HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 3 (2022)

Are Musicians Being Replaced by Artificial Intelligence?

Matthew Chen[1], Ganesh Mani[#] and Alfred Renaud[#]

[1] The Harker School, USA
[#]Advisor

ABSTRACT

The use of machine learning in music interpretation is a tricky subject due to human subjectivity and personal preferences. Magenta's PerformanceRNN provides a benchmark with which musicianship and human performance can be replicated and thus evaluated.

Introduction

Since the 1950's, artificial intelligence has had an invisible grasp on the music industry as a whole. From lyric and melody generation to song recommendations, much of the music we experience today has been mechanized to meet consumer standards and appease the general audience. However, despite the fact that machines are able to observe and predict human preferences to a tee, the days of a stereotypical media-portrayed, technologically-advanced society where robots take musicians' jobs still seems to be a far-fetched future. Or are they?

Artificial Intelligence in The Music Industry

The introduction of artificial intelligence in creative industries is a sensitive topic. Because of its creative nature, art has been able to withstand the automation occurring in other lines of work. Compared to other trades, machine learning is used more as an aid rather than a replacement. While other careers are becoming more and more computerized, a musician might use AI to help make a jazz beat for his song. Recently, this conversation regarding artists and artificial intelligence increased due to the sudden popularity of OpenAI's DALL-E machine learning models which are able to generate images based on descriptions provided by their users. With recent advancements in art generation, style transfer, and composition, many fear that machine learning will eventually be able to compete and even overtake professional artists. But is this fear justified? Others believe that the human elements of art are too complex to be imitated. Music in particular is an unmistakably "human" product, yet it is universally understood and appreciated.

Definitions

Two words that are commonly referred to in musical discussions are dynamics and tempo. The terms essentially just tell a musician how loud they should play and how fast they should play.

Dynamics

The word dynamics refers to the variations of volume that a musician plays. Dynamics are often used to elicit emotion or emphasize/reduce certain notes.

Tempo

Tempo refers to the rate of speed of a musician's performance and dictates the intensity of certain musical passages.

Challenges and Roadblocks

Classical Music

When conducting my research, I decided to focus on classical music: its longevity and ambiguity make it the most antiquated genre of music, and the vast amounts of information online mean I won't need to worry about lacking data. One characteristic of instrumental music is the level of creative reign given to performers: dynamics and tempo are relatively left up to the musician's interpretation. In fact, many of Bach's compositions lack dynamic markings because he wanted performers to interpret the music in their own way. However, musical freedom does not mean performers have free reign. Musicians still have to rely on their musical ear and taste to determine the intensity and speed of their playing. One of the more complex problems during research was figuring out the difference between a good performance and a bad performance; there is no ruleset for an aesthetically pleasing interpretation of a piece (i.e. there is no correct speed, volume, or style to play classical music); however, musicians can still inherently tell if it is musically tasteful or not.

