INTEGRATION OF MACHINE LEARNING ALGORITHMS IN THE COMPUTER-ACOUSTIC COMPOSITION GOLDSTREAM VARIATIONS

Scott Deal
Indiana University Purdue University Indianapolis
Department of Music and Arts Technology

Javier Sanchez
Indiana University Purdue University Indianapolis
Department of Music and Arts Technology

ABSTRACT

This paper presents an implementation of a musical interface that utilizes machine learning (ML) attributes in real-time performance. The objective behind the work is to empower performers with an expanded musical palette. This is achieved by employing a variation on a delay effect implemented with neural networks. In this scenario, the delayed signal is an echo of previously performed motifs based on new inputs categorized by an ART (Adaptive Resonance Theory) neural network. An overview of the piece used to test the musical range of the effect will be given, followed by a description of the development rationale for the project. The paper will conclude with a qualitative evaluation of the usability and responsiveness of the effect, as well as its contribution to the aesthetic quality of the composition.

1. INTRODUCTION

Traditionally, Machine Learning (ML) methods have been applied successfully to many problems in engineering and science. Currently, developments in both software and hardware have made the use of such methods feasible in other domains, including music performance [4].

One of the particular problems of implementing such systems lies in the difficulty of integrating the mechanics of ML into the aesthetics of music. In this work we developed a machine learning system that helps musicians augment their performance. In turn, the involvement of a human performer then helps to guide the aesthetic relevance of the ML system.

Goldstream Variations (Deal, 2012) was created with this ML project in mind. Both the composition and the system were created in tandem in order to effectively integrate ML into the musical fabric of the piece. This approach gave insight into how closely artistic expectations can be matched by the chosen design methodology.

The ML interface was constructed by incorporating *ml.**, the *Machine Learning Toolkit for Max 5+* developed by Ben Smith [7] into an *Ableton Live for Max* template, which provided a mode for the implementation of newly captured MIDI clips

The interface designed can be classified as a "logical effect", i.e., a system that analyzes the input and then outputs signals with logical meaning in the chosen context. For clarity, interface and logical effect will be used interchangeably in this paper. The term "interface" in this case refers to the whole system acting as a facilitator to the expanded musical space, rather than to a GUI.

In the interface, the inputs are sequences of pitch classes and the outputs are pre-recorded sequences from previous passages of the music. The decision regarding which pre-recorded sequence should be played at any given time is determined by an unsupervised learning system that categorizes incoming sequences and then selects material that most closely resembles the current melodic context. From the point of view of the musician, the interface acts as a logical delay effect.

2. GOLDSTREAM VARIATIONS

Goldstream Variations is a work that creates a dynamic musical environment through the integration of performance, harmonic movement and computer interactivity. For clarity throughout the remainder of the paper, the work will be listed as *GV,* with numbers following representing one of the five variations (e.g. *GV*.2 = *Variation 2*). The variations are scored for one to seven musicians on undetermined acoustic instruments, together with one to seven electronic/computer artists. Any ratio of electronic to acoustic artists is feasible. The selection of this grouping shapes the aural nature of performance space through the arrangement of performers and loudspeakers. Each page of the score constitutes one variation that acoustic and controller musicians perform soloistically, yet in heterophonic fashion with the rest of the group. These passages are captured by computer artists and the ML interface, who in turn release cascades of re-processed harmonic, timbrel, and melissmatic material that emanates throughout the space in a constantly changing sphere of energy. *GV* is designed for performance in either a single physical space, or distributed telematically between multiple sites on high-bandwidth Internet. While the score is open, suggested instruments include percussion, piano, harp, strings, guitars, woodwinds, prepared/augmented instruments, controllers, and computer interactivity. ML is introduced into the design of the work via the structural shape of the composition, in which variations consist of virtuostic passages followed by large rests, which in turn create room for liberal amounts of interactivity. The aesthetic focus of a performance lies in the timing, placement, and juxtaposition of virtual and live sound.

3. LOGICAL DELAY

3.1. Objective

The objective of the project was a processing system that would allow a performer to produce musical renditions that would normally be difficult or impossible to do by him playing alone [2]. There are several such systems used in different genres of music employing electronic means: looping devices, delays, and harmonizers. Examples of logical effects that transform the input into a relevant musical context also include arpeggiators, chord builders and scale transformers [3].

