

Do Androids Dream of Sonata in C?

Lauren Funston

Abstract

The recent advent of generative artificial intelligence (AI) has led to discussions around the role of technology within creative fields. One field of particular interest is music, which artificial intelligence researchers have been attempting to emulate for decades. Music has long been characterized as a precise and logical algorithmic endeavor, a view exemplified by the field of music theory; however, it also carries a spiritual and emotional history as a core element of the human experience. This dual nature makes music a perfect case study for examining the relationship between artificial intelligence and traditional art forms. This article explores the roles of both human cognition and mathematical logic in music composition to construct an understanding of how generative artificial intelligence interacts with traditional art forms. Ultimately, while generative technologies are now capable of producing and perhaps even enhancing musical compositions, they continue to fundamentally lack the emotional nuance necessary to produce truly moving works of art.

Introduction

When it comes to the ability to shape human emotion and behavior, nothing quite compares to music. Our representations of life via plays and movies are typically accompanied by a soundtrack for every moment, reflecting its ability to color our memories. It's hard for most of us to imagine a world without melodies and rhythm – even deaf

people can enjoy songs by feeling the beats pulse through their bodies. Human life has rarely been devoid of music – the first instruments were made over 40,000 years ago (Schyff, Schiavio, & Elliott, 2022, p. 4), and there is evidence of music-making in ancient cultures. The ancient Greeks “saw music (along with dance and poetry) as a driving force in the emergence of their people and culture” (Schyff, Schiavio, & Elliott, 2022, p. 6), emphasizing music's role not only as an illustrator of our individual lives, but as a powerful social force.

Music has been historically tied to technological advances. Instruments are tools that aid in the music creation process, and as these tools become more advanced, more musical sounds and styles are possible. A new instrument can completely change the direction and sound of a society's music. Just compare the electronically inspired beats of the new millennium with the sweetly lilting melodies of the nineteenth century – each of these styles of music was spurred by the creation of new technologies, such as the Musical Instrument Digital Interface (MIDI) standards and the harpsichord. However, these instruments still rely on human input to generate music; a guitar can't generate chord progressions or syncopated rhythms on its own.

Music's reliance on human input has led to a fascination with musical automata, and inventions like music boxes, record players, and jukeboxes have all given the illusion of technologically created

music (Roads, 1980, p. 4). However, behind all these automata remains a human composer – it is understood that these automata, advanced as they may be, are still incapable of producing original works. We don't see these computers as composers, but as mediums through which we experience music. In the past few decades, computer programs have been written that claim to generate and analyze music, with mixed success (Meehan, 1980, p. 60). Computers can take as input the rules of music theory that seem to prescribe musical success with technical precision; and yet, the music they produce can seem robotic or unemotional. But music is subjective, and one could argue that much of the music popular today could also be described as robotic and unemotional. If machine-generated music is just as valid as the simple, formulaic pop songs cluttering radios around the world, then music may not be as unique to the human experience as we once believed.

Rather than accepting that music is a wholly human or wholly mechanical process, I believe that a compromise between the two can be found. While computers are now able to produce songs that become popular and should rightly be considered part of the modern musical landscape, they are far from matching the emotionality and affection that a human musician is able to cultivate in a full body of work. Note that just because human musicians can produce this effect does not necessarily mean that all of them do so. Moreover, music, like any art form, is subjective in nature, and thus music that is generated by non-human entities should not be excluded from the canon. However, the human experience of music, being as closely tied to emotion as it is, is much harder to define in a computational sense. Though we have algorithms that attempt to define music in terms of genres and moods, these categories are hotly debated among humans ourselves; arguments over whether a band is truly “indie”, or whether a particular song is good or bad,

define most of the musical discourse in our daily lives. By exploring historical conceptions of music, the current and future capabilities of computer programs with regards to music, and how humans process music, I hope to show that while generative technologies can be used (often in collaboration with humans) to produce music, they still rely on some form of input from humans themselves to create works with real staying power. Ultimately, human musicians should be reassured that their jobs are not going to be taken by computers; at least, not any time soon.

Historical Conceptions of Music Creation and Appreciation

Debates over the classification of music as a mathematical process or natural phenomenon have permeated philosophical debate for centuries. Plato argued that music was determined by the “mathematical principles that govern the rules of musical relationships” (Schiff, Schiavio, & Elliott, 2022, p. 6) rather than the physical experience of sound. His view was influenced by the work of Pythagoras, who, along with discovering various mathematical relationships, also began to define the musical relationships that would later become the field of music theory. This rational view of music contrasts with the view of Aristotle's pupil Aristoxenus, who argued that “although aspects of music can be described in terms of abstract mathematical ratios, music cannot be reduced to them, nor are these aspects required for musical phenomena to be perceived, understood, and judged” (Schiff, Schiavio, & Elliott, 2022, p. 8). Aristoxenus highlights that many musicians, especially in the ancient era, created music without any knowledge of these relationships, and likewise, members of the public were able to appreciate this music without any formal education. Both opinions are still applicable today – though the field of music theory has certainly advanced in the past centuries,

the ability to appreciate musical works does not rely on even a basic understanding of the fundamentals of rhythm or melody. It also cannot be said that all musicians are masters of these musical relationships, as many successful musicians do not have any formal training. Despite this, a rational basis for music does indeed exist and is evident in the development of the field of music theory since the early days of Pythagoras.

