Creating a Machine Learning Assistant for the Real-Time Performance of Dub Music

Joaquín Jiménez Sauma[1],

[1] Independent Researcher
jjsauma@gmail.com

Abstract. We apply machine learning for the real-time control of a mixing console to mimic the way a dub music producer uses it to perform a new version of a musical piece. A user-trained neural network maps the meaning of semantic controls available to the user to apply rules of dynamics, mixing and audio processes in the mixer. Tests with controls such as "colour" of the mix, amount of rhythmical behaviour and proportion of dynamics showed an accurate control of the performance. Performance tests show the contribution of this project when compared with an instrument, adding variation to deterministic performances and being consistent with the aesthetics created by dub music producers.

Keywords: Music performance, Human-Computer Interaction, Generative Music, Max/MSP, Machine Learning, Neural Networks, Ableton Live, Computational Creativity, Intelligent Music Production, Dub Music.

1 Introduction

We propose this project as the continuation of the research and performance piece: "Polymer Dub" by Jimenez (2019, 2020), where the author analysed the theoretical foundation of dub techno and proposed the inclusion of machine intelligence as the next step in its evolution. We draw inspiration and influences from the works of Kolioulis (2015), who constructs the archaeology of dub techno, the work by Lee (2019), who leverages the aesthetic properties of machine learning inconsistencies and from the Nooscope Manifest by Pasquinelli (2020), where the author theoretically analyses the bias and mystifications of AI.

Dub mixing is an artistic task that is usually performed by an expert audio engineer or music producer (Cox, 2005). A typical performance of dub music consists of an audio engineer manipulating a mixing console to improvise while processing the stems of audio of a musical composition, either recorded or generated during the performance while adding or subtracting each strand of sound to generate a new composition with more impact. In this activity, the engineer uses the mixing console as an instrument for encouraging musical intervention and creation instead of cautious fine-tuning (Bengler, 2011). Dub mixing's most recurring techniques include isolating the rhythm tracks (typically drum and bass) and emphasising it beyond the rest of the sounds in the recording.

2 Design strategies and implementation

This work focuses on the process of creating Polymer Dub's main control device to assist the performer while avoiding the creation of an autonomous system. An essential part of this development is to define boundaries where the performer creates, and the assistant performs to complement the performer's actions without interfering with the artist's intentions. We apply a trained machine learning model to a control device in Max/MSP running inside Ableton Live to respond to performer instructions via a small number of semantic controls, and in turn, dynamically setting Live's mixing console parameters to process the audio in real-time. In designing this device and performance, we have paid particular attention to performance and expression, hence the importance of mapping a connection between the user input and system parameters. This strategy helps the performer to avoid the complex control mechanisms of an instrument. This concept is known as the mapping layer, defined as linking or creating a correspondence between control parameters and sound generation or synthesis parameters (Levitin, 2002) and in this project, the mixing controls of a mixing console. The mapping layer interface is presented to the user as a set of semantic controls, having a meaning that is easily understood by humans (i.e. describing high-level acoustic concepts such as "warm" or "dark"). Our definition of the mapping layer is at the point where the machine learning model predicts and transforms interaction from the user to settings and dynamic behaviour in the digital mixing console, implementing a one-to-many mapping in a way that a single knob on the user interface can be translated via the machine learning model to up to 24 control values representing the semantic value in the performance. We implemented this device using the ML.* machine learning library for Max/MSP created by Smith (2012), using the Multi-Layer Perceptron Neural Network model (MLP).

Our training dataset structure contains user-defined feature vectors that associate mixer settings and behaviour with the position of the semantic controls. Each vector is composed of 3 values representing the position of the semantic controls (input values) that we want to pair with 24 values describing the mixer behaviour (output values). The input values can be assigned different meanings depending on the performer, and the device's graphical user interface allows the user to input these values and train the model while performing (Figure 1). In empirical and performance tests, we have experienced consistent and satisfactory results with datasets having 100 or more rows of training data. However, we have yet to perform a proper evaluation of our model to determine the optimal dataset size and balance, to compute estimation error as well as leverage artistic expectations in terms of training time and accuracy.

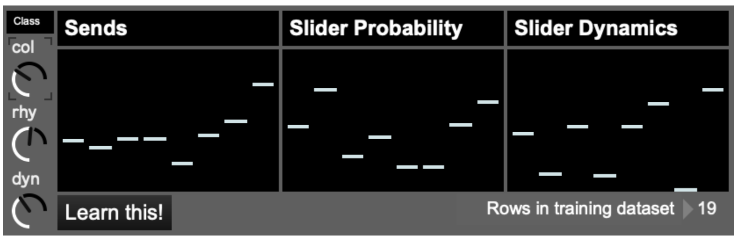

Fig 1. Module to enter rows into the training dataset. This interface allows the performer to enter new rows into the training dataset by setting the controls graphically before or during a performance.

