3-2024

Harmonic Data: AI Music, EDM Transcription, & Minimalist Jazz

Trevor Freed

Harmonic Data: AI Music, EDM Transcription, & Minimalist Jazz

Abstract

This thesis presents research in the domains of music theory and music technology. The topics of the three chapters are as follows: (1) the technical and artistic limitations of artificial intelligence when generating music, (2) the transcription of electronic dance music (EDM) using a hybrid notation system, and (3) the effective convergence of jazz and minimalism in the music of Joe Hisaishi from Studio Ghibli. Each topic investigates a different aspect of harmony and the modern ways in which humans communicate in terms of harmony. The first chapter surveys the latest research projects in AI-generated music and proposes ideas for future applications of music generation using machine learning. The second chapter presents a hybrid notation system for the transcription of electronic dance music. The final chapter identifies progressive composition techniques embodied in the works of Japanese composer Joe Hisaishi, who borrows musical influence from around the world.

Document Type

Masters Thesis

Degree Name

M.A.

First Advisor

Mitchell Ohriner

Second Advisor

Kristin Taavola

Third Advisor

Sean Friar

Keywords

Artificial intelligence (AI), Electronic dance music (EDM), Hisaishi, Minimalism, Music, Transcription

Subject Categories

Arts and Humanities | Composition | Music | Music Theory | Other Music

Harmonic Data: AI Music, EDM Transcription, & Minimalist Jazz

---

A Thesis

Presented to

the Faculty of the College of Arts, Humanities and Social Sciences

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Trevor Freed

March 2024

Dr. Mitch Ohriner

Author: Trevor Freed
Title: Harmonic Data: AI Music, EDM Transcription, & Minimalist Jazz
Advisor: Dr. Mitch Ohriner
Degree Date: March 2024

Abstract

This thesis presents research in the domains of music theory and music technology. The topics of the three chapters are as follows: (1) the technical and artistic limitations of artificial intelligence when generating music, (2) the transcription of electronic dance music (EDM) using a hybrid notation system, and (3) the effective convergence of jazz and minimalism in the music of Joe Hisaishi from Studio Ghibli. Each topic investigates a different aspect of harmony and the modern ways in which humans communicate in terms of harmony. The first chapter surveys the latest research projects in AI-generated music and proposes ideas for future applications of music generation using machine learning. The second chapter presents a hybrid notation system for the transcription of electronic dance music. The final chapter identifies progressive composition techniques embodied in the works of Japanese composer Joe Hisaishi, who borrows musical influence from around the world.

Acknowledgements

I would like to extend my deepest gratitude to all my advisors on this project, including Mitch Ohriner, Kristin Taavola, Sean Friar, Haluk Ogmen, Jack Sheinbaum, David Hanson, Eric Gunnison, Woody Colahan, Keith Waters, and Andrew Hannum. I would also like to thank all my friends and family for their continued support of my studies, composition, and research.

Table of Contents

List of Figures

Introduction

This thesis presents research in the domains of music theory and music technology. Across the three chapters, each topic investigates a different aspect of harmony and the modern ways in which humans communicate in terms of harmony.

The first chapter surveys the current climate of artificial intelligence with respect to music and visual art generation. My research responds to the contrast in public reception of audio and visual media generated by artificial intelligence. In the last five years, AI-generated visual art has stunned and excited audiences far more than AI-generated music has been able to. My research focuses on the technical and artistic limitations of the latest AI music algorithms and proposes ideas for future applications of the current technology.

The chapter identifies two fundamentally different approaches to AI music, both of which involve training neural networks on large data sets. The first kind of approach trains networks on MIDI data, where the network's output is a string of note values, including pitch, loudness, and duration. These values are given to any digital instrument synthesizer to be "performed". The second approach trains networks on datasets of audio waveforms, where the output is a song file that can be played with any loudspeaker. The music generated by MIDI networks is recognizable for its robotic quality, which sounds dispassionate and less expressive than human-made music. The waveform networks can create something closer to a professionally mixed record, but the technology has not yet advanced enough to convince listeners that no computers were involved in the production of the song. The chapter ends with my central argument: how to effectively apply these technologies toward creating more compelling music given their current limitations.

The second chapter explores the history of musical notation in order to most effectively transcribe compositions of electronic dance music (EDM). I first respond to the work of Charles Seeger, regarding the functional difference between *prescriptive* and *descriptive* musical notation. Next, I examine two case studies of transcription systems which Seeger presents in his essay: a graph from Milton Metfessel and a graph from Olav Gurvin. These graphs provide benchmarks for notating the most granular subdivisions of pitch, which is the primary goal of my research in this chapter. After discussing the strengths and shortcomings of each model, I introduce the 2010 global hit song "Levels" by Tim Bergling, also known as Avicii. Finally, I analyze a twelve-second passage of music to craft an original *descriptive* transcription with my new notation system.

The challenge of transcribing this passage involves mapping the gliding, ascending pitch of a prominent synthesizer in the music. By combining a modern piano roll notation with an audio spectrogram, my transcription is designed to simultaneously display the pitch of the gliding synthesizer along with other arpeggiated notes. The result is a musical graph which conveys the precise pitch of each synthesizer at any beat in musical time.

The final chapter presents formal analyses of two musical works by the Japanese composer Joe Hisaishi (b. 1950) from Studio Ghibli. Working closely with director Hayao Miyazaki, Hisaishi has composed the music for nearly every one of Studio Ghibli's animated films. His music borrows strongly from the Western classical tradition, but also from African-American jazz music and minimalist composition techniques.

The first piece I analyze is "One Summer's Day" from *Spirited Away* (2001) and the second is "Bygone Days" from *Porco Rosso* (1992). Both pieces present simple, modest melodies that become enriched by intimate, surreal atmospheres. Hisaishi uses extended and chromatically altered harmonies to support his melodies, often creating an idiomatic jazz quality. In terms of minimalist techniques, the composer effectively recycles and transposes short melodic fragments to create the fabric of his compositions. Finally, I argue that Hisaishi makes frequent use of a particular type of harmonic motion I call a retrogressive phrase overlap, which undermines the sense of tonal closure in his pieces and contributes to the feeling of time being suspended.

Overall, this thesis presents multiple perspectives on music production, harmony, and composition in the modern world. Audio examples of AI-generated music can be found as web links in footnotes and Chapter 1 References. Full transcriptions of the selected works by Joe Hisaishi can be found in Appendix A and Appendix B.

Chapter 1:

OK Computer, Heart & Soul:

Effective AI Music-Making

Introduction

In 1951, the first computer-generated melodies were recorded at the University of Manchester.[1] At the time, Alan Turing was conducting research at the school's Computing Machine Laboratory. He programmed a computer to generate audible tones that would indicate the machine's progress on some task. When Turing's colleague Christopher Strachey heard these tones in action, he was inspired to produce a melody using the computer's speaker. Soon thereafter, the computer convincingly hummed out "God Save The King," "Baa Baa Black Sheep," and Glen Miller's "In The Mood."

