The Music of Machine Misreading: Machine Learning Artifacts as Sources for Artistic Content and Control

Jeffrey M. Morris, DMA

Department of Performance Studies, Texas A&M University, College Station, Texas, USA

morrismusic.org

e-mail: morris@tamu.edu

Abstract

In contrast to research that seeks to use machines to faithfully replicate human performance in familiar tasks, this research project explores the native character of machine learning systems by using their errors and artifacts to create and shape content in artworks, in order to facilitate understanding the nature and creative potential of these systems more deeply and intuitively, and to develop a sensitivity to the ways they unintentionally they shape our thinking. Two musical compositions demonstrate useful techniques including a "crib sheet" structure, self-cancelling signal processing, and disregarding artificial metaphors imposed on machine learning systems.

1. Introduction

It is natural to desire for new technologies to perform tasks we already know. This happens in music too, where the ubiquitous paradigm divides software into a signal-rate "orchestra" and a control-rate "score" (i.e., sheet music) for the virtual instruments to perform, forcing new technologies to replicate historic composition and performance practices and overlooking more digital-native approaches to creative work (Morris, 2019c).

Since the field of machine learning is evolving rapidly and receiving much attention from industry, it is natural to dismiss undesired results (toward practical goals) as worthless and to move on. However, human culture increasingly needs a deeper and more intuitive understanding of how these systems work, so we can better understand their hidden potential and how they are unintentionally shaping our thought processes. Scholars including Janelle Shane are giving increasing attention to the out-takes or “waste” from machine learning systems (Shane, 2018, 2019). The line of research described here directly exploits machine error as a source of content and control in artworks, in order to give voice to the native character of machine learning systems.

2. Related Work

In music, John Cage's work has been the leading example of compositions that interrogate their own situations and break basic assumptions, to reveal new aesthetic experiences, new modes of creation and expression, and new realizations to be had while experiencing art. For example, *Inlets* (Cage, 1977) features material that is outside the composer's or performer's control by having performers tilt conch shells filled with water that will “gulp” unpredictably. In technology-based music, Milan Knížák's *Broken Music* (1979) involves pre-recorded vinyl LP records that have been cut and reassembled, not to feature the originally recorded music but rather to give voice to the sounds of the phonograph medium itself as it fails. More recently, Art of Failure's *8 Silences* (2013) lets sound emerge by sending a silent audio signal through streaming networks, so that the only sounds heard are the errors introduced by the network itself. Relating to machine learning, Ryo Ikeshiro's interactive audiovisual installation titled *Ethnic Diversity in Sites of Cultural Activity* (Ikeshiro, 2015) uses a camera-fed classification system to herald the ethnicity of viewers by playing music associated with that culture, however incorrect or unconfident the computerized classification may be. As in

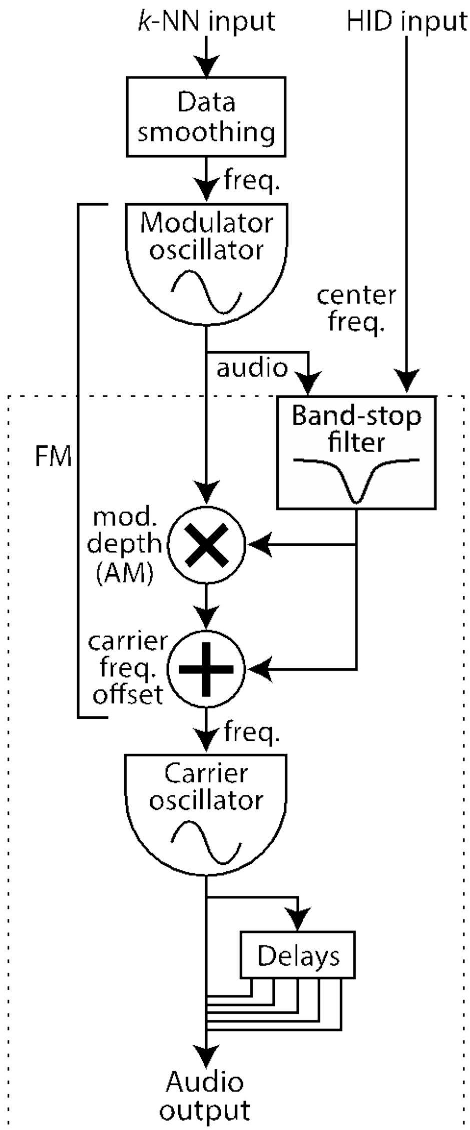

Figure 1. *Divining Rod* signal flow “crib sheet” structure that self-cancels when both inputs match and exaggerates any errors. The dashed box shows the region in which every stage would yield silence when both inputs match.

Knížák's work, the pre-recorded material becomes a medium for highlighting the software's disjunct style of compositing those materials, driven by its tenuous grasp of the world around it.

