Week 9 Report

Kieran Grant - 2357351G

What I've done this week

- Completed '*Generating Sound with Neural Networks*' tutorial series.
  - Implemented a VAE for MNIST speech dataset.
  - Uses STFT and encodes/decodes the (log)-spectrograms. ISTFT to transform generated spectrogram back to audio.
  - Used Weights and Biases during model training to track model configs/runs.
- Downloaded Style Transfer code.
  - Ran inference on some example audio - uses pretrained weights.
  - Downloaded the music audio dataset - yet to try preprocessing and training.
- Created plan for rest of the year and some targets.
  - Main aim is to have a modified version of the Style Transfer network which can use SPSA for numerical gradient estimation of arbitrary audio effects with a VAE replacing the controller network.
  - As well as this, a plan for the training data, losses, DAFX to be used and evaluation metrics.
- Started a doc to organise/group papers I've read.
  - Also started planing a 'narrative' for the interim report.
- Read *β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework (2017).*
  - Didn't have time to create summary/takeaways this week - will prepare this for next week.
- Updated architecture diagram with feedback from last week.
  - Appended to end of report.

Questions

- As far as I understand, if the latent space dimensionality using β-VAE is 'too-big' there will be dimensions which won't 'do much' - i.e. changing the values in these redundant dimensions take don't really change the reconstructed signal.
  - Is there a way of using this to tune the dimensionality of the latent space? I.e. a measure of the variance of the output w.r.t each dimension?

Plan for next week

- Perform preprocessing on audio dataset.
- Train the model on small subset of processed audio.
- Look at inference script to see how pre-trained weights are being loaded into the model.
- Pick library to use for loading in DAFX (Pedalboard or similar)
- Look at current SPSA implementation and whether it can be used directly with the above library.
- Write up summary/takeaways of β-VAE paper.
- Read *A Feature Learning Siamese Model for Intelligent Control of the Dynamic Range Compressor (2019)* and write up summary/takeaways.

Current state of project

- Have much clearer idea of the scope of the project.
- Aim to have a prototype model which is at least trainable by the end of the year.
  - Have a plan for work that needs to be done to achieve the above.

Architecture Diagram

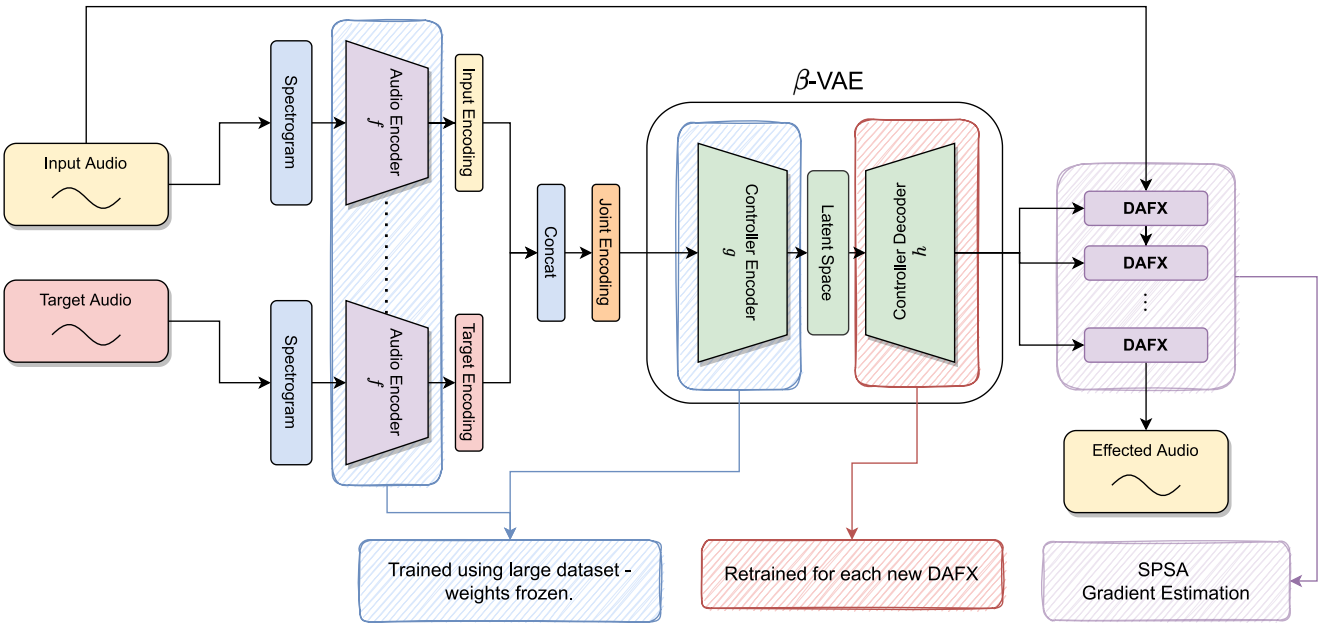

Figure 1: Updated architecture diagram from last week. The audio encoder (f) and controller encoder (g) networks are pre-trained with a large dataset and weights frozen. The controller decoder (h) network is retrained for each new effect (and model weights stored) which can be used for multiple style transfers.