The invention of musical notation in the West provided an encoding for music, encouraging a more systematic approach to music composition. As Roads (1980) notes, "...one of the major tasks of composition is creating a rule system... for a piece" (p. 14). In the Western world, musical composition began to be seen as a process of defining rules and composing musical material that follows these rules. A particularly creative composer may define a uniquely complex set of rules, including key, chord sequences, rhythm scheme, and more, but these rules are always present in any meaningful piece of work. Even in works that attempt to circumvent such rules, such as atonal or arrhythmic pieces, the strict avoidance of form is a "rule" of its own. This view of music creation appears especially amenable to the ways in which computers operate. Computer programs may be seen as similar sets of rules that are written by humans, and therefore it does not seem impossible for a computer to carry out the task of composition based on a set of provided rules. However, as Schyff, Schiavio, and Elliott (2022, p. 18) point out, the "traditional focus on the relationship between composed form and expression" is not necessarily present in non-Western or other traditionally marginalized musical communities, many of which produce music in a more communal, unstructured process. Jazz improvisation is a prominent example of this type of creative process, with jazz bands creating music on the spot in a "call-and-response" pattern reminiscent of conversation, requiring on-the-spot collaboration

between all musicians involved (Smithsonian, n.d.). These processes involve interactions between many different creative agents, and the result is music that in some cases cannot be defined by precise rules and notation. It is less clear, then, how these processes could be mimicked by a computational process.

Current and Future Capabilities of Computational Music

For better or worse, much of the efforts towards computational music generation have indeed been centered around the Western view of composition, utilizing modern staff notation to provide clear inputs and operators for the computational "problem" of music generation. But despite what seem to be clear links between musical "language" and natural languages, surprisingly few efforts to generate music have been successful or generalizable on a wider scope. Meehan (1980, p. 6) notes that although computer programs existed as early as the 1950s to generate nursery rhymes, cowboy songs, and other simple tunes, many of these programs did not rely on music theory as much as the specific structure of type of music they were trying to replicate, resulting in highly specialized systems without much opportunity for generalization. Indeed, it appears that there is not a similarly robust model of music as exists for human language.

Of course, these simplistic programs are not entirely unimpressive. Meehan (1980, p. 64) also writes that the I-vi-IV-V7 chord sequence used in many pop songs of the 1950s is similarly repetitive and uninspired, yet it produced some of the most popular songs of the decade. Clearly, this style of songwriting follows strict rules that could easily be followed by a computer program, and the music it creates is widely appreciated and adored. Thus, it seems that a more generalizable model of music might not be necessary for computers to create music that is enjoyed by human listeners.

And indeed, in recent decades, computer

generated music has exited the realm of research labs and entered the public consciousness. The phenomenon of “Vocaloid” singers in Japan – virtual idols whose voices are entirely computer-generated – have topped charts and held concerts across the country for the past decade (Collins, 2011, p. 36). More recently, the phenomenon of producing “AI covers” has taken off across social media. If you’ve ever wondered how your favorite singer would cover a certain song, you now use the power of AI to cook up a sufficiently accurate representation. Nevertheless, it’s important to note that for all these technologies, human input is still required. Samples of real human voices are used to create Vocaloids, and their songs are still written by human composers; the same is true for AI covers, which use the voices of human singers to make covers of human songs. When it comes to a computer’s ability to independently produce a song, the results are mixed and often come across to human listeners as robotic and emotionless.

The ability for computers to analyze or appreciate music is even more fraught, and most early work in this field did not even attempt it; as Meehan wrote in 1980, “There are very few analysis programs, good, bad, or otherwise...” (p. 60). This deficit could once again be attributed to the lack of a robust model for musical “language”, but it likely runs deeper than the simple lack of a model. The experience of music is largely irrational, and the interpretation of a piece of music depends on the individual experience of the listener in a way that differs from language. While language can (and does) have different individual interpretations at times, there is only so much that you can dispute the meaning of a simple sentence such as “That dog is brown.” By contrast, what a musical stanza “means” is constantly up for debate, and it is extremely difficult to create a model that adequately ties the semantic primitives of music (notes, rhythms, etc.) to higher-level interpretations (Meehan, 1980, p. 64).

In this way, the problem of interpreting music is a significantly harder problem than the computational generation of music.

Study of Human Cognition Related to Music

Research into large language models has typically included examinations of how the human mind understands and interprets language; thus, it should be no surprise that understanding how humans interact with and process music is necessary for building a more comprehensive model of music theory. As discussed earlier, music has been categorized by some as an extension of human language, as it can be broken down into syntactical elements that represent semantic meanings. For example, a chord played on the piano is syntactically made up of a series of notes played at the same time; however, the emotions evoked by that chord are dependent on the combination of notes that are played and the context they appear in (Meehan, 1980, p. 62). This has traditionally been the dominant view of explaining human cognition of music, and drawing parallels to language processing can be helpful, especially because it allows us to draw upon previously defined models. However, this view is not without its drawbacks. As Schyff, Schiavio, and Elliott (2022, p. 15) note, significant differences in musical and linguistic perception have been observed. Thus, it is unlikely that relying solely on the same model of language perception to accurately describe musical perception will be sufficient to produce musical models of the same caliber as current large language models.

