Abstract

Music composers commonly compose music that is inspired by the sounds of another composer. When a composer writes a new piece of music, sometimes the rhythm or choice of pitches can be similar to the existing songs. However, the song can still sound original. In this paper, we introduce a similar approach by using two base songs to create a new melody. One base song provides the rhythm for the new melody. Conditional probabilities of note transitions are extracted from the other base song. The melody generation algorithm uses a probabilistic approach to generate a new melody with similar conditional probabilities of note transitions. The outputs of this approach are evaluated by a human being to measure how human-like, repetitive, and random they sound. Findings were that there is a consistent amount of randomness and repetitiveness required in order to consider a song human-like while not sounding too similar to the original base songs.

Introduction

When it comes to music compositions the melodies that are written are made up of many previous observations of rhythm and melodies. Musicians observe patterns, and based on these patterns, they are inspired to make something new and original. It has happened time and time again where a label wants to sue an artist for using something that is too similar to one of their own artists. More often than not, these situations could be subconscious accidents where an artist creates an original piece that maybe has its own melody but the same rhythm from another song.

This begs the question, can we use AI to come up with something new and original using components from other songs? Is it possible that new genres and musical ideas can come from the combination of parts from multiple song sources? After all, this is an important component of the creative process for a musician, to be inspired by others ideas in order to create their own.

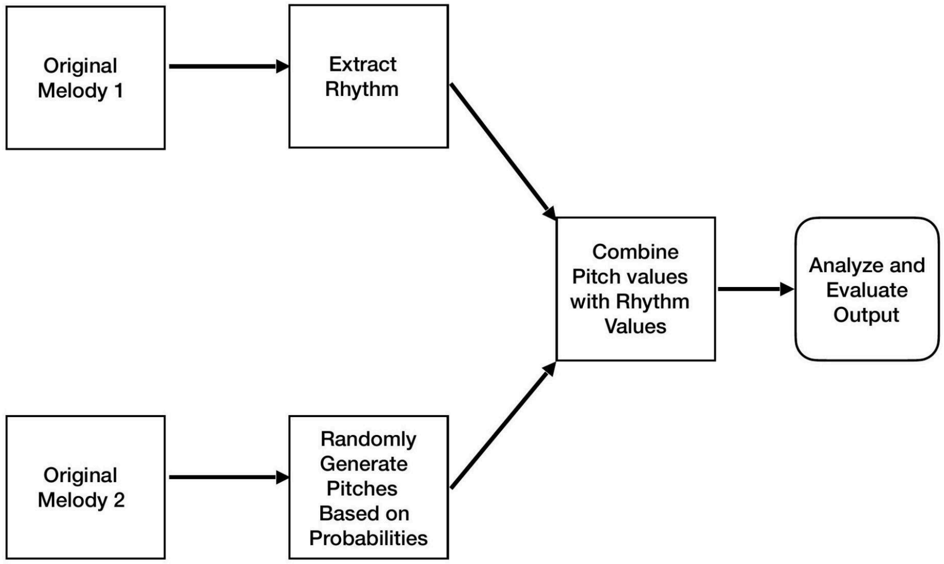
In this paper, we will explore and propose an algorithm to do this; generate melodies using a generative bayesian algorithm with the rhythms of other songs. The algorithm will randomly generate and select notes based on the probabilities of one original song, and then apply them to the original rhythms of another song. Once we have generated the melodies, we can make judgments on the melody based on how well the phrasing matches with the rhythm, the structure of the entire piece, and how human the melodies generated sound. This idea could be used to effectively generate style specific melodies for any choice of genre.

Results showed that human beings require a certain amount of sense of randomness and repetitiveness in order for a song to be recognized as a human generated melody. If we have a feeling that we can predict everything that we are about to listen to, then the song will sound less human and the same with randomness If we can predict nothing, then the song will sound less human. The songs that sounded more human-like also generally did not have a high similarity rating to the original which shows that this method of using another musician's rhythm with our own generated melodies can still produce authentic original sounding results.

