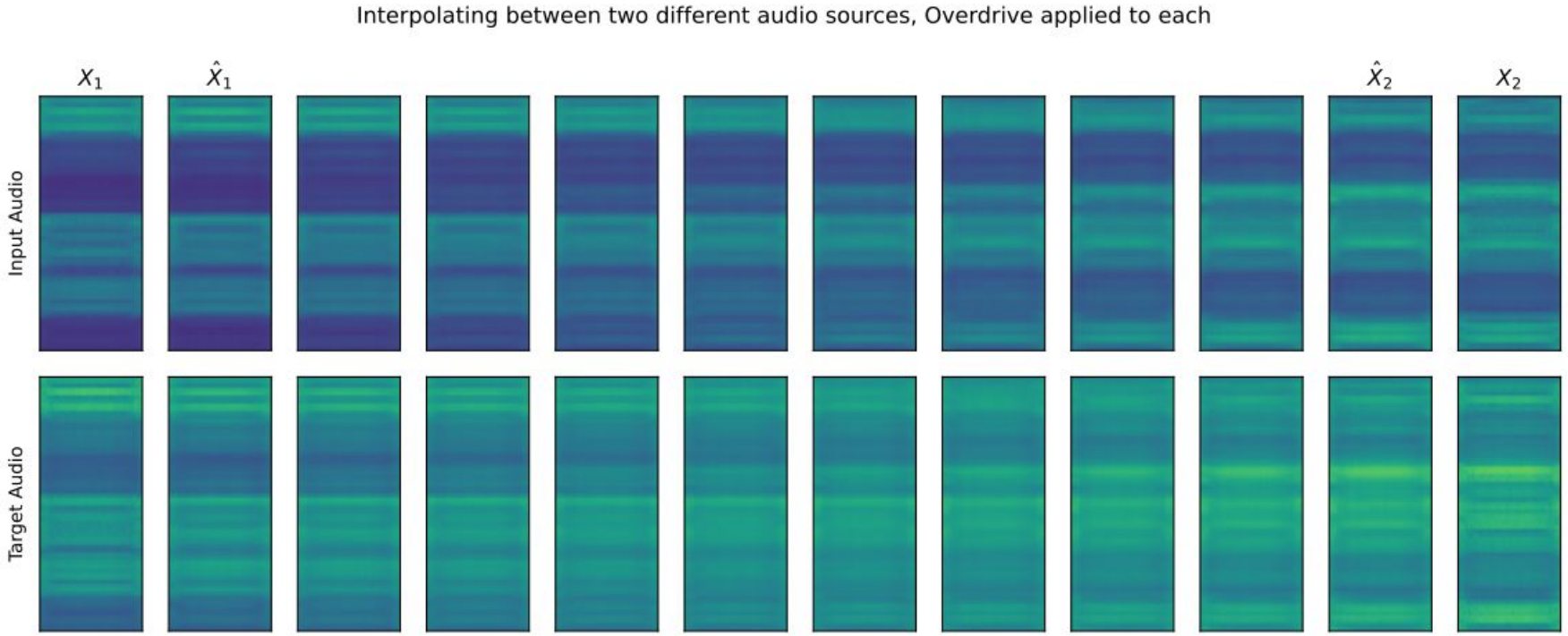
Spectrogram Reconstruction (Delay)



Interpolation (same audio, different FX)



Interpolation (different audio, same FX)



Questions

- Now that I have a simple trained Spectrogram-VAE - I'm wondering about the best method to map this down to a low dimensional (2-8D) latent space that can be 'played with' to find similar parameter settings. Would UMAP work, or should I have another bottleneck in a linear network and use similar sampling/KLD as in the autoencoder before mapping to parameter settings?

Plan for next week

- Debug end-to-end system.
- Run training on end-to-end system without bottleneck.
- Implement latent bottleneck for controller network.

Where I am in schedule

- Spectrogram VAE looks like it is performing a little better, should hopefully unblock me for the moment so I can work on the controller network.
- Hopefully end-to-end system should cause fewer issues with sensible embeddings.