Bowed Stringed Instrument Raw Audio Synthesis with Generative Neural Networks

DeVries, Benjamin (School: Woods Cross High School)

Neural networks are utilized for machine learning. This project attempted to recreate the sound of stringed instruments similar to the violin using neural networks. Current computer generation techniques yield unrealistic results. The major limitation in this project was computation power. As a result, lower quality 16000 Hz 8-bit audio was used. Furthermore, a Generative Neural Network (GNN) with Dilated Convolutional layers was used in order to work directly with raw audio, which is a difficulty with traditional neural networks. The GNN was modeled on a current text-to-speech network. The success of the network was measured with a "loss" function, with lower loss corresponding to better results. After initially training the network on 306 notes simultaneously, it was decided that that this simultaneous method required too much computation time to be feasible, and the loss function had not dropped low enough to yield good results. As a second attempt, the network was trained on single notes at a time, and this time yielded much better results. The network took about 2 days (8000 training steps) of training for a single note, and the loss was .001 at the end, instead of approximately 2.5 from the previous attempt. The results demonstrated that the network was capable of generating identifiably stringed instrument notes. Future networks with revision could theoretically generate entire musical phrases, and more computation power would yield better products. Applications include electronic keyboards, the recording and film industry, as well as artificial voicing and wherever sound needs to be generated realistically.