Performing with a Generative Electronic Music Controller

Charles Patrick Martin[1]

[1] *The Australian National University, Canberra, Australia*

Abstract

Generative electronic music is, by and large, old news; however, despite ever more convincing composition systems, less progress has been made in systems for live performance with a generative model. One limitation has been the focus on symbolic music, an imperfect representation for musical gesture, another has been the lack of interactive explorations of co-creative musical systems with modern machine learning techniques. In this work these limitations are addressed through the study of a co-creative interactive music system that applies generative AI to gestures on an electronic music controller, not to creating traditional musical notes. The controller features eight rotational controls with visual feedback and is typical of interfaces used for electronic music performance and production. The sound and interaction design of the system suggest new techniques for adopting co-creation in generative music systems and a discussion of live performances experiences put these techniques into practical context.

Keywords

interactive music system, mixture density recurrent neural network, performance

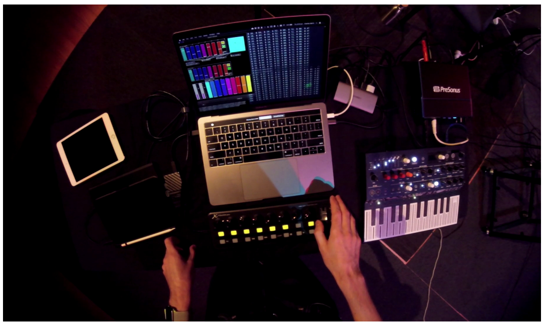

Figure 1: Performing with the generative electronic music controller. The Behringer XTouch Mini (lower centre) is the main musical interface for this system while the laptop screen shows the synthesiser and generative system state. In the performance, both the performer and generative system have control over the eight knobs of the controller. A performance video is available at: https://youtu.be/upHSIpiGYVg

1. Introduction

Generative music is well-established as a component of contemporary composition, with proponents in the experimental music scenes of the mid-20th Century among other earlier examples [1]. Current explorations of deep neural networks for generating music [e.g., 2, 3, 4] are enjoying success in terms of convincing output, but, perhaps, not in terms of application where much simpler rule-based music generators are more common. Two issues facing musical performance with generative AI are that such systems tend to generate symbolic music, ignoring the gestural and non-note-based aspects of present electronic music performance, and that co-creative interactions for musical AI systems have not been explored to the same extent as the generative models. To address these issues, different types of musical models must be explored, and the interactions between performers and generative models must be considered as a first-class problem.

In this work, a somewhat different kind of musical AI system is presented: a physical electronic music controller with eight knobs allowing direct control over a synthesiser program is backed by an AI system that attempts to continue interactive gestures from the human performer using the predictions an artificial neural network. This system explores an approach to embodied co-creation, where interactive gestures, rather than musical notes, are generated and collaboration of performer and generative model are expressed through live performance[1].

Rather than a note-driven aesthetic, the musical context is improvised electronic sound with gestural control over synthesis parameters. The neural network has been trained on this gestural performance data, collected from the controller during rehearsals and performances, to predict the next interaction, both in terms of quantity of controller movement, and the amount of time before this movement should occur. This sets this work apart from other interactive generative music systems related to music production [5] and MIDI-note performance [6] as well as non-neural-network systems such as *Continuator* [7],

[1] A video of this system in performance can be found at https://youtu.be/upHSIpiGYVg

Joint Proceedings of the ACM IUI Workshops 2022, March 2022, Helsinki, Finland

charles.martin@anu.edu.au (C. P. Martin)
https://charlesmartin.com.au (C. P. Martin)
0000-0001-5683-7529 (C. P. Martin)

or *Voyager* [8] that generate MIDI notes. While gestural predictions have been studied in a minimal musical instrument [9], this work involves a more complete musical interface capable of driving a complete performance.