Midi Files

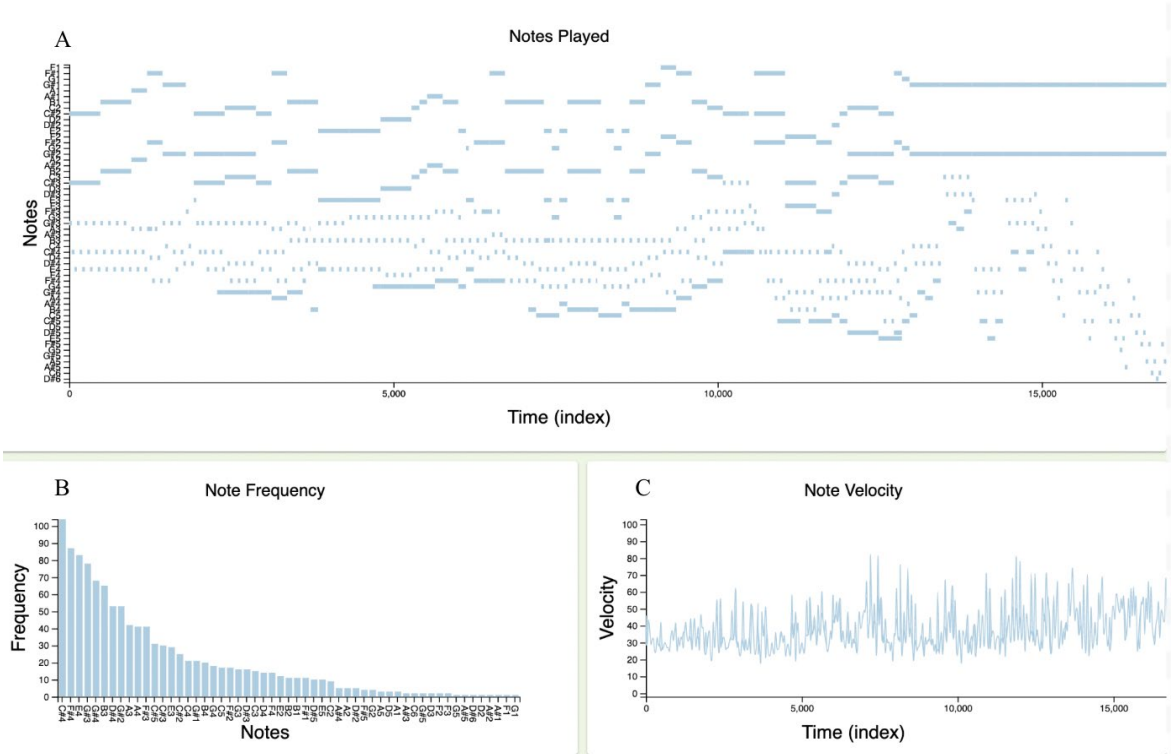

Figure 1. Graphical representation of Beethoven's Moonlight Sonata as a MIDI file.

Interestingly, taking a look at my file type helped me narrow down my objective from adding human interpretation to just dynamics and tempo. A MIDI file or Musical Instrument Digital Interface holds instructions rather than sounds, leading to precise but robotic sounding music. MIDI's can hold song, sequence, and track structures, tempo, and time signature information, essentially acting as a compact digital musical score. While these files are compact and take up little space, they don't convey emotions the same way a file such as an mp3 does. As seen in figure 1A, the two axis are labeled notes and time, meaning dynamics are neglected in favor of storage. The problem with midi files is that they offer no interpretation; they're black and white instructions on how to play the song. The musical aspect of interpretation and performance omitted from a MIDI file is the human inconsistency that comes with playing music. This means that being MIDI files suffer from the cardinal musical sin of being "too perfect." When a musician plays a piece perfectly, they run the risk of sounding too robotic and losing their performance's artistry. In truth, beautiful music is created through inconsistencies. Whether it's playing a note softer to achieve a sadder tone or holding a note longer than it's actual value to emphasize it, interpreting a piece is just as important as getting the notes right. This becomes tricky when we consider it from a machine learning perspective: when everyone has a different interpretation of a piece, how can the computer notice a pattern?

Magenta

Magenta is Google's open source research project that explores artificial intelligence in music and art. Ironically, their github repository states that it is "also an exploration in building smart tools and interfaces that allow artists and musicians to extend (not replace!) their processes using these models." Though the functions of these Magenta models range from composition to language modeling to style transfer, the most relevant model to this paper is PerformanceRNN.

PerformanceRNN

According to the repository, PerformanceRNN "applies language modeling to polyphonic music using a combination of note on/off, timeshift, and velocity change events." In other words, PerformanceRNN is able to compose piano music while adding expressive dynamics and tempo, sort of like stylized improvisation. However, PerformanceRNN differs from standard machine learning composition models in a few important ways. First, the model is trained on around 1400 performances by pianists from the Yamaha e-Piano Competition dataset, lending it an innate ability to interpret music just like a human would. To put it another way, human input equals human output. Importantly, repertoire from the Yamaha Competition was also all classical music. Finally, the model puts its entire focus on the piano and places its note on/off switches based on human performance.

Temperature

One way to modify the output of PerformanceRNN is using the temperature variable. At a base of 1.0, the model's predicted event distribution is outputted without any modification. By increasing temperature, the randomness of the model increases and thus so does the sporadicity of the music. By decreasing the temperature, the music becomes monotonous. The introduction of this variable allows the program to control and mimic human irregularity. Personally, I found that a temperature of 1.2 to sounded the best.