Methodology

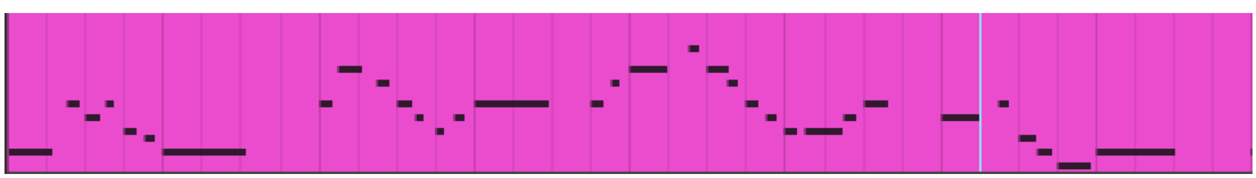
In order to generate human-aware melodies the method we have used is a generative bayesian approach based on using probabilities of notes that follow one another. Because we are generating monophonic melodies, the first part of the process is to pick an appropriate data set that would give probabilities which would be most similar to what a human being would be most likely to pick. For example, using a dataset of popular mainstream songs would not be a good choice because of the amount of repetition within the melodies. Using these probabilities would generate melodies that have a high amount of repetition that they sound less human-like. We needed to find a dataset that accurately would give us probabilities of how a human might phrase their music. Therefore, we have chosen a dataset of scores of solo pianist performances and then we use the top line melodies from these pieces.

Figure of Process to Generate Melodies

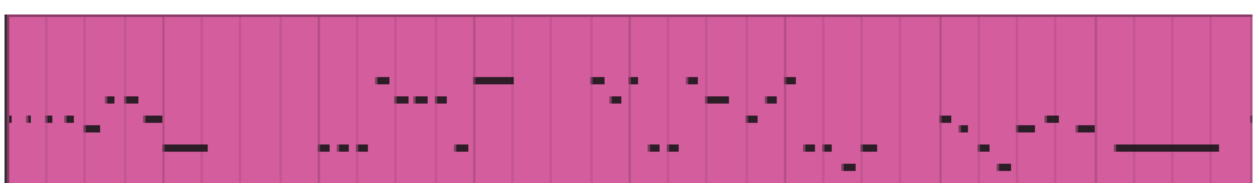


The next part of our process is that we will randomly choose two songs from our dataset. One of our songs will be used to extract the rhythm to use for the melody we will generate. For our other randomly generated song, we will extract probabilities of how likely the distance is between two pitches to occur. Using other methods such as the probability of three notes in a sequence did not seem to work as well and sounded less human. This was because the distribution of probabilities had a high likelihood for some sequences and a low likelihood for other sequences with not much in between. It is common for musicians to use arpeggios (a three note sequence of a broken down chord) when they play music but the other sequences are often more unique in their occurrence. Therefore the choice of using probabilities of one note following another gave us the most human sounding results in our pitches.

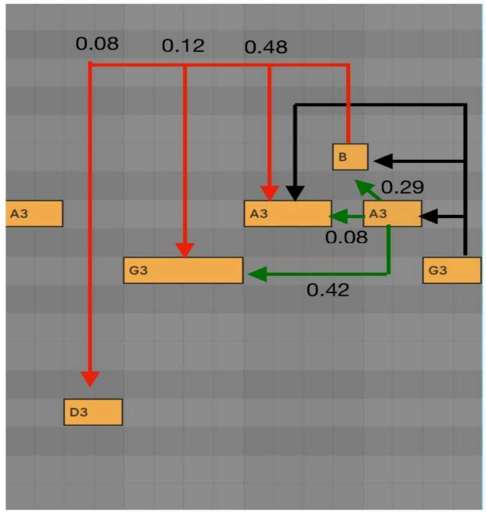
Midiclip of original Rhythm to be used



Midiclip of probabilities to use



Transition probabilities of note transitions



Upon generating these pitches we found that there was an issue with resolve in the generated melodies. When musicians play music, the pieces they will play often resolve the end of a song on a root note (the first note in the scale of the current key) or sometimes on the 5th. To address this issue we reversed the melody and then generated the probabilities so that the song would land a note that would give a sense of resolution. After generating a list of these pitches, we can apply our rhythm, and we have our generated melody.

Table of Note Durations and rests to be used

	duration	rest
0	567.0	0
1	166.0	153
2	206.0	74
3	127.0	34
4	167.0	113
5	155.0	73
6	1025.0	85
7	177.0	895
8	318.0	63
9	155.0	162
10	194.0	85

Generated Pitches to be applied

	pitch
0	63
1	65
2	67
3	67
4	70
5	72
6	70
7	62
8	60
9	60
10	60

While we have generated this melody based off of these probabilities and rhythm we come across the issue of there being a lack of human phrasing within the melody. In order to address this we generated a number of songs. Then we ran an algorithm on the pitches to see what the longest repeating substring would be. The idea behind running this algorithm would be to see how the phasing sounded in a song and then associate that with the length of the substring. Substrings that were longer (of length greater than about 10) tended to be far too repetitive. Substrings that were too short did not seem like they had a sense of phrasing and sounded too random in respect to rhythm and choice of pitch. The optimal length seemed to be around a length of less than 10 and greater than 5. Melodies that had a length of this size tended to sound more human as they had a more human-like movement to the phrasing, such as a phrase moving down to a note and then reversing itself and moving back up the set of notes.

