Art and Machine Learning
Project 3

Saniya Singh
Gaia Rajan
Malavika Krishnamurthy
Anna Gerchanovsky

Industry Plant

A. Group Member Backgrounds

Anna Gerchanovsky is a third year student in the IMB program for ECE in CIT.

Saniya Singh is a third year student studying Electrical & Computer Engineering in CIT.

Gaia Rajan is a first year student double majoring in Computer Science and Creative Writing.

Malavika Krishnamurthy is a fourth year student in the IMB program for ECE in CIT.

1. Concept

Over the last few decades, the process behind writing and recording music has drastically changed. For most of music history, creating songs began with lyrics and chord progressions; phrases of melodies started out being played on a piano or guitar and eventually formed into fully-realized songs with arrangements of multiple instruments and vocal layers. With the increasing popularity of electronic music genres such as techno, house, and hyperpop, the conceptualization process no longer relies solely on classic music theory (chord progressions and scales) as the foundation for writing a song. The practice of building songs off a beat has been around in the hip hop community for quite some time, and as the EDM community has grown, this method of developing music has become more popular. One of the major developments in this evolution was the introduction of the electronic drum kit, namely the Roland TR-808.

At the time of its release, drum machines such as the 808 were broadly seen as a cheaper, inferior alternative to recording a live drummer. However, the 808's accessible price and ease of use allowed aspiring musicians and artists to create music independently, and quickly gained popularity. The ability to program in rhythms manually rather than use preset loops allowed artists much more versatility and maintained a human touch in a series of sounds that was entirely electronically generated. In our project, we wanted to experiment with reversing this evolution in music creation; we wanted to take modern digitally-produced pop music and see if we could use machine learning music generation tools to convert them to use more traditional instrumentation.

2. Technique

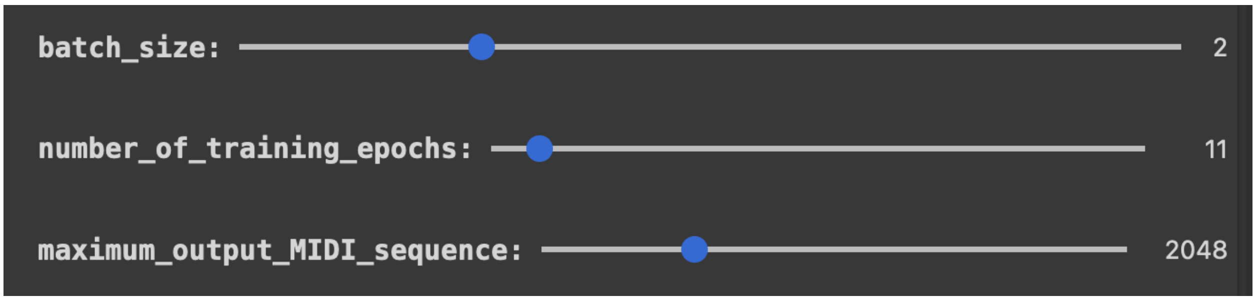
In our initial ideation for this project, we decided to focus on creating sound files that come close to the original file in melody and overall sound, but estranged from their original instrumentation. Specifically, we wanted to be able to feed in pop music to a machine learning model and emerge with similar music in classical instrumentation.

We originally searched for machine learning models that handle mp3 files as inputs, since it's relatively easy to extract mp3 files from videos or other online content. We found a model called WaveGAN that can be trained on raw audio and outputs raw audio files: the model trains on random slices from a directory of audio. We started training the model, then, during the setup process and research, realized that it was impossible to use mp3 files to isolate different instruments in a recording. In order to only access the primary instrument or output a certain instrumentation (such as piano), we'd need MIDI files, which is a filetype that separates audio into instructions. When a MIDI file is opened, it separates audio into the main instrument parts, such as backing, vocals, piano, etc. Since we wanted the output to feel melodically like the data, but to warp in instrumentation, we pivoted to using MIDI files as inputs.