Conclusion

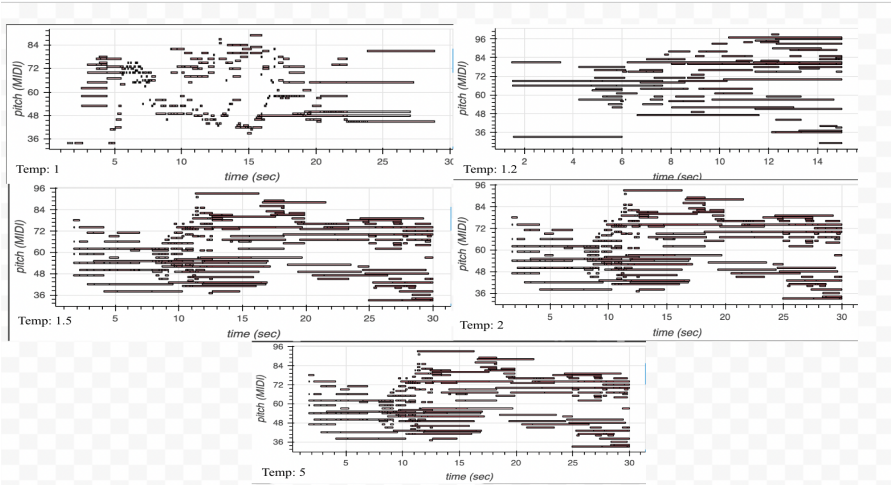

Figure 2. Graphical representation of PerformanceRNN's output at temperatures 1, 1.2, 1.5, 2, and 5.

From a classical music fan's perspective, PerformanceRNN does an amazing job of mimicking humans. It captures dynamics really well, using them in the correct places and at the correct intensities. While not on the level of Beethoven's Moonlight Sonata in Figure 1, PerformanceRNN's output sounds coherent and musically sound. As you can see in Figure 2, at higher temperatures, the notes become more random and sporadic while at lower temperatures, musical motifs are repeated as the notes are less random and more repetitive. Finally, if you want to come to your own conclusion, the first citation in my references leads to a website with audio clips demonstrating PerformanceRNN's output at various temperatures.

Discussion

So can machines interpret music as well as humans?

Overall, yes, the output of Magenta's PerformanceRNN is able to emphatically mimic the incongruencies of human interpretation: the outputted rendition is only made possible by thousands of data entries of real people. Ironically, while the models dependence on a human dataset for musical interpretation is what makes its performance so compelling, this reliance on humans is its own weakness. Though the capabilities displayed by PerformanceRNN are extremely impressive and rival those of real musicians, this likely won't be a problem until we no longer have to rely on a material repository. For the time being, musicians are safe; however, artificial intelligence is slowly creeping up on creative industries, and it's only a matter of time before artists must come face to face with increasing automation in the industry.

Limitations

Although this project focused on classical music, exploring other genres is just as important and is something I want to do in the future. One of the main problems benefits of working with a piano dataset is how easy it is to read and use MIDI files, due to the notation. Other genres of music might not offer this same benefit due to MIDI files' limitations. Finally, classical music is an extreme example of human interpretation: the ease of use associated with dynamics and tempo might not be present in other musical categorizations, meaning it may be harder for a computer to recognize patterns and thus its output won't be as clean.

Acknowledgments

Thank you to my advisors Ganesh Mani and Alfred Renaud for guiding me through the project!

References

Performance RNN: Generating music with expressive timing and dynamics. Magenta. (2017, June 29). Retrieved August 31, 2022, from https://magenta.tensorflow.org/performance-rnn

Futura-Sciences. (n.d.). MIDI file. Futura. Retrieved August 31, 2022, from http://www.futura-sciences.us/dico/d/computer-science-midi-file-50004825/

Editor, M. (2022, June 29). Why artists fear rise of ai synthetic performances? mfame.guru. Retrieved August 31, 2022, from https://mfame.guru/why-artists-fear-rise-of-ai-synthetic-performances/#:~:text=A%20performing%20arts%20union%20is,AI's%20impact%20on%20the%20workforce.

Magenta. (n.d.). Magenta/Magenta: Magenta: Music and art generation with machine intelligence. GitHub. Retrieved August 31, 2022, from https://github.com/magenta/magenta