The last part of the process is to generate a handful of these midi clips with this methodology. Then we pick a criteria to judge the songs on how human-like they sound. Using a 1 - 10 scale where 5 would be the optimal range to be in. For example in terms of how repetitive a song sounds, we don't want a song to not have any sense of repetition, but we also do not want it to have too much repetition, so therefore, having a score of 5 would be the best score.

The criteria used is as follows:

- How random did the pitches sound?

  How well did the generated pitches match with the existing rhythm? Would the song be likely to be striked with copyright?

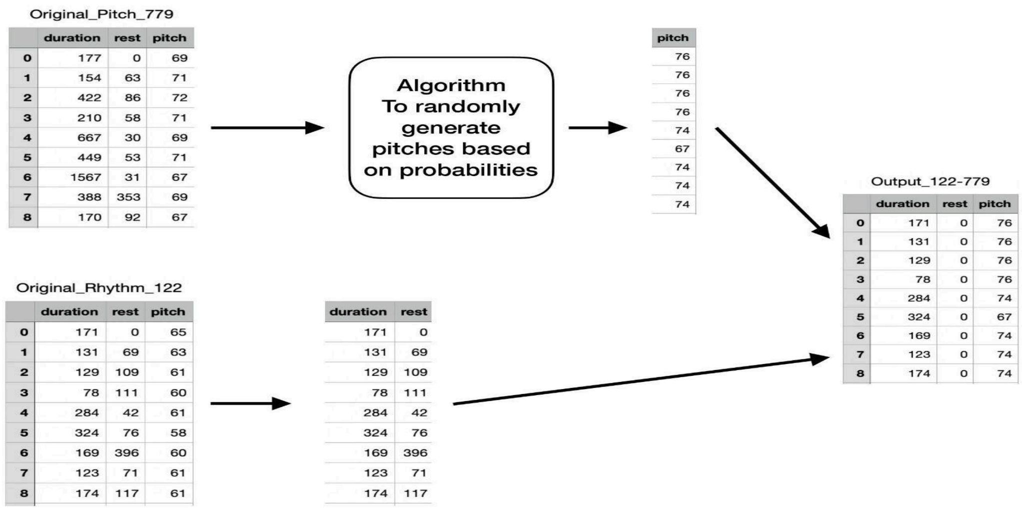
- How repetitive did the pitches sound?

  Was there too much repetition to the point of the melody not sounding human? Would the song be likely to be striked with copyright?

- How well did the pitches match with the rhythm?
- How human-like did the phrasing sound in the melody?
  - Based on the previous two questions as well as the length of the longest repeating substring, how human-like did the phrasing sound in the generated melody? How did the beginning and the end of the song sound…? Was there resolve?

- What is the over score for how human the new melody sounds? (10 being the most human)

Side by side comparison of original melodies and new melody:



Literature Review

When it comes to generating melodies, our approach is simple and effective, and logically quite similar to how a human being may create these melodies. A musician comes up with a note to start with, then imagines what note to move to next based upon previous experiences with music, and then repeats this step until there is a sense of resolve in which then the idea is complete. But human beings are unique and have their own form of creativity which is what we need in our AI for it to be considered more human-like. There is also the issue of style in that musicians have to adapt to the situation and sometimes drastically change the choices of notes that they make for whatever the particular situation might be.

There have been some other much more complex approaches to try to solve these problems. Specifically to address the issue of a style, reinforcement learning has been used to create a neural network that is style-specific (a style can be chosen as a preset to be used). An example of this would be a combination of a GAN (General Adversarial Network) and a probabilistic network. The idea behind this is that the probabilistic network acts as a “knowledge of theory” and the GAN critiques the output based upon whether or not it is of the style that the user has chosen. GANs in the past have had the issue of generating music that was too similar. To solve this issue, the reward function can be used to optimize the probabilistic network in order to create a healthy diversity of output. While this is an effective method for generating music that is qualified for a specific style, when judged by a human, there have only been about thirty percent of acceptable samples.

Similarly with reinforcement learning, RNN’s (Recurrent Neural Network) have been used in RL-Duet to generate music as an accompaniment to a human being. This algorithm learns by generating a note based on the previous context of the previous note created. This algorithm also considers how compatible the note it generates with the human generated note. There are multiple reward models used in order for this algorithm to be able to perform in real-time. While previous models have been based on composition rules and heuristics hand-crafted by human specialists, these reward functions are based in deep learning neural networks that require no prior musical knowledge to create its models. Due to the design of these reward models, this algorithm has been successful in generating quality melodic, harmonic, variety rich parts that perform better than previous models.