In the seven decades since, computer music has evolved in astounding ways. In the mid-to-late 20th century, intricate processes of algorithmic composition were of great interest to both computer scientists and musicians alike–compiling sets of rules that could be followed (by human or computer) to produce interesting kinds of music. In the early 80s, the composer David Cope developed a software called *Experiments in Musical Intelligence* (EMI), which could deconstruct scores of famous musical works, then

[1] Jack Copeland and Jason Long, "Restoring the First Recording of Computer Music," *British Library* online, accessed February 16, 2024, https://blogs.bl.uk/sound-and-vision/2016/09/restoring-the-first-recording-of-computer-music.html.

recombine the extracted material to develop original works in a similar style.[2] By the 1990s, a new era of computing was emerging, involving "deep learning" with convolutional neural networks (CNNs). These machine learning algorithms are trained on large datasets from which they extract properties, often called features, shared among entities in the set. In the visual domain, the task of *object-detection* employs this technology and serves as a foundation for more complex systems like self-driving cars. The same technology also underlies new applications in the automated generation of original content, as well as, perhaps, expressive creativity demonstrated by machines.

Over the last decade, ML- and AI-focused research into music generation continues to grow quickly. Major software companies, including Sony, Google and Spotify, have invested millions of dollars into automatic music generation projects. So why are these music projects not arousing more enthusiasm from the public? Where are all the computer-generated songs topping Billboard's Hot 100 chart? The development of new neural networks and deep learning architectures have led to ground-breaking applications in the realm of digital art creation. However, the visual art generated by these systems continues to be received far more positively by audiences than any auditory or musical art. Why is AI music falling behind the curve? Perhaps the answer to this question can be found in the nature of human expression. While scholars in academia relentlessly strive to understand expression and creativity, our machines are beginning to demonstrate creativity of their own. If we study AI-generated and human expression alongside each other, this could potentially reveal some powerful insights into our own human nature.

[2] David Cope," Experiments in Musical Intelligence," *UCSC Artsites,* accessed February 16, 2024, http://artsites.ucsc.edu/faculty/cope/experiments.

My research has been motivated by a desire to understand this current state of affairs– in just the last five years, the evolution of music generated by artificial intelligence has reached a latent plateau. Although machine learning technology and literature have been proliferating immensely, the public reception of even the most advanced AI music has been lukewarm at best. This chapter explores potential reasons for the stagnation in this field, including the "uncanny valley" of robotics, as well as present limitations, both technical and artistic, of AI music technology. The chapter's main purpose is to propose alternative applications for music generation using machine learning technology. These applications are (1) the integration of source separation into the training of large datasets, and (2) the human performance of AI-generated instrumental solos.

Machines Making Visual Media

In 1978, author and art critic Jasia Reichardt published a book titled *Robots: Fact, Fiction, and Prediction.*[3] This was the Western world's introduction to a new term in robotics: the "uncanny valley." The concept was first postulated by Japanese robotics professor Masahiro Mori in 1974, and Reichardt then translated his idea into an English phrase. Mori's research focused on individuals' reactions to stimuli that are perceived as human-like, yet not quite human. He theorized that once robots surpass a certain resemblance to human form, any affinity we have for them will become negative. The uncanny valley can be described as "the proposed relation between the human likeness of an entity, and the perceiver's affinity for it."[4]

---

[3] Jasia Reichardt, *Robots : Fact, Fiction, and Prediction* (New York: Viking Press, 1978).

[4] Masahiro Mori, Karl F. MacDorman, and Norri Kageki," The Uncanny Valley [From the Field]," *IEEE Robotics & Automation Magazine* 19, no. 2 (2012): 98–100.

Lately, the idea of the uncanny valley has been given new relevance regarding our emotional responses evoked by computer-generated media. Until recently, humans have been quite adept in discerning between media made by a human and media made by a computer. Alas, studies increasingly support the proposition that people are losing the ability to distinguish these. One such study was conducted in 2017 by a team from Facebook AI Research and the Art and AI Lab at Rutgers University.[5] The team built a new kind of neural network, called a Creative Adversarial Network (CAN), which generates original works of digital art. Participants were shown a series of images. Some images were made by the CAN and some images were human-generated and shown at Art Basel 2016, a modern art fair of international acclaim. When asked if the images were made by a human artist or by a computer, participants decided 53% of the CAN images were made by a human, compared to only 41% of the Art Basel images.

In 2018, a computer-generated portrait sold at a Christie's auction in New York for a winning bid of $432,500.[6] The portrait was generated by a Generative Adversarial Network (GAN), built by a French collective called Obvious. Interestingly, a signature was included in the bottom corner of the portrait, which was actually just the algorithm used to create the artwork. When one member of Obvious, Hugo Caselles-Dupré, was asked about the authorship of the piece, he remarked, "If the artist is the one that creates

---

[5] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone," CAN: Creative Adversarial Networks, Generating 'Art 'by Learning About Styles and Deviating from Style Norms," June 2017, https://doi.org/10.48550/arXiv.1706.07068.

[6]" Is Artificial Intelligence Set to Become Art's Next Medium?," Christies, December 12, 2018, https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx.

the image, then that would be the machine… If the artist is the one that holds the vision and wants to share the message, then that would be us."[7] This perspective highlights some nuances of human-computer interaction in the realm of creativity. I return to the idea of authorship as it relates to AI music in a later section of the chapter. Moreover, the GAN and CAN architectures mentioned are two powerful types of adversarial networks. In this chapter's section on Waveform Networks, I discuss how adversarial networks are currently the most effective models for AI-music generation.

In popular culture, the emergence of modern DeepFake videos highlights the implications of AI-generated visual media on a larger scale. The term "deepfakes" was coined in 2017 by a Reddit user of the same name.[8] Using deep learning technology, this user effectively doctored videos to swap the faces of celebrities onto other people's bodies. This trend has evolved to produce eerie social phenomena like DeepTomCruise, a TikTok account with dozens of videos starring an actor who has been given Tom Cruise's face.[9] User Bam4lam left the highest rated comment on one such video, saying: "These deep fakes are getting worryingly good. How the heck can we trust what we see in TV and camera is who we think."[10] The comment received over 65,000 likes, signaling that many people may be feeling the same way. This social experiment would also suggest

---

[7] Ibid.

[8] Yisroel Mirsky, and Wenke Lee," The Creation and Detection of Deepfakes: A Survey," *ACM Computing Surveys* 54, no. 1 (2021): 1–41, https://doi.org/10.1145/3425780.

[9] Tom (@deeptomcruise), 2021, "Sports!," TikTok, February 22, 2021, https://www.tiktok.com/@deeptomcruise/video/6932166297996233989.

[10] Ibid.

that neural networks are already producing realistic video footage that can convincingly pass as human-generated.

Online digital art platforms powered by machine learning are also producing astounding results, and many of them can be accessed on sites that are free to the public. The biggest players in this field today are Dall-E 2 by OpenAI, Stable Diffusion by Stability.ai, Imagen by Google, and a few others. These are all online platforms with powerful text-to-image functionality. The systems work by prompting users to input just a few words describing what they would like to see. Within seconds, the programs generate multiple original images to match their description. The outputs can resemble oil paintings, cartoons, or stunning photographs, depending on the user's text prompt. Some examples of text prompts on Google Imagen's website include "a small cactus wearing a straw hat and neon sunglasses in the Sahara Desert" or "a brain riding a rocket ship heading towards the moon."[11] The prompts can be totally abstract and surreal, and yet the outputs often give the impression that a professional artist might have produced them.