Throughout performance with this system, the neural network can take control of the interface, continuing the performer's actions, transforming them into a "predicted reality", or overriding the performer in real-time. The performer can see these actions represented visually on the controller interface and must tune their inputs to guide the neural network towards musically acceptable behaviours. The goal is to set up a feedback loop between human and generative neural network model where the process of co-creation leads to transformed interactive experiences [10].

This work is part of an ongoing process of artistic research studying how a ML model might evolve over time as part of a computer music practice. Over the development of this work, the ML model has been re-trained as more training data has been collected. The affordances of the neural network change (sometimes dramatically) when it is re-trained with more or different data. This changes the possible interaction between performer and instrument and demands negotiation and improvisation from the performer in each performance to learn and exploit new behaviours. The instrument itself is an experiment in co-creation. Through it, this work highlights the tension between the machine learning algorithm's role as a component within a musical instrument, and as a distinct agent that shares musical control with a human performer.

2. Generative AI System

This system uses a mixture density recurrent neural network (MDRNN) within the context of a live computer music performance. This algorithm is a variant of the deep neural networks often used to compose text or symbolic music but allows learning and creative generation of continuous data such as synthesiser control signals, and absolute time values.

The generative aspects of this system use the Interactive Musical Prediction System (IMPS) [11] which implements the MDRNN in Python. In this context, the MDRNN is configured with two 32-unit LSTM layers and an MDN layer that outputs the parameters of a 9-dimensional Gaussian mixture model: one dimension for each knob on the controller and one dimension for the number of seconds in the future that this interaction should occur. The input to the MDRNN is, similarly, a 9D vector of the location of each knob and the time since the previous interaction. Although this is a tiny model by comparison with other deep learning models, it is appropriate given the size of the dataset involved and the strict time requirements for an interactive application.

This ML model learns to reproduce how a human plays a musical instrument in terms of physical movements rather than what notes should come next. As a result, this musical ML configuration could be termed embodied musical prediction. This style of musical ML is ideal for application in a live electronic performance system, where embodied musical gestures with a new interface are often more important that traditional musical notation.

3. Sound and Interaction Design

The synthesised sounds are created by eight sound generators, each operated by one knob of the controller. Two sound options are available, a sine-tone oscillator and a looped sample player (granular synthesiser), these can be switched by clicking a knob. Turning each knob changes the main parameter of each sound generator, these are the oscillator pitch or looped sample section depending on which sound option is selected.

Each sound generator has volume set to zero (silence) by default, but changing the main parameter triggers a short volume envelope (a note). The buttons below each knob allow additional control over each generator's volume: the top button triggers the short envelope without changing the main knob and the bottom button turns the sound on continuously.

The eight sound generators are mixed together and sent through distortion and reverb effects which can be controlled through the computer interface. The large slider controls the main volume allowing the performer to start and end the performance. The sound design and MIDI interfacing with the XTouch Mini is implemented in Pure Data which runs on the performer's laptop.

The knobs controlling the synthesis tuning parameter are the main focus of the performance and it is this part of the system that is controlled by both the performer and generative AI system. The LED indicators on each knob show the latest update to the parameter, either from the performer turning the knob or the generative system adjusting it in software.

The IMPS system is set to function in a call-and-response manner. When the performer is adjusting the knobs, their changes are driven through the MDRNN to update its internal state but predictions are discarded. When the performer stops for two seconds, the IMPS system takes control of parameter changes, generating predictions for the parameters continually from where the performer left off and updating the eight synthesis parameters in real time. The generative system's changes are displayed on the LED rings on the control interface as well as on the computer screen. The performer has control of the diversity controls (prediction temperature) allowing a