Characteristically, these systems provide an interface for the musician to access different possibilities of the musical space. These systems work well within contexts that feature non-improvisational aspects, as their features allow the performer to predict fairly easily what the output of the system will be. By using ML, however, both musician and computer share artistic contexts that require a higher level of indeterminacy. As a result, it is possible to produce sonic material that is more conducive to the context of spontaneous selection between man and machine [5]. However, one of the issues with this approach is delimiting the boundaries between performances perceived as being produced by a human and a machine, and performances considered a product of human interaction with electronic tools. This distinction can have an impact in the actions of the musician and also in the perception of the piece by the audience [5].

There are numerous examples of ML in particular, and AI in general, being used in the context of interactive music performance [1]. These works utilize different methods to produce systems that range from fully independent, to input-output processors. The system presented here aims to retain a certain level of connection with the player, as opposed to creating a fully independent system.

Accounting for these issues, the program was designed in such a way that it allowed the production of note sequences that have a loose contour relationship with the material that is being played at the moment. These sequences are short enough so not to be perceived as being produced by an entirely independent identity; rather, a spatial relationship is established. Since the material produced by the ML system is a rearrangement of the notes played by the performer, ownership of the final product is not entirely removed from the musician.

3.2. Logical structure of the interface

The system consists of several elements, as shown in Figure 1. The encoder is in charge of preparing the input notes for the neural network, which is the "categorizer" in the figure. The categorizer detects, identifies, and classifies new categories of melodic contours as they are performed, after which the system stores the phrases in memory to be used later.

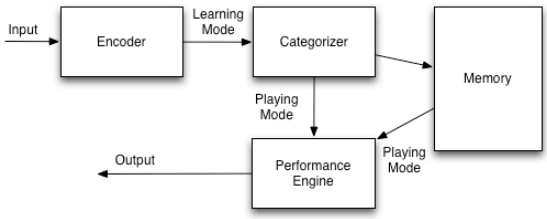

Figure 1. Block diagram for the logic of the interface.

Refer to Figure 2 for the description of the algorithm used for manipulating the material.

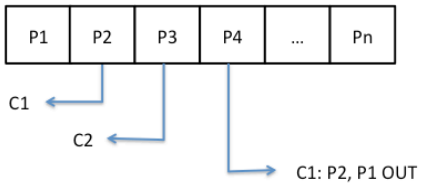

Figure 2. Pitch manipulation algorithm

Assuming a sequence of pitches P1, P2, P3, …, Pn, the system produces an event every time a new note is presented at the input, and a category change is detected. This event could be the storage of the notes held in the circular buffer of the system (which retains the last M notes played), or the playback of one of these previously stored sequences. The main element controlling these manipulations is the category number as presented by the ML system. In Figure 2, an example can be seen of a series of events where two categories are generated by the time P3 is presented to the system. When P4 is presented, the system identifies that string (P4, P3, P2, P1) as belonging to the same category as (P2, P1), which results in the system playing back this last sequence (Category 1).

3.3 Learning and playing mode

The interface is controlled using two macro parameters that determine the mode of operation of the effect. These two parameters are learning mode and playing mode. When learning mode is engaged, the system tries to classify any new inputs, effectively training the neural network. This may be useful as the performer may consider that certain inputs are not relevant for the detection algorithm, and thus it may be irrelevant to train the categorizer with these particular inputs.

For example, *GV.2* relies heavily on rhythmic and dynamic modulation, yet not so much on melodic contour; therefore it makes sense to stop the input flow to the categorizer so it won't generate too many irrelevant categories for the piece. Playing mode also works by feeding the current stream of notes to the categorizer, the difference here being that all of the categories detected by the neural network will trigger an action (not only the ones created for the first time). In

this case, every time a change of category is detected, the system retrieves the sequence that is being used as a reference for that category. The system is designed to have both modes engaged at the same time, playing back sequences from categories it recognizes, and storing new sequences from recently learned categories.

3.4 Implementation details

For the creation of the system, Max for Live was utilized. The access to the Live API was helpful as it provided control over clip slots that worked as the memory element for the interface. This in turn allowed for the easy transformation of the stored reference sequence for each one of the categories. It is simple for example, to transpose and alter the dynamics of the recorded sequences using the built-in MIDI facilities within Live, without the need of creating more Max objects.