3 Performance

During the performance, the performer continuously defines the values of the semantic controls, in Figure 2 example: *col* = colour of the mix, *rhy* = rhythmical content, and *dyn* = dynamics of the mix. The machine learning model predicts the position of the send controls as well as the dynamic behaviour of the volume faders. The performance module displays the current value of these controls and transfers these settings and behaviour to the actual Live mixer using the Live Object Model API. Each fader control performs independently according to a combination of the predicted settings. Changes in the state of each volume fader happen in time divisions, in synchrony with Ableton Live's metronome (i.e. quarter notes, eighth notes.). Each volume fader behaves in a non-repetitive fashion due to stochastic programming. The device contains special parameters such as the *heart control* to allow the model to set the amount of control of the drum and bass channels, following the dub producers ideology about these instruments representing the heart and brain of the music (DJ, 2008). Also, a *dry/wet* control to allow the performer to decide the amount of intelligent control, or when to take full control of the mixer.

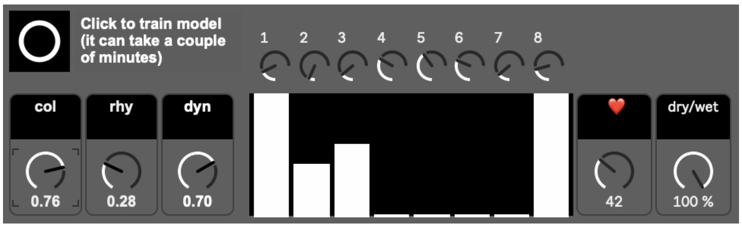

Fig 2. Performance module. This interface shows the parameters available to the user as well as the current state of send and fader controls.

4 Discussion and conclusions

We explored the idea of using machine learning to perform dub mixing by controlling a digital mixing console via an intelligent, user-trained assistant. This design can control several parameters simultaneously with a single gesture, this is more natural and is consistent with the way real musical instruments behave as it involves the perception of correlated dimensions (Levitin, 2002). The intelligent device presents a novel approach for the automatic control of a mixing console in real-time for dub mixing. To the best of our knowledge, the concept of implementing a dub music performance to create a new version of musical composition using a machine learning model is an unexplored problem and warrants a continued investigation in terms of scientific evaluation and theoretical validation. Further work in regards to expression, would be to adopt behaviour in automatic response to changes in short-term audio features to involve perception and cognition issues. As proposed by Jimenez (2020), a machine learning device seems to represent the natural progression in the use of technology for dub music production, and we are interested in exploring the impact of this implementation in the established circle of producers who look for innovative tools to develop their craft.

References

Bengler, B. (2011). The audio mixer as creative tool in musical composition and performance. Institut für Elektronische Musik und Akustik an Der Universiät für Musik und Darstellende Kunst Graz.

Cox, C., & Warner, D. (2004). Audio culture: Readings in modern music. New York: Continuum.

DJ, S. T. S. K. (2008). Sound Unbound: Sampling digital music and culture. Cambridge, Mass: MIT Press.

Jimenez, J. (2019). Polymer Dub Live Set Barcelona [Video file]. Retrieved from https://youtu.be/u8kY-aYkvHE?t=344

Jimenez, J. (2020). Polymer Dub: Urban soundscapes, evolution and cultural values. Proceedings of the Eighth Conference on Computation, Communication, Aesthetics & X. Graz, Austria. https://2020.xcoax.org/jjs/

Kolioulis, A. (2015). Borderlands: Dub Techno's Hauntological Politics of Acoustic Ecology. Dancecult: Journal of Electronic Dance Music Culture 7(2): 64–85 ISSN 1947-5403 2015. http://dx.doi.org/10.12801/1947-5403.2015.07.02.0

Lee, R. (2019). Aesthetics of Uncertainty. In Proceedings of the Conference on Computation, Communication, Aesthetics & X (Vol. 7, pp. 256-262). Universidade de Porto, Portugal.

Levitin, D., McAdams, S., & Adams, R. (2002). Control Parameters for musical instruments: foundation for new mappings of gesture to sound. Organised Sound 7(2): 171–189 2002. Cambridge University Press.

Pasquinelli, M., & Joler, V. (2020). The Nooscope Manifested: Artificial Intelligence as Instrument of Knowledge Extractivism, KIM research group (Karlsruhe University of Arts and Design) and Share Lab (Novi Sad), 1 May 2020 (preprint forthcoming for AI and Society). https://nooscope.ai

Smith, B., & Garnett, G. (2012). Unsupervised Play: Machine Learning Toolkit for Max. New Interfaces for Musical Expression (NIME). Ann Arbor, MI: ICMA, 2012.