So far, this chapter has considered an AI portrait which sold for nearly half a million dollars, an AI social media account with over three million online followers, and several digital art platforms producing professional-quality images at the free click of a button. Now, the question becomes: where are all the newest songs written by AI, and is anybody listening to them?

---

[11] " Imagen," Google, accessed February 16, 2024, https://imagen.research.google/.

Machines Writing Music

Musical machine learning projects are making some light noise on the scene, but not on the scale of their visual art counterparts. A significant gap in the research remains to be filled, namely in addressing society's reception toward and evaluation of AI music. Music generated by artificial intelligence does not seem to captivate us yet in the ways that generated visual media does. The next section of the chapter provides some brief context for the latest developments in AI music.

In 2016, researchers based at Sony's Computer Science Laboratories published the first pop song written by artificial intelligence.[12] The song is called "Daddy's Car," a Beatles-inspired track composed by FlowMachine, Sony's musical machine-learning algorithm. However, the algorithm was responsible for writing the score only. French composers Benoît Carré and François Pachet wrote the song's lyrics, arranged each vocal and instrumental part, and engineered the audio mixing as well. As far as the track's social impact, it remains unavailable on streaming services like Apple Music or Spotify and has not quite reached three million plays on YouTube.

In 2021, a Toronto-based organization called Over the Bridge published their AI music project, *Lost Tapes of the 27 Club*.[13] This project featured new "original" songs in the styles of famous artists who died at the age of 27, including Kurt Cobain, Jimi

---

[12] Glen Tickle," Daddy's Car, A Song Composed by Artificial Intelligence Created to Sound Like The Beatles," *Laughing Squid*, September 22, 2016, https://laughingsquid.com/daddys-car-a-song-composed-by-artificial-intelligence-created-to-sound-like-the-beatles/.

[13] Kory Grow, "In Computero: Hear How AI Aoftware Wrote a 'New' Nirvana Song," *Rolling Stone*, April 5, 2021, https://www.rollingstone.com/music/music-features/nirvana-kurt-cobain-ai-song-1146444/.

Hendrix, Jim Morrison, and Amy Winehouse. One of the generated songs is called "Drowned in the Sun," written in the style of the grunge-rock band Nirvana.[14] The composition process was similar to that of Sony's track "Daddy's Car." The team used Google's AI program Magenta, which trained a neural network how to write guitar riffs and vocal melodies in MIDI, rather than generating audio waveforms. Sean O'Connor from Over The Bridge explains their process:

> …We took 20 to 30 songs from each of our artists as MIDI files and broke them down to just the hook, solo, vocal melody or rhythm guitar and put those through one at a time. If you put whole songs through, [the program] starts to get really confused on what [it's] supposed to sound like. But if you just have a bunch of riffs, it'll put out about five minutes of new AI-written riffs, 90 percent of which is really bad and unlistenable. So you start listening through and just finding little moments that are interesting.[15]

O'Connor's description highlights a number of interesting themes that are quite common in AI music. First, he states that the program trained on MIDI files, which are symbolic representations of musical information created for efficient computer processing. MIDI (Musical Instrument Digital Interface) data is usually limited to three values (pitch, duration, and loudness), but other values can be represented beyond those. David Cope, the algorithmic composer mentioned earlier, designed his software to analyze hundreds of Bach chorales in the form of MIDI representations, rather than audio recordings or visual scores. Using MIDI is quite common for a lot of electronic music producers and even some film composers. However, there tends to be an unnatural

---

[14] Ibid.

[15] Ibid.

quality to MIDI music, because so many elements of expressive nuance are lost in the reduction of acoustic information.

Another theme found in O'Connor's description is a technical limitation of the program when learning to write an entire song. O'Connor states that they had to train the algorithm on fragments of the song one at a time, because training it on whole songs made the program get "really confused" about what the song should sound like. Algorithms training on visual media do not seem to exhibit this same sort of confusion. Visual media models can train on entire images, even videos, which do not have to be deconstructed into constituent parts. In addition, O'Connor claims that even when they break the song down into parts, the algorithm still produces a significant quantity of undesirable material, which he describes as "really bad and unlistenable." This requires an additional step in the composition process, where a human must sift through heaps of less preferable material just to find music that is on some level compelling.

In fact, developers of musical neural networks have realized just how challenging it is to produce full-length songs–it has become far more common to see AI projects that merely *assist* humans in the composition process. In 2018, Google released their AI music platform Magenta Studio to the public, offering it as a "collection of music creativity tools."[16] Magenta Studio contains five audio applications, called plug-ins, which can produce musical content based on datasets that were provided to the different machine learning models. The plug-ins are called *Generate, Continue, Interpolate, Groove,* and *Drumify*. Even though these apps have been available to the public for

[16]" Magenta Studio," *Magenta*, February 12, 2019, https://magenta.tensorflow.org/studio-announce.

several years now, Google has done very little to advertise what they are capable of. This is most likely due to the following reasons: (1) the machine learning algorithms are all trained on MIDI data, (2) none of them can generate more than a few seconds of music, (3) the music that is generated requires additional context and is not properly engaging or compelling on its own.

BebopNet / MIDI Jazz

In 2020, a team at the Israel Institute of Technology published a project called BebopNet, which trained a neural network to generate "improvised" jazz solos based on a large MIDI dataset.[17] The algorithm trained on hundreds of bebop solo transcriptions, and learned how to predict the next note of a musical "sentence" based on factors like tempo, chord progression, texture, density, etc. The training of the model also involved iterative input from a user, where the user is asked to rate generated solo samples based on preference. Thus, each new sample is, in theory, generated according to the user's taste.

However, the results exhibit a patent scarcity of expressive qualities that both musicians and non-musicians are very sensitive toward. These include changes in dynamics and tone color, varying rates of vibrato, extended instrumental techniques, and motivic development to name a few. What we hear instead is a stream of mechanically played notes without any discernible narrative or intention behind them.

---

[17] Shunit Haviv Hakimi, "BebopNet: Neural Models for Jazz Improvisations," *Towards Data Science*, April 29, 2020, https://towardsdatascience.com/bebopnet-neural-models-for-jazz-improvisations-4a4d723d0b60.

Here's one example of improvised material to the tune of "Just Friends" by Charlie Parker. (https://on.soundcloud.com/8Gsep)[18] The program begins with the head, or main thematic material, then follows that with generated improvisation. The first improvised notes can be heard at [0:40] in the track. Granted, there are certainly some aspects to the music that work well for the bebop style. For one, the melodic content all fits within the appropriate key center and moves through viable jazz modes as the accompaniment plays through the harmonic changes. There is enough melodic and rhythmic variation where the music does not sound overly repetitive. However, the playing does sound automatic, unfeeling and uninspired, like a warmup drill or a Hanon exercise.

Now I ask the reader to recall Hugo Caselles-Dupré's comments on authorship and artistry. He suggested a perspective that discerns between who or what holds the artistic vision and message, and who or what produces the result of those intentions. In this case, the message does not seem to be an emotional one, but perhaps a purely logical one. Therefore, the spirit behind the music lacks a depth of character or nuance, which some may consider necessary in order for music to be expressive. Furthermore, consider the repeating motif heard at [1:11] in the track.[19] This 4-note riff plays ten times in a row with only a few slight variations. As soon as this sequence begins, the illusion that a human might be playing is completely lost. The program sounds like it has become stuck on a loop, and the rigid quantization of the rhythm sounds dispassionately robotic.