For the learning network, an implementation of a Fuzzy Adaptive Resonance Theory network developed by Benjamin Smith was used [7]. The Max object is part of the Machine Learning Toolkit, which also includes a spatial encoder that was used also for the encoding section of the interface.

The *ml.art* object has several configurable parameters that can be used to adjust the rate of learning, and the number of categories created. The outputs of the object show the currently identified category and the resonance measure, i.e. how well the current input matches the learned categories. Full details for the object can be found by consulting [7, 8].

4. TUNING OF THE SYSTEM

When using the interface, there are two sets of adjustments that can influence the behavior of the system as well as the aesthetics of the entire performance.

The first set deals with parameters of the learning network itself. In the context of the *GV*, the network was calibrated to produce around 30 categories (with their respective reference sequences) per performance. Information on how to do this can be found in [7]. The number was determined through an analysis of sequence length and overall musical structure of each variation. When playing mode is engaged, the number and frequency of released sequences are a matter of careful consideration. Too many will overwhelm the balance of virtual and acoustic; too few will seem empty. Several characteristics of the learning network are relevant for the decision making process, including the fact that the order in which the input vectors (sequences) are entered into the system affects the creation of the categories [9]. This implies that different performances will yield different reactions from the system, which, in this case, is a desired attribute.

The second set of adjustments has to do with planning when to engage the different modes of the interface. These adjustments are best when made in advance as well as in discussion with performers using the interface and ensemble performers. The design of *GV* is created for the inclusion of electronics; in this case, the performer using the interface must account for the space taken by the effect and the interrelationship with their own playing and that of the other performers.

5. DEVELOPING ML OPERATIONS FOR *GV*

When developing ML for *GV,* it was decided to engage the effect only in certain sections of the piece. A different combination of the learning and playing modes were used for each one of the variations, as follows.

GV.1 begins the computer learning process, as a few phrases are played by each one of the performers, creating several opportunities to capture the phrases then fully utilize the sequence potential of the pitch set and the subsets. For this variation, learning mode is engaged and the network generates 6 to 8 categories.

In *GV.2*, the learning mode is disengaged since the content is primarily dynamic and rhythmic. It is not desirable to generate too many categories that will lack relevance later in the piece, since in the sequences obtained from this section do not fit aesthetically in the rest of the variations. However, learning mode can be reengaged almost at the end of the section to generate 2 or 3 categories that can be useful later on.

In *GV.3,* the score calls for a solo instrument that plays the melody while the other instruments play a sparse accompaniment. Here the playing mode is finally engaged and can be combined with the learning mode to create even more categories. The effect fits aesthetically regardless of the role the performer using the interface has in the ensemble; as a soloist, the effect forms part of the accompaniment and, as a background player, the system will play equally sparse sequences.

GV.4 presents new contour material that makes the ideal configuration for the effect to be playing/learning. The system adds to the texture of the variation, but care must be taken when selecting an appropriate synthesis engine for the playback, since timbre qualities too similar to the other instruments in the ensemble can lessen the clarity of the performance. Modifications to the source's spectrum can help overcome these problems, and can help small ensemble configurations fill more spaces of the global spectrum.

GV.5 contains some phrases reminiscent of *GV.1*. In this variation, learning mode can be disengaged while the system keeps playing with the material obtained from the interpretation of the previous sections. Alternatively, the system can be turned off completely to give a greater sense of embodiment between the effect and the performer by making a clear connection of the control exerted by the musicians over the system.

6. EVALUATION

Goldstream Variations received its premiere in a telematic performance between the Budapest Palace of the Arts (MUPA, Hungary) and the Tavel Arts Technology Lab at IUPUI (Indianapolis, USA) in November 2012. Musicians in Budapest included a percussionist and a harpist; who were joined in Indiana by a musician feeding the ML toolkit by performing on a *MalletKat* percussion controller and one laptop performer facilitating the ML system and audio processing. The performance was received well by the

audience, and contained successful and less successful outcomes. Most effective was the juxtaposition on the four musical voices in the Budapest space. The musicians successfully navigated themselves through the variations while cultivating an effective spatial and time contour within the ensemble, which is also to the credit of the sound engineer running the mix in Budapest. The least effective was outcome was the lack of timbrel variation/interest as a result of artistic real-time audio processing. However, this attribute was not directly relayed to the ML system, but can be attributed to the fact that the only person operating a computer during the performance had his hands full making sure the ML system worked as planned. The conclusion reached after the performance was to provide for both greater spectrums of musical parameters that can be manipulated by the ML system, plus the inclusion in the score instructions for more laptop performers to supplement real-time audio parameters.