Another interesting approach to generating unique style specific music is Open AI’s Jukebox. Unlike how most models are designed to train and produce midi, this network uses audio files of other songs as its data. This network uses a hierarchy of VQ-VAE’s that can adapt to any style of music that is provided. The AI can generate music in two ways; it can generate entirely from scratch with a specified style, or it can be given a part of a song and generate the rest. What is also interesting about the music it will generate, is that it will generate each component of the music from similar artists from that style. For example, if you gave the AI the first intro and verse of a Nirvana song, it might generate new vocals using a voice similar to a similar band to Nirvana such as maybe Alice in Chains.

This design is quite similar to how many musicians have created their musical style in their projects. For example if we look at the style of the band Greta Van Fleet it’s clear that the lead singer Josh Kiszka developed his style listening to lots of Robert Plant. However the guitar, bass, and drums have a more modern sound to them, sounding something more like the band Jet. But even though we might be able to say that the vocals are so similar to Robert Plant and the rest of the instruments sound similar to Jet, they still have an original sound.

These methods have also been a common issue between labels and artists where a producer uses something that is so similar to another producer, that they found it necessary to take legal actions against the other party. A classic example of this is the case between Vanilla Ice’s song “Ice Ice Baby” vs Queen and David Bowies “Under Pressure”. Comparing the two songs it doesn't take a musical genius to see that they both used the exact same bass line in the song. Vanilla Ice, when he produced “Ice Ice Baby”, he did use exactly the same bass line where he in fact sampled the part and placed it in his song. The two parties decided to settle outside of court where Vanilla Ice decided that it would be cheaper to just purchase the copyright instead of spending the money on court fees.

This idea that Vanilla Ice had of resampling was something that would be used as a foundation in creating hip hop and commonly used today in hip hop and rap as well as many other genres. Most hip hop songs mainly consist of resampled music but to avoid copyright the samples are changed or used in a context that they are hard to recognize. Another interesting example of this methodology of producing is the well known EDM Artist Madeon who first went viral posting the song “Pop Culture” which he produced out of 39 samples pulled from other songs. This style of production has influenced other new sub genres in the EDM community as well.

Another interesting case to look at is the case between The Taurus’s “Spirit” and Led Zeppelin’s “Stairway To Heaven”. This case was different because the claim came from Taurus claiming that Jimi Paige used the same arpeggios from the beginning of their song to create the beginning of Stairway to heaven. This is something that can subconsciously happen where if two artists happen to use the same chord progression and they choose to play the arpeggios of those chords, they can easily end up sounding similar. Therefore this case was dropped and Led Zeppelin kept all the rights to their song.

Sony’s Flow Machines have done something somewhat similar using chord progressions or similar chord progressions of another artist to generate music. For example, these machines generated the song “Daddy’s Car” in the style of the Beatles. The song sounds quite similar to the Beatles, however the song still sounds original in its nature. While the song had interesting structure as well as melodic composition, the song still didn't get much attention by the public eye.

Sony’s Flow Machines have been using similar methods to generate music using a human to input a couple of measures of notes (chords or melody), pick a certain mood, and then it outputs a melody that can be used. Similar to methods previously mentioned, the models work using a combination of probability networks and generative models. These machines along with the collaboration of human beings performing the music have created the first AI generated album “Hello World” by SKYGGE”. The album has some sounds that are similar sounding such as vocals and chord progressions from the beatles. But the result of the generated melodies gives an unexpected but not unpleasant experience. It could be the first glimpse into a new genre as the music sounds quite unique with the interesting combination of rhythms and melodies. Oftentimes, this is how new genres are created, a combination of styles combined into one new style.

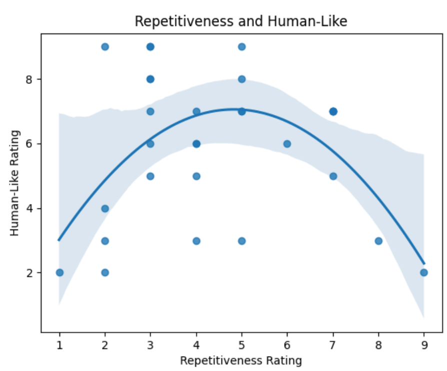
This could push music production into a new direction as we know it automating some of the process of music production. For example Sony’s Flow Machines have software where if a producer can't come up with any ideas for a chord progression they like, they just need to input the chord progression and level of energy of what they have and then if they are satisfied with the result and decide to keep it, they have overcome one of the most time consuming hurdles of producing music. All the while keeping an original authentic sound to the production of their music.

