Shiv Nadar University

MUSIC COLLAB: An IoT and ML Based Solution for Remote Music Collaboration

Nishtha Nayar, Divya Lohani
Shiv Nadar University, Greater Noida, India

Abstract

- Communication using mediums like **video and audio** are essential for various professions.
- This project looks into the interaction with **real-time audio transmission** using the tools in the domains of IoT and machine learning.
- Transport layer protocols, **TCP** and **UDP** are examined for audio transmission quality.
- **RNN models** are examined for their efficiency in **predicting music** and being used as a **substitute** in case of loss of packets during transmission.

Introduction

- The **entertainment industry** is growing at a rapid rate, so is the need for collaboration.
- There are a lot of **financial and logistical challenges** for artists and musicians who're seeking to collaborate across geographical boundaries.
- **The motivation** is to make use of the advancements in IoT and ML to help the creative process of musicians across the globe.
- There have been applications like Jacktrip [1] which aim to enable audio collaboration. Researchers at CCRMA have also tried to analyze and test audio transmission over physical distances [2]. Further, researchers at Stanford have studied various ways of music composition using Naive Bayes and neural network models [3].
- This use case can further be applied to areas like real-time video game streaming and similar multimedia applications.

Methodology

Transport Layer

- For the transport layer, five different server-client cases using TCP and UDP for audio transmission were coded using **Java** socket programming. These cases were then run on several machines and the packets were captured by **Wireshark.**

Machine Learning

- **Three RNNs – Basic, Lookback and Attention** were trained on Google environment Magenta. The flow of the models is depicted in Figure 1. These models were then used to predict 3 bars of a MIDI primer – 'River Flows in You' and then evaluated as a substitute for music produced by humans.

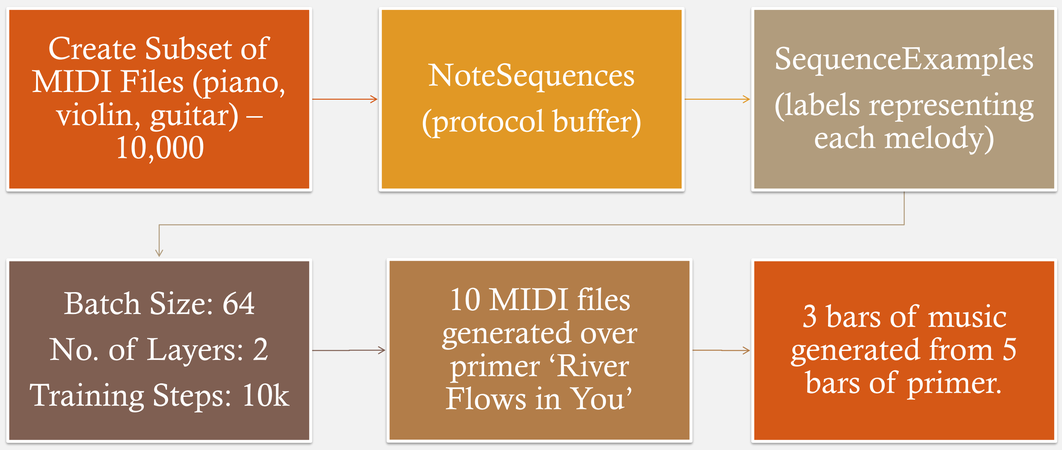

Figure 1.

Results

- The two protocols were compared using parameters: **RTT, Throughput, Latency and Packet-Loss** from the graphs obtained (Figure 2 & 3).
- As per Table 1, it was easier to capture and analyze packets for TCP than for UDP. We were not able to study UDP properly with the given tools.
- **TCP** hence seems as a better fit on the basis of the analysis and metrics.
- Performance over TCP can be further improved for seamless transmission.
- The three models were evaluated using **TensorBoard** (Figure 4 & 5). As per Table 2, the numerical metrics are almost the same.
- Basic RNN gave gaps, repetitive melodies. **Attention and Lookback** RNNs gave better sounding and more consistent melodies.

Table 1.