While the ML system performed as planned, it can be improved by reconsidering the system input and output. Specifically, having the possibility of processing both audio and MIDI inputs would make the system more flexible. This could be achieved relatively easily, as the nature of the effect alleviates the latency problems usually associated with pitch tracking algorithms. An additional advantage of including audio inputs for the ML system would mean that all of the musicians could perform on acoustic instruments. Under the current version, a musician performing on a MIDI controller of some sort must be part of the ensemble in order to feed sequences into the ML system, or pre-made sequences need to be employed. Likewise, allowing the output to be modified by other logical modules would help the system to be relevant in more diverse musical styles. An example of this would be a randomizer for note durations and pitch values that could suggest variations while still retaining the overall logic provided by the categorizer.

7. FUTURE WORK

Future *Goldstream Variations* efforts with ML will include expanding on the ability of the system to create sonic variety by encoding other musical features, such as rhythm, timbre, or even more abstract elements, such as intervallic spread. Given the openness of the score as it relates to instrument/voice selection, many different versions of the work could create a large range of sonic material suitable for manipulation in a variety of performance scenarios.

When implementing the variations, aesthetic considerations stem from the arrangement of performed material by artists and the ML system. Audiences classify different musical features unequally, so the more abstract implementations of the effect can present difficulties when conveying the intended meaning. This will remain a problematic factor unless a clear method of articulating the ML voice, or sound is established in a performance context. More performance research and rehearsal is needed to establish best practices for this kind of aesthetic communication.

Additionally, over time, the methodology employed in this work implies that the assembly/rehearsal process will also draw inspiration from the mechanics of the ML system. The performance instructions could then be adapted/expanded to explore additional musical ideas allowed by the capabilities of the effect.

8. CONCLUSION

ML algorithms provide a way to add variation to static logical effects. Challenges encountered while following these approaches include the establishment of logical connections between the output of the system and the musical context, and the controllability of the system. Further efforts in broadening the pallet of options for the ML system, combined with more rehearsal/exploratory hours with live musicians performing the variations will yield a broader range of musical possibilities. Solutions to these issues shape the aesthetic outcome for the whole performance, as one issue affects the perceived integration of the effect by the audience, while the other affects the interaction between system and performer.

9. REFERENCES

[1] Baird, B., Blevins, D., Zahler, N., "Artificial Intelligence and Music: Implementing an Interactive Music Performer". *Computer Music Journal*, 17:2, 73-79, 1993.

[2] Collins, N. "Generative Music and Laptop Performance". *Contemporary Music Review, 22:4, 6-79*, London, UK, 2003.

[3] Ramirez, R., Peralta, J., "A Constraint Based Melody Harmonizer". *Proceedings of the Workshop on Constraints for Artistic Applications" (ECAI '98)*, Brighton, UK, 1998.

[4] Roads, C. "Research in Music and Artificial Intelligence". *ACM Computer Surveys, Volume 17, Issue 2*, New York, USA, 1985.

[5] Roederer, J. *The Physics and Psychophysics of Music. An Introduction*. Springer, New York, 2008.

[6] Rowe, R. "The Aesthetics of Interactive Music Systems". *Contemporary Music Review, 18:3, 83-87, 1999*

[7] Smith, B., "Machine Learning Toolkit for Max 5+", http://ben.musicsmiths.us/ml.phtml

[8] Smith, B., Garnett, E., "Machine Listening: Acoustic Interface with ART", *Proceedings of SIGCHI Intelligent User Interfaces*, Lisbon, Portugal, 2012.

[9] Smith, B., Garnett, E., "Unsupervised Play: Machine Learning Toolkit for Max", *Proceedings of the 2012 International Conference on New Interfaces for Musical Expression*, Ann Arbor, USA, 2012.