---

[18] Shunit Haviv, "Just Friends - Jazz Improv AI," SoundCloud audio, 1:26, 2020, https://soundcloud.com/shunit-haviv/just-friends-jazz-improv-ai.

[19] Ibid.

Waveform Networks

Thus far, the musical neural networks discussed in this chapter have been Recurrent Neural Networks (RNNs), which train on character-based data in the form of MIDI. RNNs are designed to predict the next element that is likely to appear in a given sequence, such as a written sentence or a musical melody. A distinctly separate class of networks, Generative Adversarial Networks (GANs), are designed to train on multi-dimensional datasets, such as digital images or audio waveforms. The musical results produced by this class of networks are compelling in their own ways, although the uncanny valley remains glaringly wide and eerie.

In 2020, OpenAI published a project called Jukebox, which trained a neural network to generate original songs in a number of different styles. Their website describes the program as such: "Provided with genre, artist, and lyrics as input, Jukebox outputs a new music sample produced from scratch."[20] For example, one generated song they shared online is called "Rock, in the style of Elvis Presley." (https://on.soundcloud.com/V1ms)[21] First, the network trained on a large dataset of rock songs, which included songs by Elvis Presley. The algorithm trained on music and lyrics separately, but then learned to also match the generated lyrics with the generated music. The result is a new original track with a unique formal structure, guitar riffs, vocal melodies, and lyrics sung in the style of Elvis Presley.

---

[20]" Jukebox," *OpenAI*, accessed February 16, 2024, https://openai.com/blog/jukebox/.

[21] OpenAI, "Rock, in the style of Elvis Presley - Jukebox," SoundCloud audio, 1:11, 2020, https://soundcloud.com/openai_audio/rock-in-the-style-of-elvis-4.

There are some peculiar qualities to the music that stand out as soon as the track begins. Perhaps the first and most apparent element to be heard is the grainy white noise that shifts around the audio mix. The noise is actually not a result of older, lower-quality recordings as one might expect–it is found in all of Jukebox's samples from each genre. In fact, the white noise is a result of randomization in the architecture of the GAN. The algorithm is *guessing* how to fill in the cracks of the audio waveform, and we are left with punctuated moments of disorganized data that our ears perceive as static-like noise. This is another technical limitation of the current technology, although one that has a clear path towards its resolution. When the training datasets of songs become large enough, the next generation of algorithms will be *guessing* with more informed accuracy, producing less white noise as a result. Finally, the vocals in the song are consistently warped, muffled, and lacking clear articulations. At [0:28] in the track, we hear what is functionally the final stanza of the first verse, and yet most words are either unclear or unintelligible.[22] Although the artificial voice sounds like a fairly convincing Elvis impression, the unrefined scrambling of syllables entirely dispels the illusion that a human could actually be singing.

Future Research / Alternative Applications

Based on my findings in this field, I return to my main argument: the most promising application of machine learning toward music-making will involve the interaction of two independently robust technologies. The first is, namely, the training of musical neural networks as discussed in this chapter thus far, training the networks on large datasets of

[22] Ibid.

songs in specific styles in order to generate original material. The second is another type of machine learning algorithm with a distinctly different objective, which is commonly called *source separation*. Source separation refers to the isolation and extraction of individual instrument parts from a song's overall mix. In 2019, the French online music streaming company Deezer released a source separation software called Spleeter.[23] The software features a machine learning algorithm that is trained on a dataset of audio from isolated instrument sources. In this way, when it is asked to analyze a song with multiple instruments, it effectively separates the song into audio stems of individual instruments.

My first claim is such: the quality of AI music will see dramatic improvement, both technically and artistically, when source separation is incorporated into the pre-processing stage of training musical neural networks. When each instrumental track is treated individually by the machine learning algorithm, the generated outputs will be more refined, sounding cleaner and more authentically similar to the original sources they are emulating. For example, there can be one network training on isolated drum stems, another training on guitar, and another on vocals. When each network produces new original content on its own, the constituent tracks can eventually be mixed together to create the final waveform, much like the process of mixing and mastering in a recording studio. In the case of the Elvis-style song by OpenAI's *Jukebox*, each instrument would likely become clearer in the song's mix, no longer warping or phasing in and out. The form and structure of generated songs will also become more defined, as the networks

[23] "Spleeter," *Deezer,* accessed February 16, 2024, https://research.deezer.com/projects/spleeter.

will learn more acutely the moments and sections of a song when an instrument is either sounding or remaining silent.

My second claim is that generative MIDI platforms like BebopNet will become far more likely to captivate their audiences under one condition–they leave the musical performance to humans. The solos generated by BebopNet are complex and interesting on an abstract level, before they have been interpreted by any digital instrument. The lack of expressive timing and articulations in the computer's performance, however, leave the music sounding robotic and lifeless. Until now, every instrumental solo in history has been written by a human and likely played by a human. Perhaps, not too far in the future, we will hear virtuosic human interpretations of full-length art songs written by machines.

Conclusion

This chapter has discussed the technical and artistic limitations that AI music technology currently faces. For audience reception of AI-generated music to match the level exhibited by that of AI-generated visual media, these limitations must be overcome. Machine learning algorithms will inevitably improve over time, but we can also explore alternative applications that could yield better, more musical results in the meantime.

For the generation of fully mixed songs, I argue that the integration of source separation technology could likely lead to higher-quality results. I have yet to see any research projects where these technologies have been combined, so I have started working toward the implementation of this development for my latest research project. For MIDI solo generation, we need to integrate a wider array of expressive qualities into the neural network training models. On the other hand, audiences might ultimately prefer to hear humans perform complex instrumental solos written by computers.

As machines learn to become more creative, it seems we can also learn about our own artistic nature and our capacities for creativity. I hope this chapter has identified the prevailing technical and artistic limitations of AI music generation that remain. I also hope to have inspired the reader to consider alternative applications of these technologies for the near future. There are exciting advancements for this field on the horizon, but the how and when remain up to ambitious engineers. Until then listeners await, either eagerly or fearfully, the day when they realize their new favorite song was written by a machine.

References

Cope, David. "Experiments in Musical Intelligence." UCSC Artsites. Accessed February 16, 2024. http://artsites.ucsc.edu/faculty/cope/experiments.

Copeland, Jack, and Jason Long. "Restoring the First Recording of Computer Music." British Library. Accessed February 16, 2024. https://blogs.bl.uk/sound-and-vision/2016/09/restoring-the-first-recording-of-computer-music.html.

Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. "CAN: Creative Adversarial Networks, Generating 'Art 'by Learning About Styles and Deviating from Style Norms." Paper presented at International Conference on Computational Creativity (ICCC), Atlanta, GA, June 2017. https://doi.org/10.48550/arXiv.1706.07068

Grow, Kory. "In Computero: Hear How AI Aoftware Wrote a 'New' Nirvana Song." Rolling Stone, April 5, 2021. https://www.rollingstone.com/music/music-features/nirvana-kurt-cobain-ai-song-1146444/.

Hakimi, Shunit Haviv. "BebopNet: Neural Models for Jazz Improvisations." Towards Data Science, April 29, 2020. https://towardsdatascience.com/bebopnet-neural-models-for-jazz-improvisations-4a4d723d0b60.

