

Week 2 Report

What I've done this week

- Created separate GitHub repository for dissertation.
- Read Hall of Fame papers and marking scheme for MSci project.
- Read and summarised *Differentiable Digital Signal Processing (2020)*.
 - Short summary of takeaways below.
- Started a summary document of different approaches to neural synthesis.
 - Strided Convolution Models (SING, MCNN, WaveGAN)
 - Autoregressive Waveform Models (WaveNet, SampleRNN, WaveRNN)
 - Fourier-based Models (Tacotron, GANSynth)
 - Oscillator models (vocoders, synthesizers)
 - Spectrogram models (SpecGAN)
 - Differentiable DSP models (DDSP)
- Added some more potential datasets and evaluation methods to relevant documents.
 - NSynth
 - MusOpen
 - Multi-scale Spectral Loss (MSSL)
- Watched/coded alongside YouTube tutorial series *Audio Signal Processing for Machine Learning*.
 - <https://www.youtube.com/playlist?list=PL-wATfeyAMNqlee7cH3q1bh4QJFAaeNv0>
 - Covers practical methods for feature extraction and common problems with audio data in ML.
- Read *Differentiable Signal Processing With Black-Box Audio Effects (2021)*.

What I didn't manage to do

- Haven't had time to make music in the past week.
- Not looked in to unsupervised tagging for data such as sample packs etc.

Development of research question

Lots of demos of audio effects (guitar pedals, plugins etc) available online. DDSP paper showed that reverb could be extracted from a violin (over many recordings) and transferred to another audio signal. Could there be a way to make a model which could accurately extract the 'effect' from a dry/affected signal comparison without the signal being one-to-one? Would be cool to be able to quickly try out effects or mimic effects on your own recordings/instruments. Need to think about:

- Dataset to use (YouTube A/B comparisons, SPICE models, anything else?).
- Inference time (if required online).
- If user would be able to control the effect at all?

Questions

- Overall marking is done by the supervisor and an assigned reader. Will this just be John and Johnathan, or will someone else also mark the paper?

- If there is lots of potential model training, are there facilities at the school that can be used (GPUs etc)?

Plan for next week

- Try to narrow down to a couple more fleshed out project ideas.

Current state of project:

- Still in the background research phase.

Main takeaways from DDSP paper

- The main advantage of the DDSP library presented in the paper is that it implements differentiable versions of classic DSP modules (filters, additive synthesis, reverbs). This allows for the parameters of these models to be trained end-to-end (rather than relying on expert knowledge to tune these from scratch) and for priors around how sound is produced to be factored in to the model (similar to convolution for images).
- A few applications are shown such as:
 - Taking a voice as input and outputting the same melody transformed to a violin sound.
 - The reverb from violin recordings being extracted and applied to a voice.
- Seems like this library could be very useful in any direction that the project might go.
- There wasn't a lot of discussion about real-time inference, and the audio signal has a low sample rate (16kHz). However, the final model had roughly 10x fewer parameters than comparable models and so should improve inference performance.
- Overall, it seems like this will be an important paper in the field and I've already seen a few papers which have implemented the library for audio effect synthesis.