3. Divining Rod

In the live audiovisual performance work *Divining Rod* (Morris, 2019a, 2019b), a *crib sheet* structure—in which the computer's guess and the actual answer are both known—serves to reliably expose erroneous and wavering classification decisions by the machine learning system. A camera is focused on a computer's (typing) keyboard, feeding a *k*-nearest neighbor (*k*-NN) classifier running in the Wekinator (Fiebrink, n.d.), which reports its results to Max (Cycling74, n.d.). The classifier's guess regarding what key is currently being pressed is mapped to the frequency of a sine wave. On the other side of the crib sheet structure, the keyboard's actual state is polled via the HID protocol and mapped to a band-stop filter that would cancel out the sine wave when both sides of the system agree. That output is used in both amplitude modulation and frequency modulation on the sine wave, so that the result will still output silence when the classification algorithm guesses correctly but any error will create sound that grows in amplitude (loudness) and is spread across the frequency spectrum. This is further sent into a network of shifting delay lines that build a rich contrapuntal background out of any errors that are sonified (see Figure 1). The camera signal is also fed into a feedback-based graphic processing system that is shaped by the sound, creating an audiovisual portrait of the machine's errors and wavering confidence.

In order to give the software the best chance of guessing correctly, so that the resulting errors are the ones most native to this situation, a percussion mallet was modified with a matte blue head and a matte black stick, it is held with a matte black glove, and the keyboard is uniformly lit with an LED lamp next to the camera over the keyboard (see Figure 2). The video signal is processed to remove the image of the empty keyboard and glare (see Figure 3). All of these measures allow the software to most confidently read the blue color channel and ignore superfluous objects and glare in the camera's view. This work was premiered at the 2019 Generative Art international conference in Rome (Morris, 2019a, 2019b).

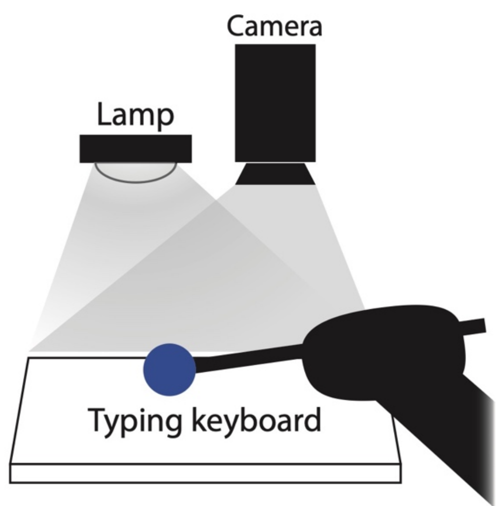

Figure 2. *Divining Rod* hardware configuration for performance.

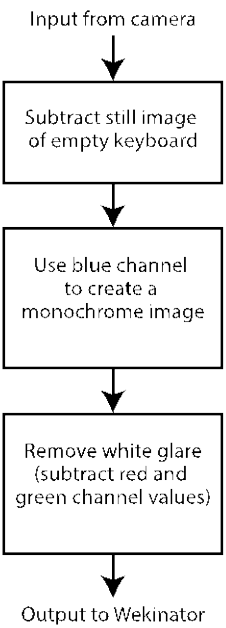

Figure 3. *Divining Rod* video pre-processing.

4. Adapting the Snake Game for Sound Spatialization

The next phase of work is an audio-only composition for a massively multichannel speaker system, such as the 93.8-channel system in György Ligeti Hall at the University of Music and Performing Arts Graz, Austria. In this work titled *The Chances*, a deep Q-learning network (DQN) trained to play a simple snake game (as popularized on Nokia cellular telephones in 1997) is used to spatialize the sound material through the speaker system. A model adapted from de Ponteves (2019) was trained and modified to output its gameplay decisions as text, which in turn is read by sound playback and spatialization software created in Max.

Because the speaker is system is not a 2D plane (as in Figure 4) but a dome plus a “sky” layer—a partial, circular 3D space with varying depth (as in Figure 5)—the four possible movement choices made in the snake game, i.e., whether to move up, down, left, or right, are mapped to have the sound move to the nearest or the second-, third-, or fourth-nearest speaker, respectively, as time passes. This breaks the physical metaphor of the snake game, e.g., the walls are not mapped to static positions in the performance venue, but the character of the trained DQN model is preserved and merged with the peculiarities of the venue as it wiggles through the terrain of the speaker system. Aesthetically, the character of the trained DQN model's steering patterns is the important element to preserve. The composition is not about a snake or a walled garden but rather the decision-making process of the DQN model. So, breaking those aspects of the physical metaphor focuses better on the characteristic nature of the movement.