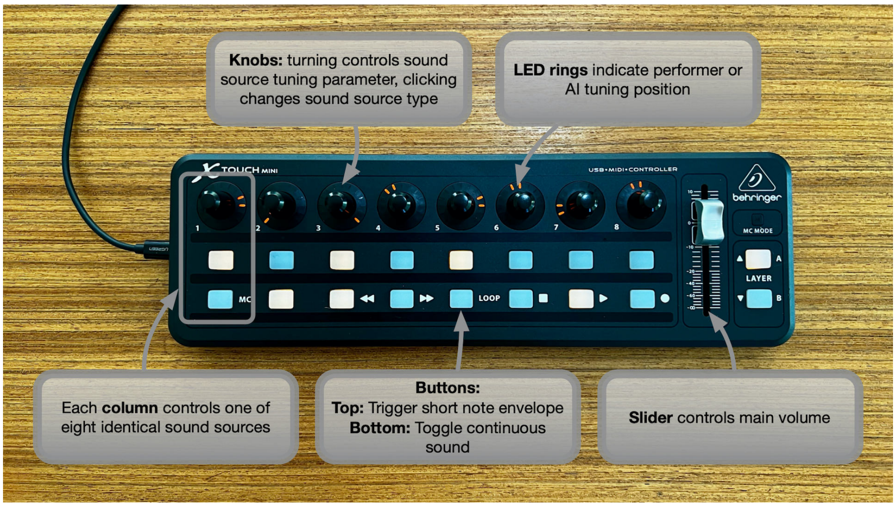

Figure 2: The XTouch Mini MIDI interface used in this performance system. Each column of controls is mapped to a separate sound generator. Both the performer and generative system can adjust the parameter knobs. The performer has access to other controls to steer the performance.

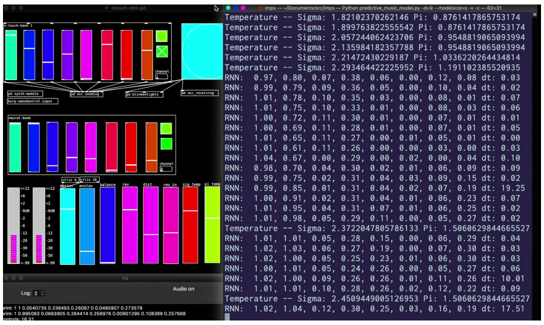

Figure 3: The computer screen view during performance showing the state of each control knob from the performer and generative AI system. The RNN system runs in a terminal window on the right. This screen is shown to the audience during performance.

degree of influence over generated material.

While "call-and-response" might suggest that the performer can do nothing while the generative system is operating, in fact, this setup allows the performer to adjust other aspects of the performance; for instance, the buttons changing the envelope state, the sound generator type as well as the computer-based controls for effects. In this type of performance, it is advantageous to allow a degree of generative change to one part of the musical system to continue while focusing on other parts.

4. Performance Experiences and Conclusions

This system has been deployed in live performances since 2019. These experiences demonstrate that the generative system works and makes a practical contribution to the performances in terms of creating plausible adjustments to the synthesis parameters. A deeper question then is whether the MDRNN generative system offers a level of co-creative engagement above what could be offered, for instance, by a simpler random-walk generator. From the experience of these live performances, it does seem that the generator can be influenced simply through the style of adjustments that the performer is making (e.g., it tends to continue adjusting the knobs that the performer previously was using). Different behaviours in between the eight knobs, e.g., adjusting just one, changing multiple, pausing in-between adjustments or making continual changes, appear in the generator's changes. These behaviours appear "for free" with the MDRNN, that is, they are learned from the dataset, whereas they would need to be encoded into a rule-based generator manually.

Whenever the system is used, either in rehearsal or performance, the performer's interactions are captured to continue building a set of gestural control data for the XTouch Mini controller. As the system is retrained with new data, it "learns" more behaviours, just as the performer adjusts their style in between performances. In this way, this system could be said to be co-adaptive [12], although this is yet to be studied in a rigorous way. From the experience of working with this system, it can be reported that features such as the buttons controlling synthesiser envelopes were added in order to give the performer control over the sound while allowing the generative system to operate. Even though there is the potential for direct interplay between the performer and generative system, it seems to be important to have some different roles to play, and to allow the performer to listen and interact without interrupting the generative model.

