# Improvising with MANDI the AI Drummer

Patrick Hutchings Faculty of Information Technology Monash University Melbourne, Australia patrick.hutchings@monash.edu Toby Gifford Faculty of Information Technology Monash University Melbourne, Australia toby.gifford@monash.edu

#### Abstract

The Musically Attentive Neural Drum Improviser (MANDI) is a system that generates drum patterns in real-time to accompany improvised performances with human musicians. MANDI is powered by a Temporal Convolutional Network trained on recorded duet improvisations of professional musicians. It demonstrates responsive musical behaviour when used by improvising musicians, useful for practice and performance. We discuss the system design and present recorded performances of MANDI in use.

## **1** Introduction

Music improvisation involves a combination of skills that are difficult to reproduce algorithmically. Automated music composition, co-operative agency, responsive behaviour and expressive performance are individually complex to model, and particularly so in combination. The lofty ambition of developing musical AI systems that are in some sense 'creative equals' with their human collaborators continues to drive several research programs [1]. However, there is also an important role for 'somewhat intelligent' systems in music improvisation [2, 3]. By providing the improviser with dynamic inspiration and a shifting framework to build their solos around, such systems can stimulate and surprise both audience and creator [4].

When designing music systems for use with human performers, the selection of constraints and learning techniques applied should be considered in the design and collection of data as well as system architecture [5]. The Continuator [6], Reflexive Looper [7] and Jambot [8] have demonstrated the performance and practise-support potential of improvisation systems that work alongside human performers, when developed with performance data.

Deep learning techniques have been used successfully in automatic music generation based on large music corpora [9] and drums specifically [10, 11]. However, the large amount of data required to train deep learning networks has meant that music applications of deep learning typically rely on written scores or recordings of pre-composed works. While patterns useful in the composition of music may be learned from these datasets, specific characteristics important to the development of collaborative, spontaneous composition seen in improvised performances are never exposed to the system.

# 2 MANDI

In order to leverage the power of deep learning for an improvising music agent, we developed a drum improviser – MANDI – by first creating a dataset of improvised performances between pairs of musicians: one drummer and one melodic instrumentalist. This dataset was then used to train a Temporal Convolutional Network (TCN) [12, 13] to attend to musical input and produce drum patterns for new performances.

Workshop on Machine Learning for Creativity and Design (NeurIPS 2018), Montréal, Canada.

MANDI was designed to generate drum patterns taking into account both its own recent history of musical output and input from an improvising instrumentalist. By constraining MANDI to a single tempo (120 beats per minute), quantised to 12 clicks per beat, the learning capabilities of an artificial neural network were focused on rhythmic interplay and real-time responsiveness with musicians.

#### 2.1 Data Collection

Two melodic instrumentalists and two drummers performed short improvisations with a 120bpm click-track in separate drummer-instrumentalist paired sessions. Sessions from the four pairings produced 150 minutes of data, recorded in MIDI format. Drummers used Roland TD-50 electronic drum-kits that recorded stroke positions on drums and cymbals as well as note timing and velocities. Onset, offset and velocity of notes performed by instrumentalists were added to this data.

#### 2.2 System design

MANDI samples the output of a trained TCN, converts selected notes to MIDI messages and sends them to a Roland TD-50 drum physical synthesis module for performance. The TCN is fed three measures of quantised performance data, encoded as text strings, from both MANDI and the human improviser, and generates one measure of drum patterns to add to the performance.

The TCN had seven hidden layers with dilation in each layer to produce an effective memory covering the 196 timesteps of input sequences. The network was designed to achieve optimum predictive accuracy with inference time less than a single timestep (4ms) on a desktop computer with GTX 1080 GPU. Softmax activation was used on the final network layer to treat outputs as a categorical distribution and sampled to generate performance data. The probability of any of the 451 words in the corpus dictionary being played at any timestep in the upcoming measure was used to make a weighted selection.

# 3 Results

A per token perplexity of 5.87 was achieved on the (15%) validation set after 64 epochs of training with batches of 32 sequences.

Recorded performances using MANDI can be heard at: https://goo.gl/1V2Y53 with full hyperparameters listed.

While an extensive hyper-parameter search was conducted, the most musically significant changes were observed through adopting different sampling strategies from the output layer.

Multiplying tensors by a scalar before the final softmax activation during inference affected entropy and was used as an effective 'predictability dial' resulting in performances with compositional structure not learned by the TCN due to its limited memory. (See recordings).

Currently applied constraints in the form of quantisation, use of a click track and short memory prevent the system from learning a range of important aspects of improvised performance. Within these constraints, however, the system demonstrates interesting and nuanced performances by utilising a wide spectrum of velocity and stroke positions in contextually appropriate ways.

MANDI is able to attend to these details, not available in standard MIDI datasets while having real-time responsiveness not currently possible with audio-based models. In this way, MANDI demonstrates the effectiveness of capturing detailed performance data from improvised performances for training intelligent music improvisation systems.

## 4 Future Work

To relax some of the constraints currently palced on MANDI, a significantly larger dataset is desirable. Longer sequence lengths and increased variation from micro-timing that come from finer quantisation makes sequence prediction difficult with small datasets. We are currently expanding the dataset by recording 50 hours of duet improvisations to assist training with higher temporal resolution.

#### References

- [1] Jon McCormack and Mark d'Inverno. Designing improvisational interfaces. In *Title: Proceedings of the 7th Computational Creativity Conference (ICCC 2016). Universite Pierre et Marie Curie*, 2016.
- [2] Toby Gifford and Andrew R Brown. Cybernetic configurations: characteristics of interactivity in the digital arts. 2013.
- [3] Oliver Bown. Performer interaction and expectation with live algorithms: experiences with zamyatin. *Digital Creativity*, 29(1):37–50, 2018.
- [4] Andrew R. Brown, Toby Gifford, and Bradley Voltz. Stimulating creative partnerships in human-agent musical interaction. *Comput. Entertain.*, 14(2):5:1–5:17, January 2017.
- [5] Toby Gifford, Shelly Knotts, Stefano Kalonaris, Jon McCormack, Matthew Yee-King, and Mark d'Inverno. Computational systems for music improvisation. *Digital Creativity*, 29(1), 2018.
- [6] François Pachet. The continuator: Musical interaction with style. In *Proceedings of ICMC*, pages 211–218., Göteborg, Sweden, 2002. ICMA.
- [7] François Pachet, Pierre Roy, Julian Moreira, and Mark d'Inverno. Reflexive loopers for solo musical improvisation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2205–2208. ACM, 2013.
- [8] Toby Gifford. Appropriate and complementary rhythmic improvisation in an interactive music system. In *Music and human-computer interaction*, pages 271–286. Springer, 2013.
- [9] Jean-Pierre Briot, Gaëtan Hadjeres, and François Pachet. Deep learning techniques for music generation-a survey. *arXiv preprint arXiv:1709.01620*, 2017.
- [10] Patrick Hutchings. Talking drums: Generating drum grooves with neural networks. *arXiv preprint arXiv:1706.09558*, 2017.
- [11] Dimos Makris, Maximos Kaliakatsos-Papakostas, Ioannis Karydis, and Katia Lida Kermanidis. Conditional neural sequence learners for generating drums' rhythms. *Neural Computing and Applications*, pages 1–12, 2018.
- [12] Colin Lea Michael D Flynn René and Vidal Austin Reiter Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *IEEE International Conference on Computer Vision* (ICCV), 2017.
- [13] Shaojie Bai. Sequence modeling benchmarks and temporal convolutional networks (tcn), April 2018. https://github.com/locuslab/TCN.