Haviv, Shunit. "Just Friends - Jazz Improv AI." SoundCloud audio, 1:26, 2020. https://soundcloud.com/shunit-haviv/just-friends-jazz-improv-ai.

"Imagen." Google. Accessed February 16, 2024. https://imagen.research.google/.

"Is Artificial Intelligence Set to Become Art's Next Medium?" Christies, December 12, 2018. https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx.

"Jukebox." OpenAI. Accessed February 16, 2024. https://openai.com/blog/jukebox/.

"Magenta Studio." Magenta, February 12, 2019. https://magenta.tensorflow.org/studio-announce.

Mirsky, Yisroel, and Wenke Lee. "The Creation and Detection of Deepfakes: A Survey." *ACM Computing Surveys* 54, no. 1 (2021): 1–41. https://doi.org/10.1145/3425780

Mori, Mashiro, Karl F. MacDorman, and Norri Kageki. "The Uncanny Valley [From the Field]." IEEE Robotics & Automation Magazine 19, no. 2 (2012): 98–100.

OpenAI, “Rock, in the style of Elvis Presley - Jukebox.” SoundCloud audio, 1:11, 2020. https://soundcloud.com/openai_audio/rock-in-the-style-of-elvis-4.

Reichardt, Jasia. *Robots : Fact, Fiction, and Prediction*. New York: Viking Press, 1978.

“Spleeter.” Deezer. Accessed February 16, 2024. https://research.deezer.com/projects/spleeter.

Tickle, Glen. “Daddy's Car, A Song Composed by Artificial Intelligence Created to Sound Like The Beatles.” Laughing Squid, September 22, 2016. https://laughingsquid.com/daddys-car-a-song-composed-by-artificial-intelligence-created-to-sound-like-the-beatles/.

Tom (@deeptomcruise). 2021. “Sports!.” TikTok, February 22, 2021. https://www.tiktok.com/@deeptomcruise/video/6932166297996233989.

Chapter 2:

Read Between the Waves:

A Hybrid Notation System for Electronic Dance Music

Introduction

Over the last twenty years, electronic dance music (EDM) has emerged as one of the most popular genres of music in the world. However, the strong influence of this genre has coincided with a major shift into a new era of musical composition–an era where many traditional practices of Western music notation are being left behind. For hundreds of years, much of the creation, performance and formal analysis of Western tonal music has relied upon written notation and sheet music. In recent decades, however, many new musical genres have been born in the United States, especially after the Vietnam War, the counterculture movement, and the Digital Revolution. New genres are inevitably characterized by new styles, techniques, and even artistic materials. In this vein, composers of electronic music, often called producers, have become masters of nontraditional instruments, such as laptops, synthesizers, drum pads, and sound mixers. As a result, EDM composers are often depicted to be writing music in front of a computer screen, twisting knobs, sliding faders, and dragging MIDI notes around an elaborate digital audio workspace (DAW).

At electronic music concerts today, crowds in the tens of thousands gather to see a single music producer on stage with the same laptop they used to compose their most recent hit song. That song may never have involved a written note on a page, or even a live instrument. In his book "Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music," professor of music theory and cognition Dr. Mark Butler acknowledges that electronic music is hardly ever composed on paper or via written notation.[24] In most cases EDM producers do have something *like* a score, which is the computer file they work with in their DAW. These look nothing like traditional sheet music and are more akin to grids where notes can be stretched or dragged around. Later in this chapter, my proposed notation system utilizes this kind of grid framework for its equal spacing of chromatic pitches. Furthermore, even after an EDM track has been published, a full transcription of the music is rarely given life. This is hardly the case for other genres of pop music. For many popular genres like jazz, rhythm & blues, musical theater and rock, major music publishers continue to hire transcribers and sell books of sheet music. These books are purchased not only by musicians for recreation and performance, but also by music scholars for formal analysis. Dr. Butler argues that, even for electronic music, written notation "facilitates the close engagement with musical sound that is an essential part of music-theoretical research."[25] In order to conduct the most thorough analysis of a piece of music, a music theorist must rely upon more than just their ears. However, due to the largely digital nature of EDM, there are some significant challenges that arise when attempting to craft an accurate transcription in the style of EDM.

---

[24] Mark Jonathan Butler, *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music*, Bloomington: Indiana University Press, 21, 2006.

[25] Ibid., 22.

This chapter explores the various barriers that currently exist regarding the transcription of electronic dance music. First, I provide a background of written musical notation in common practice and discuss the functionality of traditional symbols. Then I examine historical examples of progressive transcription systems and how they were conceived. Next, I propose how certain elements from these systems could be borrowed for the synthesis of a new system. My argument is framed around a case study of the famous EDM track *Levels* by Tim Berling, also known as Avicii. In doing so, I argue for the necessary components of a standardized notation system that could be used to effectively analyze future catalogues of EDM.

Notation

Every culture has seen the creation of music *without* any sort of notation. Written sheet music, though, has often been an essential resource for both performers and scholars of Western music. Typically, its role is to communicate the most fundamental elements of a musical work to the reader. However, one might encounter a written score in a variety of contexts. In the early stages of working on a piece, a composer might use sheet music to sketch ideas and arrange them in a particular order. In preparing the performance of a piece, musicians might be given sheet music as a means of instruction for how their part should sound. After a work is performed or published, perhaps there are some who would like to craft their own transcriptions without the manuscript of the composer. These are all efforts to engage one's visual sense with music outside of time. In the creation of any transcription there lies a balance between presenting the most relevant information for the reader and navigating available space on the page.

There is perhaps one distinction, however, that unifies the lot of these cases.[26] Namely, this difference is in whether the sheet music is being utilized before or after the realization of a musical work. Composers, performers, and arrangers might consult a written score before a work is realized, while music theorists, musicologists, and other scholars might consult the sheet music after the work has been realized. In a 1958 essay titled "Prescriptive and Descriptive Music-Writing" in *The Musical Quarterly*, the famous musicologist Charles Seeger argued that traditional practices have largely failed to distinguish between these two primary functions of notation. The key difference, he claimed, is "between a blue-print of how a specific piece of music shall be made to sound and a report of how a specific performance of it actually did sound."[27] Seeger proposed the terms *prescriptive* and *descriptive* to differentiate between these two paradigms. For EDM, there are usually no performers to whom this blue-print could be given, except to the computers, since the performances and recordings are fully realized by synthesizers and other computer algorithms. Therefore, this chapter focuses on *descriptive* notations which aim to visually report musical events that have already happened.

In the world of music, any musical tone can be said to exhibit four main attributes– (1) pitch, (2) duration, (3) loudness and (4) tone quality (also called timbre). Let's consider a typical classroom melodic dictation, where students are asked to transcribe the notes they hear while listening to a passage of music. The instructor will either sing or play an instrument, and the task of the student is to craft a visual report of the pitches

---

[26] Charles Seeger, "Prescriptive and Descriptive Music-Writing," *The Musical Quarterly* 44, no. 2 (April 1958): 184, https://www.jstor.org/stable/740450.

[27] Ibid.

heard, including the durations of all notes and rests. It is worth stating here that any consideration of tone quality or timbre is usually omitted from this report–the primary goal is to reflect the (1) pitches and their (2) durations, and the task of notating timbre in symbols without words is a recognized challenge. Most commonly, students complete the activity by writing out notes on a musical stave. The flagging, stemming, and filling-in of any note head determines that note's duration, relatively speaking. The relative height of that note head on the staff will determine the note's musical pitch name. With sharps and flats, this Western system lets us agree upon twelve chromatic tones within one octave, sometimes referring to half and quarter tones found between those twelve.

