Challenges in Modeling Audio for Realistic Music Generation

Sam Hinds
Texas State University
San Marcos, Texas
CS434

KEYWORDS

datasets, neural networks, ADAs, autoregressive models, long range correlation, audio signals

1 INTRODUCTION

When attempting to recreate organic sounds or instrumentation a lot of truly continuous parameters of physics are at play. In this theory-oriented approach I will attempt to deconstruct and analyze current methods of generating realistic forms of music in the audio domain and review the current state of progress in this field of generative modeling.

2 PROBLEM DESCRIPTION

Typically in generating music or other forms of sound in a computational sense we rely on high level representations of what the real physical sound is intended to indicate. MIDI data to some degree is able to replicate a various forms of live performed music but still lacks a degree of context and accuracy in subtle details. Because audio signals are typically sampled above 16hz the size of the problem was thought to be too big of a problem.

Details such as timbre, velocity, and sustain are currently examples of low level structures that can be generated well but capturing more high level elements involved is a much harder task. This problem has recently started to become something that can be solved thanks to progress in Auto regressive modeling allowing us to generate realistic forms of audio wave forms rather than the context-lacking forms of MIDI data or spectograms.

3 CURRENT RESEARCH

Current research involving models capable of generating audio signals through autoregressive models is only recently starting to gain traction as using the (AR) model can better predict audio waveforms in each individual timestep at a time. We are currently reaching the point where (AR) models are more commonly used as text-to-speech models which are easier to generate in comparison to the dynamic nature of music.

4 PRELIMINARY PLAN

Currently my approach to exploring solutions to this issue will be to first conduct more in depth research on (AR) models and understand how their structure is the most logical approach to working well with this specific issue. I will then conduct various experiments from examples of this model based off the paper on advances in neural information processing I have to chosen to research *The challenge of realistic music generation: modelling raw audio at scale (2018)* in order to determine how effective these advances in audio generation have been in practical application.

I am also planning to do a good amount of research into the stacking of of hierarchical unconditional models in order in order to sample code sequences and turn them into rendered audio. These samples are then quantitatively compared in signal fidelity and musicality.

5 REFERENCE

DIELEMAN, S. VAN DEN OORD, A. SIMONYAN, K. *The challenge of realistic music generation: modelling raw audio at scale (2018)*

- MILLS, TERENCE C. *Time Series Techniques for Economists. Cambridge University Press. (1990).*