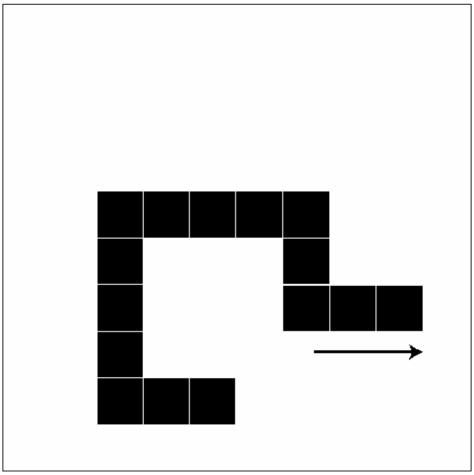
Figure 4. A 2D snake game.

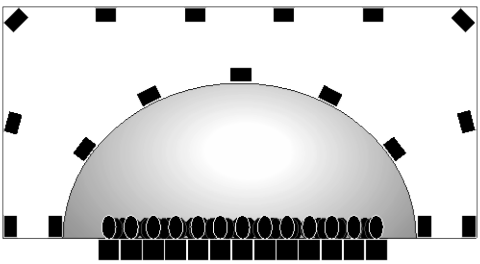
Figure 5. Cross-section of the speaker arrangement in a 3D surround sound speaker system (audience is seated at the bottom).

5. Conclusion and Future Work

Whereas *8 Silences* (Art of Failure, 2013) naturally allows rich sounds to emerge from silence, a self-cancelling crib sheet structure proved useful wherever a computerized guess and definite correct answer can be compared, and wherever audiovisual processes can be put in place that would cancel each other out when the system is correct while giving a rich voice to any errors that occur. With *The Chances*, the notion of breaking the physical metaphor of the original premise has emerged as helpful for keeping focus on the voice of the system itself rather than the fiction that has been fabricated around it.

A future phase of this line of inquiry will investigate the creative potential of applying this approach to a highly modifiable video game such as the classic first-person 3D game *Doom* (id Software, 1993). Whereas the stock imagery and sounds of the game establish a metaphor of a soldier exchanging weapon fire with a variety of adversaries while navigating a walled maze, this project will view the system more neutrally and abstractly as what it is, more basically: an interactive platform to composite and arrange foreground *sprites* and background images and to trigger sounds, on a 2D screen, in an idiomatic way. Disregarding the physical metaphor of the game will not only yield novel audiovisual compositions that expose something more true to the nature of the platform's structure; training a deep learning network based on such imagery (i.e., a 2D abstract image compositing platform with arbitrary rewards and penalties, versus a narrative, representational 3D maze metaphor with mission objectives and injuries) is also anticipated to give discussion-worthy results.

References

Art of Failure. 2013. *8 Silences*. Musical Composition. https://freemusicarchive.org/music/Art_of_Failure/Failure/RAD0004

John Cage. 1977. *Inlets*. Musical Composition. Edition Peters EP 66787, Leipzig, Germany.

Cycling74. Max. Software. https://cycling74.com/

Hadelin de Ponteves. 2019. AI for Games. In *AI Crash Course*. Pakt, Birmingham, UK, 287–326.

Rebecca Fiebrink. The Wekinator. Software. http://www.wekinator.org/

id Software. 1993. *Doom*. Video Game.

Ryo Ikeshiro. 2015. *Ethnic Diversity in Sites of Cultural Activity*. Installation. https://ryoikeshiro.wordpress.com/2015/01/08/ethnic-diversity-in-sites-of-cultural-activity-2/

Milan Knížák. 1979. *Broken Music*. Music Album. Multhipla Records n. 5, Milan, Italy.

Jeffrey M. Morris. 2019a. *Divining Rod*. In *Proceedings of the XXII Generative Art Conference (GA2019)*. Domus Argenia, Milan, Italy, 513–514. https://www.generativeart.com/GA2019_web/JeffMorrisLP.pdf

Jeffrey M. Morris. 2019b. *Divining Rod*. Video. https://www.youtube.com/watch?v=Boj30ARHm3o

Jeffrey M. Morris. 2019c. Gaming Concepts for Analysing Techniques and Aesthetics in Bytebeat Performance. In *Proceedings of the XXII Generative Art Conference (GA2019)*. Domus Argenia, Milan, Italy, 177–189. https://generativeart.com/GA2019_web/28_JeffMorris-168x240.pdf

Janelle Shane. 2018. Machine Learning Failures—For Art! Keynote Speech. Strange Loop 2018, September 27–28, 2018, St. Louis, MO, USA. https://www.youtube.com/watch?v=yneJIxOdMX4

Janelle Shane. 2019. *You Look Like a Thing and I Love You: How Artificial Intelligence Works and Why It's Making the World a Weirder Place*. Litte, Brown, and Company, New York, NY, USA.