While searching for midi files freely available on the internet, we found that it was difficult to find large official sets for popular commercial music. Since midi files are often used to create karaoke versions of songs or to make playable versions in games such as Guitar Hero, we went searching for those. We were able to download a few individual tracks from varying places, but it was often unclear whether the files were from some official release of the music. At one point while searching through forums, I came across a few threads on Reddit discussing a large dataset of 130000+ midi files of (relatively) sorted popular music. Unfortunately it seemed that all of the links were outdated or broken, but after finding a more recent forum post regarding this lost dataset, I followed the suggestions in the thread and was able to download the set using the WaybackMachine to search through old snapshots of the site and find the magnet link.

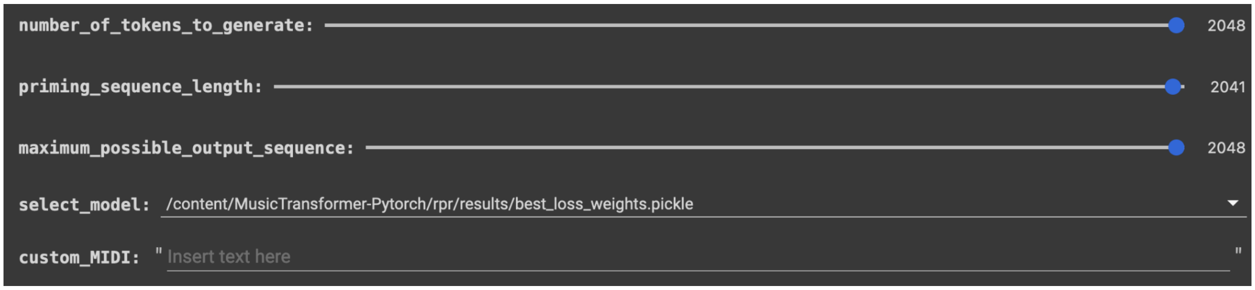
3. Process

In order to explore the effect of combining pop rhythms with classical instruments, I decided to use the SuperPiano algorithm. Although I knew this was not going to be my final choice, I began by training on the built in MAESTRO dataset. To begin, I trained it for 10 epochs, and with a batch size of 2. I raised the maximum output from 1024 to 2048 in order to allow for more flexibility in the outputs.



I tested all the weights on Toxic by Britney Spears. Toxic is a quintessential, iconic pop song, and we always want to explore exactly this while using machine learning— the popular, the iconic —as well as how it interacts with popular opinion, past and present. So, by taking a popular song and altering it with an algorithm learned from even more songs, we almost automate the process of pop— learn what to do from what has succeeded in the past. In addition, the instrumentals of Toxic were interesting to us. The song is electronic, which provides difficulty for the distinctly separated notes of piano performance. At the same time, I was interested to see how the algorithm would be able to do it, and I thought the lack of instrumental style in the electronic notes could be interesting to see altered. It also has violin instrumental parts, which both go back to our theme of classical instrument performance, but also provide an interesting challenge to convert to piano.

The results we generated were as expected: they most closely resembled piano arrangements of both songs. I generated longer output sequences rather than shorter ones, so I could hear a larger portion of the song be altered. In addition, adjusting the priming sequence length greatly changed the output. The shorter the priming sequence, the more of the output was unguided by the primer (the chosen MIDI file). I ended up recording this result a few times, and the results tended to not be very melodic. I found this interesting, but the goal of the project was more to keep altering songs to a pop standard and classical style, rather than generate completely new songs, so I stuck to longer priming sequences. After some experimentation, I used the following parameters:



After testing the algorithm on the default dataset, I wanted to try it on custom datasets. Most importantly, I was curious to see what would happen if the files it was trained on contained non-piano music. After gathering a large collection of Midi files, I ended up spending most of the time focusing on a dataset of pop rock guitar music. This was chosen for a few reasons. First, it actually provided a somewhat large variety of guitar music that was not on any genre extreme. It included the classic rock "All Night Long" by ACDC, old school rock "Johnny Be Good" by Chuck Berry, modern pop love song "Just the Way You Are" by Bruno Mars, folk rock "Cecelia" by Simon & Garfunkel, funk and boogie "Uptown Funk" by Bruno Mars , and psychedelic pop "Beautiful Stranger" by Madonna. All of these differ widely in their style and use of guitar. This variety, as well as the fact that their focus was not piano, was meant to shift the training model from a classical piano to a more amalgamated version of popular music rhythms and patterns, even if it was still performed by a piano.