In addition, tying music processing to language processing doesn’t account for the multitude of factors besides traditional mental processes that influence how music is experienced. The most obvious example is motor processes – music is often accompanied by dancing, and research has shown that there is a significant connection between

musical cognition and motor processes (Schyff, Schiavio, & Elliott, 2022, p. 13). Reconciling the process of understanding music and translating it into movement is something that a simple language model might not be able to account for. Sociologists also tie musical understanding to the lived experience and social circumstance of the individual listening (Schyff, Schiavio, & Elliott, 2022, p. 17), indicating that an individual's understanding of a musical work is formed on a personal basis. There is no one "true" meaning of a piece of music, and attempting to define one may cause more harm than good. Finally, humans tend to understand music as primarily emotional rather than logical, and it takes a substantial amount of training for humans to recognize the rational structural elements that make up music (Schyff, Schiavio, & Elliott, 2022, p. 20). This is not to say that these elements don't matter, but rather that they are not strictly necessary or even influential in determining the emotional response an individual has to a piece. Because the experience of music is made up of so many different parts, many of which occur on an individual scale, it is hard to imagine a program capable of analyzing music in terms beyond pure rational structure. Even as humans, the task of defining a common interpretation of a musical piece to apply to all situations and contexts appears impossible.

Conclusion

We have seen that the mathematical relationships guiding musical composition are indeed real, and computers can use them to generate songs that become immensely popular. Even just this year, "Heart on My Sleeve", a song written by a human "ghostwriter" but sung by AI-generated voices of The Weeknd and Drake, went viral online and was submitted for a Grammy award (Shanfeld, 2023). Although this example may seem to represent a gross overreach from new generative technologies, I believe it highlights the relevance

of human musicians even in this emerging musical environment. The song was still composed by a real human writer, whose real human experiences undeniably helped to propel the song to success, despite the use of AI-generated voices.

Indeed, it is hard to imagine a full system capable of all that humans can experience and understand through music. An individual's understanding of music is influenced by their current environment, lived experience, socioeconomic status, background knowledge, and more. Even if a computer were able to "understand" music, it would be similarly influenced by its own experiences, and while these experiences may be far removed from those of a human's, they will nonetheless shape the analysis of such a machine. A universal analysis or understanding of any musical piece is impossible, as so many of these factors are non-generalizable. Thus, while computers can perform the act of generating music, they are unable to fully understand and resonate with it in the same way as humans, and thus their ability to create pieces that will truly resonate with human emotions and experiences is unquestionably stunted.

Thus, while the prospect of computer-generated music may cause fear for human musicians, afraid that these AI-powered machines may be able to make music at rates and quality far surpassing their own abilities, I believe that because music is such a personal experience, there is still a glimmer of hope. Human musicians need not fear being completely replaced by computer musicians, because the subjective experience of music means that people will always have different music styles and musicians that they appreciate and enjoy. Rather than being replaced by machines, human and machine musicians can coexist or even collaborate. Because there is no objective measure for the quality of music, there is no reason that either type of musicians will necessarily be "better" – in fact, human musicians will continue to have the

upper hand for producing music that appeals to human audiences, for the simple fact that they are human themselves. Most importantly, there is no “turning back the clock” on this technology, and it will inevitably be a part of the future of music, whether we consent or not. Instead of running away from the idea of electronic musicians, we should work to better understand how they can be used in collaboration with human musicians to strengthen their works. In this way, we can help to usher in a new era of music, an era in which the entry of computer musicians into the musical arena is an exciting prospect that enhances the diversity of today’s musical landscape.

References

- Collins, N. (2011). Trading Faures: Virtual Musicians and Machine Ethics. *Leonardo Music Journal*, 35-39.
- Meehan, J. R. (1980). An Artificial Intelligence Approach to Tonal Music Theory. *Computer Music Journal*, 60-65.
- Roads, C. (1980). Artificial Intelligence and Music. *Computer Music Journal*, 13-25.
- Schyff, D. v., Schiavio, A., & Elliott, D. J. (2022). Getting Situated. In *Musical Bodies, Musical Minds: Enactive Cognitive Science and the Meaning of Human Musicality* (pp. 1-24). MIT Press.
- Shanfeld, E. (2023, September 6). Ghostwriter’s ‘Heart on My Sleeve,’ the AI-Generated Song Mimicking Drake and the Weeknd, Submitted for Grammys. Retrieved from *Variety*: <https://variety.com/2023/music/news/ai-generated-drake-the-weeknd-song-submitted-for-grammys-1235714805/>
- Smithsonian. (n.d.). What is Jazz? Retrieved from *National Museum of American History*: <https://americanhistory.si.edu/explore/projects/smithsonian-jazz/education/what-jazz>