

Meeting 10/10/2022

Minutes

- Went through slides and discussed progress in the past week.
- Focussed on the AudioGen paper and how it works.
- Discussed various datasets/tagging methods that might be appropriate:
 - Asking music colleagues/friends to tag.
 - Mechanical Turk
 - AudioCommons
- Also went more into depth around gradient estimation method for Black-Box DDSP paper. However, still some uncertainty on the output measure of the model and format of the input.
- Need to check that what I'm doing even make sense - i.e. is there some quantifiable way to discuss these characteristics of sound?

To-do for this week

- Think about how the (Black-box) DDSP architecture can be used alongside word embeddings.
 - Need to get an understanding of which language models may be appropriate.
 - Should make some sort of high level diagram of what the architecture might look like.
- Have another look over the Black-Box DDSP paper and understand:
 - What is the output of ∇f ? I.e. the gradient estimation should spit out a single number.
- More research into *AudioCommons* model for tagging:
 - How does it work?
 - Is it valid?
- **Priority:** understand how descriptive language about sound qualities in music can be quantified? Is there any music theory research?
- Start the *Generating Sounds with Neural Networks* tutorial series.
- More in-depth reading about how classifier-free guidance, Transformers and LSTMs are actually implemented (maybe look at DALL-E mini code?).