Report

Ideas for project:

- **Virtual analogue (VA) modelling using a deep learning black-box approach.**
- Generating drum/synthesizer sound banks using generative deep learning models.
- *Discarded/tentative ideas*:
    - Deep learning to isolate instruments in music tracks/songs.
    - ML model which can generate impulse responses from images of spaces (cathedrals/studios etc).
    - Model to generate melodies from a chord sequences.
    - Lyric-to-melody ML model.
    - Creating sound samples from text descriptions (DALL-E for sound)
    - ML for optimising parameters in traditional DSP white-box methods

Research papers read (full read/written summary)

- *WaveNet: A Generative Model for Raw Audio (2016)*
    - Breakthrough DeepMind paper that produced an autoregressive, probabilistic model for generating raw audio. Has seen widespread use in speech synthesis and generative music.
- *A Review of Neural Network-based Emulation of Guitar Amplifiers (2022)*
    - Literature review of state-of-the-art in deep emulation of guitar amplifiers.
- *Adversarial Audio Synthesis (2019)*
    - WaveGAN/SpecGAN, used to produce short (~1 second) audio signals for sound effects/foley work.
- *Deep Learning for Tube Amplifier Emulation (2019)*
    - Uses WaveNet as a basis for modelling a Fender Bassman amplifier with user conditioning (gain control).
- *Efficient Neural networks for Real-time Analog Audio Effect Modelling (2021)*
    - Also uses TCN (WaveNet) but with large dilation factors to emulate the LA-2A dynamic compressor in real-time.

Research papers read (skimmed)

- *DDSP: Differentiable Digital Signal Processing (2020)*
- *Differentiable Signal Processing with Black-Box Audio Effects (2021)*
- *Identification of Volterra Models of Tube Audio Devices using Multiple-Variance Method (2018)*
- *Volterra Series and State Transformation for Real-Time Simulations of Audio Circuits Including Saturations: Application to the Moog Ladder Filter (2010)*

Other articles/videos/resources

- *Volterra kernel based sampling and the future of convolution audio software*
    - https://www.youtube.com/watch?v=h9-pMQzPqbo
    - Talk given by Acustica Audio who use Volterra kernel based sampling to model various audio effects.

- *Designing Audio Effect Plugins in C++ (2nd edition) - Will Pirkle*
    - Currently using this book to learn basic DSP and how to implement the algorithms in C++.
- *UA's Art and Science of Modelling AUD Plug-Ins*:
    - https://www.uaudio.com/blog/ua-plug-in-modeling-story/ (Part 1)
    - https://www.uaudio.com/blog/ask-doctors-ua-modeling-plug-ins/ (Part 2)
    - https://www.uaudio.com/webzine/2004/july/text/content2.html (older regarding Dynamic Convolution)
    - Universal Audio blog about their approach to VA modelling.
- Google's Magenta project:
    - https://magenta.tensorflow.org/studio
    - Wide range of applications of machine learning in the creation of music and art.
    - Many of the implementations are available as Ableton/Standalone plugins.
- *Deep Learning for Virtual Analog Modeling with Alec Wright*
    - https://www.youtube.com/watch?v=joMXK09-lUM
    - Interview with doctoral candidate researching deep VA modelling at Aalto Acoustics Lab in Finland.
- *Virtual Analog Audio Effects Simulation with JUCE, Ivan Cohen, JUCE Summit 2015*
    - https://www.youtube.com/watch?v=l_HHJdCKcjA
    - Ivan Cohen (Musical Entropy) talk at JUCE summit around VA modelling.
- *Julius Orion Smith's website*:
    - https://ccrma.stanford.edu/~jos/
    - Well known professor at Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University.
    - Great resource for DSP theory around audio.

Pre-ML VA modelling methods:

- White box:
    - Wave digital filters (WDF)
    - State-space models
    - Modified Nodal Analysis (MNA)
    - Post-Hamiltonian formalism
- Grey box:
    - Block-oriented structures (one model for linear part of system, one for non-linear)
    - Hammerstein and Wiener models
- Black box:
    - Dynamic convolution
    - Volterra kernels

Datasets available

- SPICE
    - https://en.wikipedia.org/wiki/SPICE
    - Program used to model analogue circuits, frequently used in white-box approaches to VA modelling.
- SignalTrain:
    - https://github.com/drscotthawley/signaltrain

    - https://zenodo.org/record/3348083
    - Very large dataset containing many hours of recordings using the LA2A dynamic compressor including full songs and individual instruments/voices.
- Freesound:
    - https://annotator.freesound.org/fsd/
    - Large collection of annotated audio snippets.
    - Sometimes used alongside SPICE (clean DI'd guitar from FreeSound through SPICE simulation)
- IDMST-SMT-Guitar/Bass/Drums/Chords/Audio Effects:
    - https://www.idmt.fraunhofer.de/en/publications/datasets/guitar.html
    - https://www.idmt.fraunhofer.de/en/publications/datasets/bass.html
    - A number of very large dataset comprising of DI recordings of various instruments using different playing techniques and configurations of the instrument.
- AES Guitar Amplifier Sounds Dataset:
    - https://www.aes.org/e-lib/browse.cfm?elib=19754
    - Dataset contains five different styles of guitar sounds passing through different guitar amplifiers with 10 steps of a gain parameter.
    - Input/output also provided in matrix form.
- Other:
    - Some papers use custom datasets, either using physical analogue equipment or SPICE simulations.
    - These will be used with DI guitar/bass recordings or function generators.
    - For drum sounds, there are thousands of sample packs available online which are generally organised into categories (snare, kick, fx...) which would be suitable for training.

Evaluation methods

Losses/objective functions

- Mean average error (MAE), mean squared error (MSE) and variants
- Error-to-signal ratio (ESR)
    - Usually used alongside pre-emphasis filters (first-order HP/LP, folded differentiator) to account for real-world perceptual accuracy.
- Short-term Fourier Transform (STFT)
    - Has also been combined with MAE

Real-time capabilities

- Number of operations
- Timing inference
- Real-time factor ($RTF = \frac{\text{Processing Time}}{\text{RT Constraint}}$)

Listening tests

- MUSHRA
    - https://en.wikipedia.org/wiki/MUSHRA
- Aural comparison of prediction and target
- Web Audio Evaluation Tool
    - https://github.com/BrechtDeMan/WebAudioEvaluationTool