From here, additional symbols can be incorporated to provide a more detailed report of the musical passage. Dynamic markings, including crescendos and decrescendos, can be used to convey how the volume should change between notes and through passages. Accents are available to indicate forceful, more intense notes, while staccato and legato marks (slurs) can suggest whether notes are disjointed or smoothly connected. With a pencil and some sheet music paper, a trained musician can learn how to read and how to craft a descriptive transcription using these standardized notation systems.

What happens when more complex musical events are introduced? For example, how can one use standard symbols to report the precise rate of a note's excited vibrato, or a dramatic glide between two pitches? Traditionally, vibrato is not notated in Western music. However, some modern composers will notate vibrato by using a jagged line and appending that line to a note head, though these symbols are usually unspecific in terms of vibrato rate and range.[28] Moreover, a glide between pitches might be shown using a glissando symbol, which can be either straight or jagged and connects a starting note to an arrival note.[29] However, the glissando symbol is a loss of musical information, since the reader is left to assume the rate at which the pitch glides over the glissando's range. Later on, this chapter discusses how computers and spectrogram technology can aid in the recovery of musical information that has been lost in the notation process.

Metfessel's Graph

Therefore, if standardized symbols and traditional notation practices will not suffice to produce high-fidelity transcriptions, a new system must be made to complete the task. In his essay "Prescriptive and Descriptive Music-Writing," Seeger presents a fellow researcher's attempt at such an original system, conceived in 1928 by Milton Metfessel.[30] Figure 2.1 below is a hand graph that reports pitch information from a vocal performance of *Swing Low Sweet Chariot,* an African-American spiritual folk tune.[31] The vertical lines that spread across the *x*-axis represent seconds, and the *y*-axis shows musical pitch names. For each note in the melody, a note head is positioned at the approximate moment of the note's onset or vocalization.

---

[28] Elaine Gould. *Behind Bars: The Definitive Guide to Music Notation*. Faber Music Ltd, 146, 2016.

[29] Ibid, 140.

[30] Milton Metfessel and Carl E. Seashore. *Phonophotography in Folk Music: American Negro Songs in New Notation* (University of North Carolina Press, 1928).

[31] Ibid.

This archetype of graphical representation already provides much more musical information than a traditional melodic dictation on staff paper. Upon close inspection, a few significant features stand out.

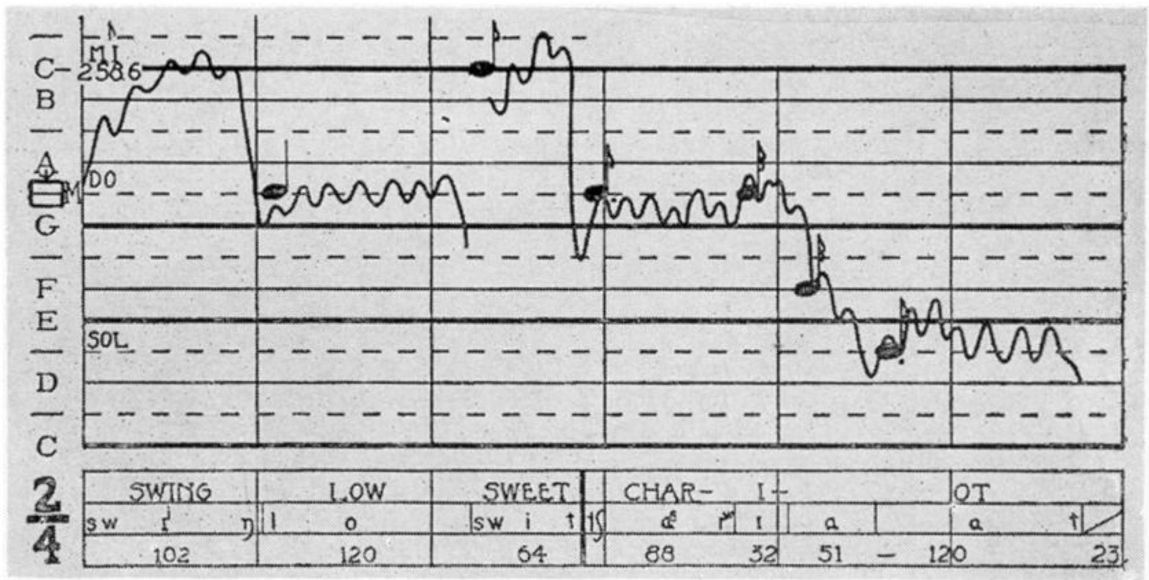

Figure 2.1 Hand Graph, made by mathematical reduction of a "sound wave photograph."[32]

First, the reader can see the contour and shaping of individual notes. Rather than seeing a generic vibrato symbol extending out beyond a note head, one sees more detailed information about the timing and range of each vibrato. One can also see, with the highest C note and several other cases, that certain notes are approached by a slight glide upward into the diatonic pitch. Although this technique is quite common in vocal performance across many genres, the symbols that are available to notate such an inflection are both limited and largely unspecific.

[32] Ibid.

Gurvin's Graph

While Metfessel's graph above demonstrates a notable improvement in accuracy, even more information can be gathered with the aid of additional technology. Figure 2.2 below shows another type of graph, made in 1953 by Norwegian musicologist Olav Gurvin, consisting of three separate systems.[33] Directly above the musical staff, the middle system resembles Metfessel's in that they both report pitch over time.

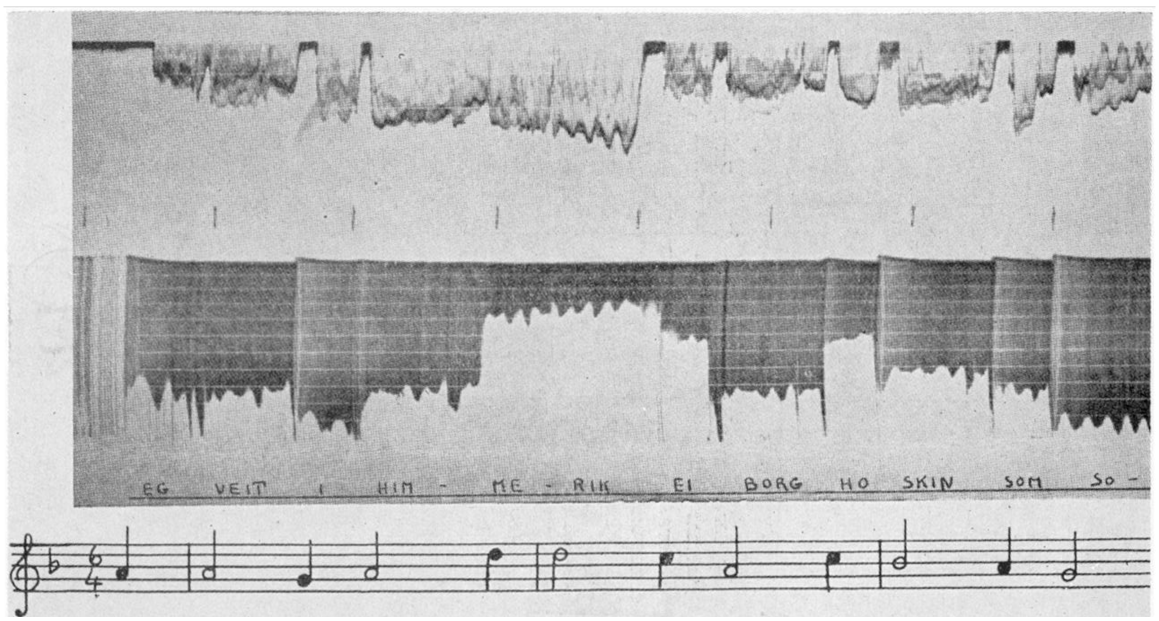

Figure 2.2 Automatic Graph (oscillogram) made by electronic-mechanical reduction, photographed on film, of Norwegian folksong, sung by woman's voice.[34]