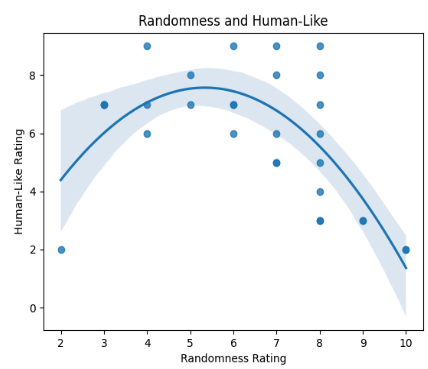
Results

Songs were rated on three dimensions: randomness of melody, repetitiveness of melody, and degree of humanness in the melody. This last metric reflects our interest in understanding if the melody is artificial and machine-like sounding, or if it sounds similar to a human composition. After rating a selection of 30 random songs, results showed that the repetitiveness and human-like ratings have a curvilinear relationship where middle scores (i.e. an average rating of 5) for repetitiveness are associated with higher human-like ratings.

Nearly the same outcome was evident with randomness. The closer to an average rating of 5 to the randomness rating was, the higher the human-like rating also was. Therefore, both randomness and repetitiveness have similar relationships with human-like ratings. Low scores and highscores on each of repetitiveness and randomness are associated with low human-like ratings. The highest human-like scores were associated with average scores on repetitiveness and randomness. See graphs below illustrating both of these patterns.

We conclude that for a human being to find this melody to be natural and human-like, there has to be a sense of something unexpected but not completely random. Repetitiveness follows a similar pattern: a moderate sense of repetitiveness is required for a song to be rated more human-like.





Conclusion and Future Work

As we can see, when we take rhythm from one song, and simply generate a new melody over it, and come up with a new melody that sounds authentic and original. Using the generative bayesian method looking at the probability of one note following another, proves to be an effective method of giving the AI a sense of “knowledge of theory”. While not all songs had a sensible sense of human phrasing, most songs did have a sense of dissonance that ended in resolve. Then we use human judgment to rate metrics such as, similarity to the original pieces, randomness, repetitiveness, and how human-like the output is.

The results showed us that for something to sound more human-like, we need to have a decent amount of randomness. However, too much randomness begins to sound nonhuman. The same correlation happened with the rating of repetition to how human-like the songs sounded. Similar to the randomness ratings, the ratings that didn't have too much randomness, and too little randomness did the best when it came to what sounded the most human-like.

For some future work something that might improve the results is to limit songs matched with each other to having a bounded tempo. For example, having songs matched together between 100 and 130 beats per a minute, or songs that are matched together between the tempos 140 and 170 might have a better yield on what sounds more human-like. This would be because in slower tempos, a musician may be able to fit more notes within a phrase which could change the distribution of notes. In a tempo like 170, there would be less notes, so note transition probabilities such as a root to a 5th, might be higher.

Using a GAN (General Adversarial Network) may also fix this issue. The GAN could be trained on these tempos due to the repetition of the probabilities in certain tempo ranges. Therefore, when something is generated that is too random or repetitive the output would be rejected by the network. This could potentially be of more use of an algorithm due to a more consistent phrasing and could be used to write full pieces with a chordal polyphonic section written by a human making the process of writing and recording music easier for the producer.

References

Alex. (2022, June 29). *Classic copyright cases – ice ice baby*. Briffa Legal. Retrieved September 10, 2022, from https://www.briffa.com/blog/classic-copyright-cases-ice-ice-baby/

Blistein, J. (2020, October 5). *A new Led Zeppelin court win over 'Stairway to Heaven' just upended a copyright precedent*. Rolling Stone. Retrieved September 10, 2022, from https://www.rollingstone.com/music/music-news/led-zeppelin-stairway-to-heaven-copyright-infringement-ruling-appeal-964530/

itsmadeon. (2011, July 11). *Madeon - Pop Culture (live mashup)*. YouTube. September 10, 2022, from https://www.youtube.com/watch?v=lTx3G6h2xyA&ab_channel=Madeon

YouTube. (2021, March 5). *Ballad of the Shadow*. YouTube. Retrieved September 10, 2022, https://www.youtube.com/watch?v=O9FUSmtP0Ug&list=PLDLD_n0Am-yNN9j-L7x7dna9qYBenQ-QQ&ab_channel=SKYGGEMUSIC