Daniel Faronbi Honors Thesis Proposal

Synthesizer Parameter Approximation by Deep Learning

Keywords: Timbral Feature Extraction, Linear Arithmetic Synthesis, Machine Learning

Abstract

Synthesizers have been an essential tool for composers of any style of music including computer generated sound. They allow for an expansion in timbral variety to the orchestration of a piece of music or sound scape. Sound designers are trained to be able to recreate a timbre in their head using a synthesizer. This works well for simple sounds but becomes more difficult as the number of parameters required to produce a specific timbre increase. The goal of this research project is to formulate a method for synthesizers to approximate a timbre given an input audio sample using deep learning. The synthesizer should be able to modify its settings (oscillators, filters, LFOs, effects, etc.) to produce an audio signal as close to the input sample as possible. A cost function will measure the difference between the outputted audio signal from the learned synthesizer parameters and the original audio signal that is being mimicked.

Introduction

Over the years, a variety of machine learning approaches have been used to approximate specific timbres. For the most part, this approach has been used to model acoustic instruments [1-2]. However, this project aims to use these techniques to recreate sounds that were original generated by a computer. The lack of robust research in this area means that there is room to explore multiple aspects of parameterized deep learning for sound synthesizers. InverSynth, a recent project completed in 2019, attempted this same idea [3]. However, only subtractive and frequency modulation synthesis techniques were used. For this project, the eventual goal is to expand the current literature by exploring synthesizer parameter learning with additive, wavetable, and linear arithmetic synthesis. InverSynth used an open source java libray, JSyn [4] to programmatically produce audio. Unfortunately, that software will not work for this project because of the added synthesis methods. As such, a custom synthesizer will be created using the JUCE framework [5]. For this semester, the goal is to replicate the InverSynth project, focusing particularly on using parameter learning with subtractive synthesis to learn an audio signal

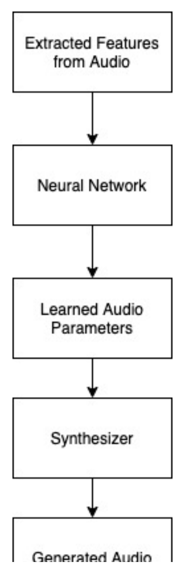


Figure 1 - Deep Learning Approach

Methods

This project will employ a variety of novel techniques in the construction of the features used. It has been observed that many music information retrieval deep learning projects focus on change in learning models rather than use of appropriate features [6]. To counter this common mistake, a significant amount of time will be spent analyzing a variety of features, so that the most appropriate features will be selected to train the neural network.

Different activation functions within the neural network will be tested including sigmoid, hyperbolic tangent, hard hyperbolic, SoftMax, and many more [7]. The varying activation functions can incorporate the best of logistic regressions styles of learning with linear regression styles of learning to model a neural network and produce the optimal result.

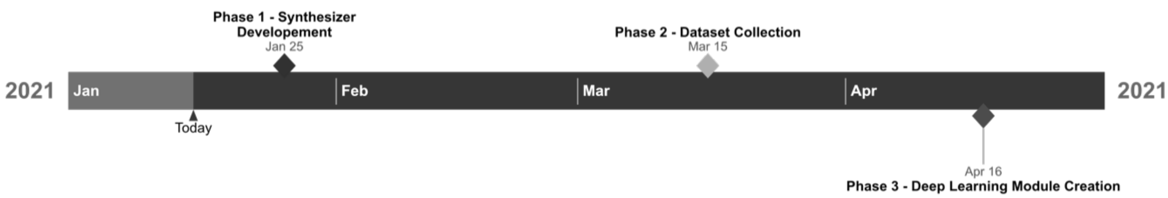
Plan/Timeline

The proposed duration of the project is 5 months. The project will be split into three different phases: synthesizer development, data set generation, and deep learning module creation.

The first phase of the project will focus on the development of the synthesizer. This will allow for the parameters to be changed programmatically. As mentioned before, the JUCE framework will be used to create this software.

The next phase will be used to gather training datasets. These will be a variety of audio samples generated from popular synthesizers with varying timbral complexity. “Low” complexity samples will be simple periodic waveforms and “high” complexity samples will be sampled audio from popular electronic songs.

The final phase will be used to develop the first iteration of the deep neural network using tensor flow [8].



Conclusion

This research plan focuses on using cutting edge machine learning approaches to music information retrieval. Based on a review of current literature, more research in this area is needed. The knowledge generated from this project would be very beneficial to others involved in music information retrieval and can be applied to other digital signal processing projects. The novel technology developed from this product makes it ideal for a product. This adds economic value to society as well as enjoyment to consumers. The new technology also allows for other products to improve by adding options to developers of similar software. These strengths make this project ideal for an honors thesis.

[1] Riionheimo, J., Välimäki, V. Parameter Estimation of a Plucked String Synthesis Model Using a Genetic Algorithm with Perceptual Fitness Calculation. EURASIP J. Adv. Signal Process. 2003, 758284 (2003). https://doi.org/10.1155/S1110865703302100

[2] Itoyama, Katsutoshi, and Hiroshi G. Okuno. "Parameter estimation of virtual musical instrument synthesizers." ICMC. 2014.

[3] Barkan, Oren, et al. "InverSynth: Deep Estimation of Synthesizer Parameter Configurations From Audio Signals." IEEE/ACM Transactions on Audio, Speech, and Language Processing 27.12 (2019): 2385-2396.

[4] http://www.softsynth.com/jsyn

[5] https://juce.com/

[6] Humphrey, Eric J., Juan Pablo Bello, and Yann LeCun. "Moving beyond feature design: Deep architectures and automatic feature learning in music informatics." ISMIR. 2012.

[7] Nwankpa, Chigozie, et al. "Activation functions: Comparison of trends in practice and research for deep learning." arXiv preprint arXiv:1811.03378 (2018).

[8] Abadi, Martín, et al. "Tensorflow: A system for large-scale machine learning." 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). 2016.