In Gurvin's method, the vertical lines above the system also represent seconds in a similar manner. Finally, the system at the top of the graph was produced by an oscilloscope and reports amplitude, or loudness, over time. If a reader saw one of these

[33] Olav Gurvin, "Photography as an Aid in Folk Music Research," *Norveg* 3 (1953): 181-96.

[34] Ibid.

systems without the others, they may very well experience some difficulty in extracting meaning from such unfamiliar representation. However, the three systems taken altogether report a wealth of information more cohesively.

At this point, the reader of this chapter should hopefully have adequate context for the various considerations and possible approaches to high-fidelity transcriptions. The next section of the chapter discusses additional challenges that arise for the genre of EDM.

Electronic Dance Music: A New System

In the musical examples considered thus far, the instrumentation for each performance has been limited to a single human voice. Imagine, for a moment, what a graph from Metfessel or Gurvin might look like for a passage where multiple instruments are sounding simultaneously. There would be far too much visual chaos on the page to comprehend any meaning. Herein lies the challenge for electronic dance music: how does one convert the most relevant sonic information from an EDM track into a legible report, where the precise pitch of each instrument is accurately mapped?

First, we must consider: what are the essential attributes of EDM as a genre? In the *Journal of Popular Music Studies* 2001 issue, Kembrew McLeod from the University of Iowa wrote a chapter exploring the multitude of subgenres within EDM. Mcleod writes:

> The synthesizer is the instrument most associated with the genre, and the music has been characterized (often negatively) by its unrelenting, repetitive beat, which often was created with drum machines, synthesizers, and other electronic instruments. [35]

[35] Kembrew Mcleod, "Genres, Subgenres, Sub-subgenres and More: Musical and Social Differentiation Within Electronic/Dance Music Communities," *Journal of Popular Music Studies* 13.1, 59-75, 2001.

As Mcleod suggests, the repetitive beat of EDM, usually created by drum machines, is one of the defining attributes of the genre. In addition, synthesizers are used in almost every EDM track–these can be modular synths, keyboard synths or any other electronic instrument. When these instruments generate an electrical signal, it is very common for producers to apply modulation effects to the signal, including reverb, delay, distortion, or a bandpass filter, which each change the timbre of the instrument significantly. As a result, the task of notating timbre is widely recognized as a difficult challenge. Dr. Butler draws attention to this challenge in his book, where he writes the following footnote:

> This deficiency with respect to timbre is hardly unique to my transcriptions, however; timbre is underrepresented in many different notational systems. It would certainly be very useful to have a system of transcription capable of representing the timbral nuances of EDM, but that difficult task must be left for a future work.[36]

The first major consideration for an EDM transcription is deciding whether each instrument should be treated in the same manner. In a traditional score written for orchestra, all parts are treated with equal significance and are presented as such on the page. Each performer relies upon the same traditional modes of musical information to realize their individual part. In the case of an EDM track, however, synthesized parts are often created by using algorithms which, while complex, may offer little relevant information to the reader for the rest of the song.

For example, an EDM song could feature one synthesizer that glides between pitches and follows a specific algorithm that determines the rate of vibrato at any moment–this information is surely significant and ought to be included in the report for this instrument.

[36] Mark Jonathan Butler, *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music*, 22.

However, if another part in the song is played by digital bells that remain static in pitch and vibrato, then the inclusion of analogous information for this part becomes a waste of space on the page. This sort of hierarchical attitude is perhaps embodied by traditional notation for an orchestral percussion section–when musical pitch is deemed less relevant, relative height of note heads on the staff is understood to indicate a different parameter, such as specific instrumentation. As such, in order to preserve visual space and report only the most relevant information, an effective system should treat individual parts respectively, according to their fundamental parameters of operation.

Here, I introduce the original research of this chapter: a case study on the famous EDM track *Levels*, produced by the late Tim Bergling, known by the stage name Avicii.[37] This track was chosen for several reasons, chief among which was its instrumentation– the song features a synthesizer that plays a rising, gliding pitch over several measures, which is nearly impossible to transcribe with traditional notation. Moreover, the track includes multiple other synthesizers that operate on distinctly separate temporal functions. This case study combines a number of approaches explored in the chapter thus far in order to establish an effective framework for formal analysis.

Original Graph

The main goal for this analysis was to create a visual report of three instrumental parts sounding simultaneously in the track. The song begins with an intricate background texture, played by a single arpeggiating synthesizer. This texture creates a polyphony of

[37] Tim Bergling [Avicii], "Levels – Radio Edit," track 1 on *Levels,* Universal Music Group, 2011, Spotify streaming audio.
https://open.spotify.com/track/5UqCQaDshqbIk3pkhy4Pjg?si=1e0ac9b9fbd248c5

several interwoven melodic lines, and the pattern remains unchanged for the entire first minute of the song.

In order to represent this first instrument, I elected to use a graphical system that resembles the *pitch vs. time* paradigm used by both Metfessel and Gurvin. The music software Logic Pro was used to plot the notes in a typical “piano roll" format, commonly seen in digital audio workspaces.[38] A piano keyboard is shown vertically on the left side of the graph, and the duration of each note is indicated by the width of each plotted note. The second instrument represented is the lead synthesizer that plays the main hook of the chorus, heard at [0:07] and throughout.[39] Within the same system, this part is shown by notes colored red. In such a format, any reader can identify pitches played by more than one instrument quickly and easily. Moreover, several instruments can be represented by distinguishable colors, so long as the visual space does not become cluttered.

Finally, the third instrument represented is the featured synthesizer that performs an extended, gliding gesture toward the end of the song while the other two instruments are also sounding. This event begins at [2:31] in the track.[40] Beginning on B2, the pitch begins gliding upward at an exponential rate, rather than linearly, and reaches A5 at its peak. The challenge of my research was to provide as much pitch information as possible about each of these three synthesizers at any moment, especially this gliding synthesizer.

---

[38] *Logic Pro,* V. 10.7.2 (Apple, 2021), macOS 11 or later.

[39] Bergling.

[40] Ibid.

First, the musical software VoceVista Video was used to create an audio spectrogram of the entire song.[41] A spectrogram effectively shows a heat map of frequencies that can be abstracted from a given waveform. In the same style as the piano roll system, the *x*-axis of this system represents time and the *y*-axis represents pitch, or frequency. The third dimension represented is amplitude, where lower activations of a frequency are shown by the presence of blue/yellow and higher activations are shown by the color red (depending on program settings). In this way, the reader can see a detailed moment-to-moment report of exactly how the synthesizer's pitch glides from the low B2 to the highest A5.

