Week 4 Meeting

2357351G - MSci Half Project

What I've done this week

- Finished the *Audio Signal Processing for Machine Learning* course.
- Read and summarised *AudioGen: Textually Guided Audio Generation (2023)*.
- Re-read *Differentiable Signal Processing With Black-Box Audio Effects (2021)* with a focus on how the gradient estimation for effect parameters work.
- Watched video giving an overview of how DALL-E/diffusion models for images work.
- Found an *AudioCommons* model which extracts various timbral attributes from an audio signal.
- Downloaded and played around with some IDMT datasets - Bass, Guitar, Piano, Drums and Audio Effects.

Questions

- The timbral descriptions I've seen in the dataset are limited to a vocabulary of around a dozen words. I was wondering if a preprocessing step for the dataset would be to add duplicates of signals with synonyms - e.g. replace "shrill" with "tinny" and have them as two separate instances of an effect.
  - Is this sensible?
  - Do you think that the language model would be able to handle this implicitly?
- Do you see any problems with using a model for automatically tagging the audio data with timbre characteristics?

Plan for next week

- Same creator of the *Audio Signal Processing for Machine Learning* course also has another course called *Generating Sounds with Neural Networks*. Culminates in a model which produces spoken digits (a bit like MNIST for computer vision). Planning to make a start on this this week.
- More in-depth reading into classification-free guidance and which language model might be appropriate.
- Spend time learning how adversarial losses, LSTMs and Transformers are implemented.
- I had started looking into how non-technical descriptive language about musical timbre can be quantified. I've added a couple of interesting looking papers to Zotero which I hope to have a look at this week.

Where I am in schedule

- Narrowed down project idea to text-to-audio-effect synthesis.
- Still getting to grips with handling audio data for deep learning.
- I imagine it might be at least another 2-3 weeks before I start any proper implementation of a model.
- First need to make sure I have a suitable dataset for training.