From a practical perspective, this system has been successful in allowing complete performances in co-creation with a generative AI music system. The generative system acts as a predictive model for control gestures and is clever enough to enable interaction and steering from the performer using only their own performance gestures. Higher level behaviours, such as long-term structure of the performance are not learned by the model but need to be controlled manually by the performer. While this could be said to be limiting, when compared to similar non-generative system, the performer in this case can switch to handling high-level changes while control over the synthesis parameters is seamlessly continued by the generative system.

This research has described a generative electronic music controller for co-creative performance. This system fits within the idiom of improvised electronic music performance and shows how a machine learning model for control gesture prediction can be applied in a typical electronic music controller allowing a very different style of music generation to symbolic music generation systems. Many other electronic music designs would be possible within this style of interaction, and we see this work as part of developing an orchestra of co-creative musical instruments that interrogate how modern music generation and music interaction can be applied together.

Acknowledgments

The Titan V GPU used in this work was provided by NVIDIA Corporation.

References

[1] C. Ames, Automated composition in retrospect: 1956-1986, Leonardo 20 (1987) 169–185. doi:10.2307/1578334.

[2] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, D. Eck, Music transformer: Generating music with long-term structure, in: Proc. of ICLR '19, 2019. arXiv:1809.04281.

[3] C. J. Carr, Z. Zukowski, Generating albums with samplernn to imitate metal, rock, and punk bands, arXiV Preprint, 2018. arXiv:1811.06633.

[4] Y.-S. Huang, Y.-H. Yang, Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions, in: Proceedings of the 28th ACM International Conference on Multimedia, Association for Computing Machinery, New York, NY, USA, 2020, p. 1180–1188. doi:10.1145/3394171.3413671.

[5] A. Roberts, J. Engel, Y. Mann, J. Gillick, C. Kayacik, S. Nørly, M. Dinculescu, C. Radebaugh, C. Hawthorne, D. Eck, Magenta studio: Augmenting creativity with deep learning in Ableton Live, in: Proceedings of the International Workshop on Musical Metacreation (MUME), 2019. URL: http://musicalmetacreation.org/buddydrive/file/mume_2019_paper_2/.

[6] T. R. Næss, C. P. Martin, A physical intelligent instrument using recurrent neural networks, in: M. Queiroz, A. X. Sedó (Eds.), Proceedings of the International Conference on New Interfaces for Musical Expression, NIME '19, UFRGS, Porto Alegre, Brazil, 2019, pp. 79–82. doi:10.5281/zenodo.3672874.

[7] F. Pachet, The continuator: Musical interaction with style, Journal of New Music Research 32 (2003) 333–341. doi:10.1076/jnmr.32.3.333.16861.

[8] G. E. Lewis, Too many notes: Computers, complexity and culture in "Voyager", Leonardo Music Journal 10 (2000) 33–39. doi:10.1162/096112100570585.

[9] C. P. Martin, K. Glette, T. F. Nygaard, J. Torresen, Understanding musical predictions with an embodied interface for musical machine learning, Frontiers in Artificial Intelligence 3 (2020) 6. doi:10.3389/frai.2020.00006.

[10] S. Jones, Cybernetics in society and art, in: Proceedings of the 19th International Symposium of Electronic Art, ISEA2013, ISEA International; Australian Network for Art & Technology; University of Sydney, 2013-01-01. URL: http://hdl.handle.net/2123/9863.

[11] C. P. Martin, J. Torresen, An interactive musical prediction system with mixture density recurrent neural networks, in: M. Queiroz, A. X. Sedó (Eds.), Proceedings of the International Conference on New Interfaces for Musical Expression, NIME '19, UFRGS, Porto Alegre, Brazil, 2019, pp. 260–265. doi:10.5281/zenodo.3672952.

[12] W. Mackay, Responding to cognitive overload: Coadaptation between users and technology, Intellectica 30 (2000) 177–193.