This is the set of weights I stuck with for longest. I trained for 10 epochs at first, but then trained between 25 and 50 epochs, saving and generating a midi file at 5 epoch intervals, and then trained once more to 75 epochs. I stopped training at this point, as the resulting midi files began to sound overall very similar between epochs.

At 25 epochs, I ran a few variations on the generation using different primer sequences again. The result on a short primer sequence is called "rand_toxic25_ps745.mid". I kept this result for the report, but I preferred the results we got for larger primer sequences. At 75 epochs, I ran these weights with a larger variety of primers, consisting mostly of pop music but also of electronic alternative songs like "Clint

Eastwood" by Gorillaz. One of the most interesting results I got was using the primer of "Bohemian Rhapsody" by Queen. The song begins acapella, which is something I was super interested in exploring. The generated piano does a good job becoming analogous to the human voice. However, once the analog piano kicks in in the primer, the resulting song is muffled— the voice and piano combine into just the one piano note. Theoretically, the resulting song could have been playing the voice and piano in parallel, creating the much more distinctive sound of Bohemian Rhapsody. I think this may be due to the training on very different songs, where the new model was not able to accurately respond to a very stylized and specific primer file.

4. Top Final Products Gallery

Final MIDI files gallery

5. Reflection

This project is arguably an extension of themes and ideas our group investigated in our previous work; namely, we're interested in the relationships between modern pop culture and older classical elements of human expression, and the results of juxtaposition of these disparate creative impulses. We're interested in how trends shape how we experience art, and we want to complicate ideas of progress or novelty by iterating with a clear eye towards history. In particular, this time we wanted to explore the process of translating modern pop songs into a classically instrumented piano sequence using machine learning on MIDI files.

When placing typical pieces from modern pop such as "Toxic" by Britney Spears side by side with more traditional classical works such as Johann Sebastian Bach's *St. Matthew Passion*, stark differences emerge and it becomes easy to identify structural elements that uniquely define the two styles. For instance, chord progressions in pop music are generally far more predictable than in classical music, with nearly 80% of contemporary pop songs following some arrangement of C, G, Am, and F chords. Classical melodies overall have more complex melodic frameworks, a wide variation of harmonic exploration, and diverse structures (sonatas and symphonies as opposed to simply ABABCB). Classical music's composition process almost always stems from a "theme," a unique, hand-produced melodic line and instrumentation, then spirals outward in a set form of variations on that theme. The composition process requires a human origin point, and usually starts with sheet music. In contrast, pop music often begins with just a beat. Often, this is not a physical beat but a preset on GarageBand; producers, not guitarists or drummers, are the first to sketch out the movement of the music. In pop, the melodic line must be simple enough for the machine to understand and for the radio to hold. Layers of instrumentation can be added or removed with a click, and ensembles can be spawned in seconds.

The prevalence of samples in hit pop songs is also relevant here: the cycle of interpolation and sampling between hit tracks in modern pop music makes the genre feel even more connected to the machine. For example, "Toxic" by Britney Spears didn't even begin with a beat: it began with a sample from 'Tere Mere Beech Mein,' a 1981 Bollywood song. The sample is chopped from the original with precision, then machine-warped in pitch, tone, and tempo. This sample anchors the entire song, and fuels one of the most recognizable hooks in modern pop music. While classical music has self-referential traditions, composers almost never use machine techniques to filter the original sound or warp the sample. These dissimilarities are enhanced when the modern pop piece is flattened to a piano melody by an algorithm that has been trained on a different primer, also a contemporary piece. The most notable transformations modern pop pieces endure when being processed by the Super Piano 3 notebook is their simplification into a more classical form, which highlights the hallmark features of modern pop in traditional instrumentation.

The role and status a composer holds in current society from the Golden Age of classical music in 19th century Europe to modern pop has also shifted. Consider, for instance, the reputation of Johann Sebastian Bach, a German composer who died in July of 1750. While he is widely regarded as the architect of German classical music, his compositions at the time of his death began to fall out of fashion until later revived by 19th century composer Fanny Mendelssohn. In essence, society did not conflate the talent of classical composers with their status and social influence.