	Case 1	Case 2	Case 5
Max Latency (packets/s)	8	7000 (sudden increase)	290 (fluctuates)
Max RTT (ms)	55	0.09 (fluctuate initially)	0.06
Max Throughput (bits/s)	2100	900	8000

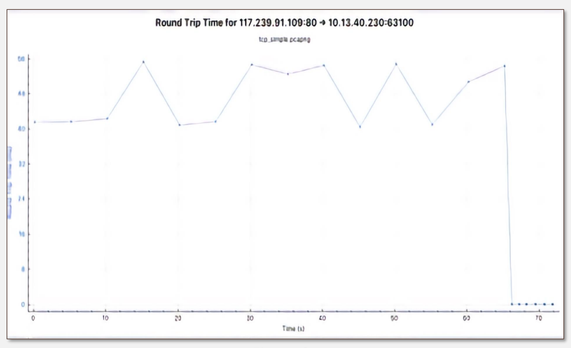

Figure 2.

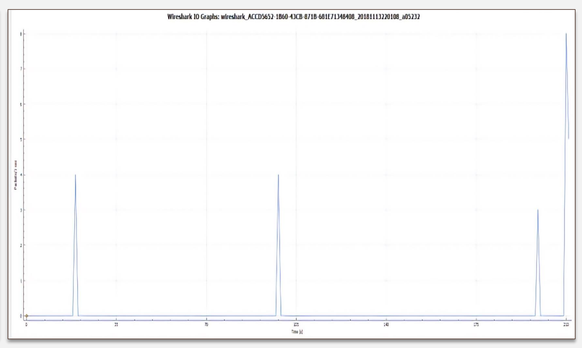

Figure 3.

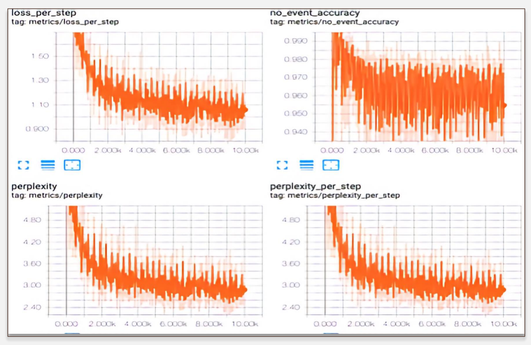

Figure 4.

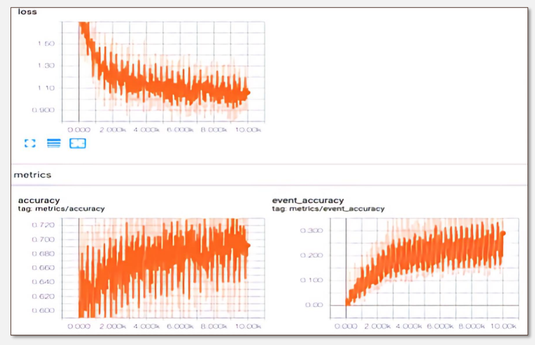

Figure 5.

Table 2.

	Basic RNN	Lookback RNN	Attention RNN
Min Loss	0.77	0.91	0.99
Max Accuracy	0.7315	0.7321	0.7131
Min Perplexity	2.5	2.55	2.71
Range of log-likelihood	-260.5 to -171.3	-274.2 to -186.1	-280.4 to -232.9

Discussion

- TCP preferred over UDP.
- Machine learning through RNN provides decent generated melodies which can be used as substitute.
- Numerically, the training accuracy for basic RNN is maximum and it's loss is minimum. However, the difference is miniscule.
- According to experimental data and evaluation of the generated music by real musicians, Attention and Lookback RNN are better than Basic RNN.

Contact

Nishtha Nayar
Shiv Nadar University
Email: nn469@snu.edu.in
Phone: +91-8286242688

References

1. Cáceres, J., and Chafe, C. 2010. Jacktrip:Under The Hood of An Engine for Network Audio. *Journal of New Music Research*, 39:3, 183-187. doi.org/10.1080/09298215.2010.481361.
2. Chafe, C. 2003. Distributed Internet Reverberation for Audio Collaboration. In *Proceedings of Twenty Fourth Conference on Advances in Engineering Software.* Banff.
3. Kang, D; Kim, J.; and Ringdahl, S. 2018. Project Milestone: Generating Music with Machine Learning. Technical Report, Department of Computer Science, Stanford University, CA.