Compare this to modern day pop artists such as Justin Bieber, Jennifer Lopez, or Britney Spears. It can be argued that these celebrities' social power follows not from their compositions themselves but with their influence. Most artists that have created a timeless brand for themselves analogous to the Classical Era's Bach have massive followings on social media platforms and contribute to the commodification of music through constant personal branding.

The mere magnitude of such celebrities' social following and idolization at events such as the Met Gala or Grammys allows them to mold aspects of social culture such as contributing to beauty standards or even shaping colloquial language. For example, American rapper Sergio Giavanni Kitchen, known professionally as Gunna, normalized the term "Pushin P" through the release of his new track *Pushin P* in 2022. The slang is now widely used by young adults and teenagers to indicate that a particular activity or individual is "cool". Titan rap artists such as Drake and Future often glamorize their private jets and exclusive access to elusive clubs worldwide, setting a gold standard for success. As music becomes more mechanized, influence and access also shift; pop music now begins with a machine, rather than a composer and a sheet of paper, changing the way we consume and engage with the art form entirely.

This project granted us the opportunity to reverse the defining evolution of modern pop music, electronic instrumentation, and return it to the classic forms of melodic instrumentation; it also allowed for a space for social commentary on how modern pop music and all its intricacies and interdependencies support social and economic structures such as capitalism and consumerism – a feature in which Classical music most definitely did not engage.

6. Future Work

In the future, we hope to expand about the styles of music that can be generated. While this project focused primarily on this "reversal" and creating pop songs with purely traditional instrumentation, it would be interesting to expand upon that by exploring computer-generated electronic music. For this project, we surveyed many different music generation algorithms. We did not have enough time to train and test all of them for our project, but we found a few observations about the methods used in this area. There were two primary types of ML algorithms: those using pure audio forms as source material and those using midi files or stem packs as source material. This shows a certain divide in the way that we currently conceptualize making music; on one hand, it may be enough to create a vague stylistic sonic imitation (take Jukebox as an example) of the source material. On the other, using midi files allows for a far more complex and structural imitation in the generated material. Since we chose to use MIDI for this reason, we also found that most projects generated outputs using classical instrumentation: strings and piano. While researching we came across one recent project entitled 'Deep House';, which seemingly generated EDM music using a complex algorithm in conjunction with Jukebox's waveform processing. Since we initially discussed more niche EDM based generations during our ideation stage, I would hope to be able to expand this study by venturing into computer-generated digital music, perhaps trained on MIDI. It has been difficult for experts to pinpoint precise formulas or structures that exist in effective, successful hyperpop music, and I am curious to see whether a learning model might be able to identify and imitate some.

We could also use a double-pronged process to generate classical arrangements of popular music: since MIDI files are often not available for free for certain pop music, we could train a model to accept MP3 files and stratify them into layers, then conduct the usual preprocessing and generation in Superpiano. For example, the Audio to MIDI package is a convolutional neural network that takes mp3 input and outputs a MIDI with the corresponding notes and durations. This has the potential to make our results more varied, since we would then have access to a far larger canon of contemporary pop music and could train and test on more recent hits. However, it can also introduce another source of noise, and if the first network imperfectly isolates the sets of sounds in the mp3, the corresponding instrumentation warp will differ from the melody. This step would bring us closer to the eventual goal of completely automated Automatic Music Transcription.

WORKS CITED

In Report:

https://interlude.hk/death-brings-success-musicians-and-artists-who-found-recognition-after-death/

In Code:

https://github.com/gwinndr/MusicTransformer-Pytorch

Songs as Primers:

Toxic by Britney Spears

I Want it That Way by Backstreet Boys

Bohemian Rhapsody by Queen

Clint Eastwood by Gorillaz

Just Dance by Lady Gaga

I Knew You Were Trouble by Taylor Swift

Turn Me On by David Guetta

Wannabe by The Spice Girls

Source For Dataset:

https://www.reddit.com/r/WeAreTheMusicMakers/comments/3ajwe4/the_largest_midi_collection_on_the_internet/

https://www.midi.org/forum/10727-midi-collection-from-reddit-is-no-longer-available