Seen on their own, neither the piano roll nor the spectrogram provides the reader with enough relevant information. The spectrogram can often show lots of informational noise that hardly improves the reader's understanding of the musical passage, and also fails to clearly show the melodic content of the first two instruments. On the other hand, the piano roll can represent the static pitches of the first two instruments, but its limited functionality is unable to represent the continuous gliding of the third instrument. In order to create the most concise report of all relevant information, I overlaid one system onto the other and resized both systems to an orthogonal scale. More specifically, the ranges of both *x*-axes were made equal, as well as the ranges of both *y*-axes. Figure 2.3 below shows an excerpt of *Levels* from [2:31] ~ [2:41] represented in this notational system.

[41] *VoceVista Video,* V. 5.4.2.5435 (Sygyt Software, 2017), macOS 10.13 or later.

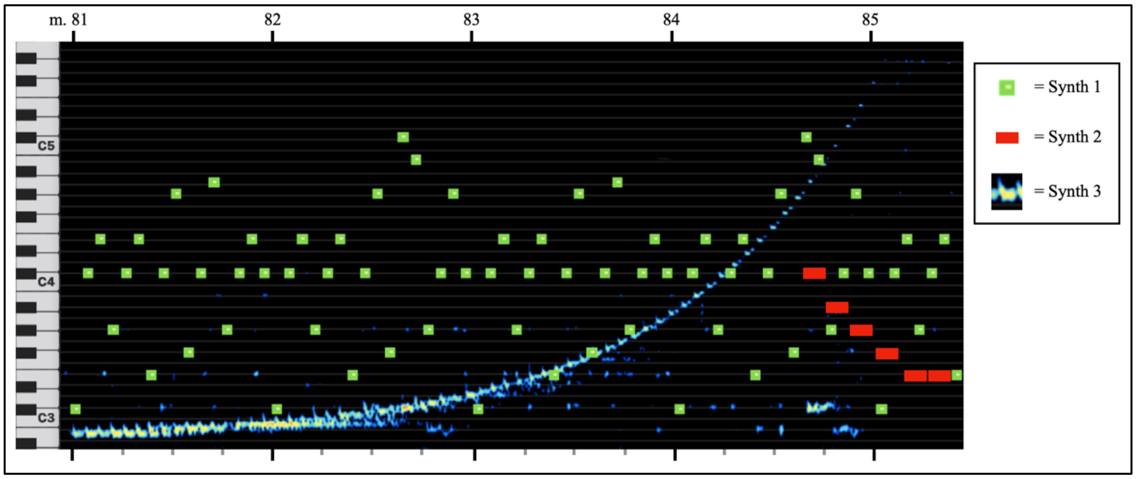

Figure 2.3 Original Hybrid Notation of Avicii's *Levels* excerpt; mm.81~85; [2:31] ~ [2:41]; time on *x*-axis, pitch on *y*-axis.[42]

Analysis

For this transcription to be as clear and insightful as possible, I needed to adjust the spectrogram's settings to minimize the appearance of irrelevant sonic information. The most important information pertains to the pitch of the gliding synthesizer (Synth 3), which is represented in Figure 2.3 by a slowly rising, jagged slope in a blue/yellow color. As such, there is some "noise" in the graph from the spectrogram, but this is simply other sonic information captured from the song that is not relevant to this discussion.

The onset of the note B2 by the gliding synthesizer (Synth 3) is indicated by the start of the blue/yellow slope in the bottom-left corner of the graph. This event coincides with the downbeat of measure 81, which is reflected by a vertical alignment with bass note C$^{\#}$3 (Synth 1) in the piano roll system. The climax of the song arrives at the start of measure 85, with an enormous "drop" where the instrumentation is at its fullest. On the

[42] This graph was created using the spectrogram view in VoceVista Video, the piano roll editor in Logic Pro, and Canva Online Photo Editor to overlay the images.

downbeat of this measure, the gliding synthesizer reaches its peak note, A5, and aligns with the note $F^{\#}3$ played by the second instrument (Synth 2). By inspecting the graph, one can identify several other moments where the gliding pitch of Synth 3 reaches momentary unison with the arpeggiating synthesizer (Synth 1). For example, on beat 2 of measure 84, both Synth 1 and Synth 3 play in unison on C#4, the most prominent pitch played by Synth 1. Without this notation system, the task of determining the exact moment when this unison occurs would be far more difficult. If tasked to do so, one could readily identify the precise sounding pitch of the third instrument at any beat of any measure, thus satisfying the proposed objective.

Conclusion / Further Research

This chapter has discussed the inherent challenges that must be faced to produce high-fidelity transcriptions of EDM tracks. Traditional practices of notation can only provide so much information, confined by the standardized systems that have pervaded Western tonal music for centuries. In order to report more detailed information, music analysts and musicologists have made comprehensive attempts to devise new visual systems for readers. By combining two or more of these systems, a new sort of framework can be conceived that presents significantly more musical information in a clear and concise manner.

With further research, this framework could be improved with visual representations of even more musical parameters, including electronic modulations like filter sweeps, delay, or stereo panning. Even with such improvements, the challenge remains of balancing information with available space on the page. Any reader should be able to refer to the graph and extract meaning on multiple levels at once, without being confused by chaotic, obscured, or absent information. Audio sampling is another essential aspect of EDM that this system can be used to analyze. Whether a sample contains vocals, sound effects, or other environmental sounds, this system can present sonic information about how the pitches in the audio sample relate to the rest of the track's instrumentation. Moreover, the original graph from my case study remains an example of *descriptive* notation, rather than *prescriptive*. The graph is intended to be used as an analytical tool, and I am not suggesting that *Levels* could be performed simply by following the graph.

Electronic dance music is still growing and evolving–slowly but surely, we will find new ways to analyze this intricate and complex genre of music. The standardized systems of traditional notation have evolved over centuries of composition and collaboration, only to be rendered largely insufficient by modern genres. Alas, the future of EDM analysis may indeed experience a similar fate as we read ever deeper between the waves.

References

Apple. *Logic Pro.* V. 10.7.2. Apple, 2021. macOS 11 or later.

Bergling, Tim [Avicii]. "Levels – Radio Edit," track 1 on *Levels,* Universal Music Group, 2011, Spotify streaming audio. https://open.spotify.com/track/5UqCQaDshqbIk3pkhy4Pjg?si=1e0ac9b9fbd248c5

Butler, Mark Jonathan. *Unlocking The Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music*. Bloomington: Indiana University Press, 2006.

Canva. *Photo Editor*. Canva, 2024. macOS 11 or later.

Gould, Elaine. *Behind Bars: The Definitive Guide to Music Notation*. Faber Music Ltd, 2016.

Gurvin, Olav. "Photography as an Aid in Folk Music Research." *Norveg* 3 (1953): 181-96.

McLeod, Kembrew. "Genres, Subgenres, Sub-subgenres and More: Musical and Social Differentiation Within Electronic/Dance Music Communities." *Journal of Popular Music Studies* 13.1, 59-75, 2001.

Metfessel, Milton, and Carl E. Seashore. *Phonophotography in Folk Music: American Negro Songs in New Notation*. University of North Carolina Press, 1928.

Seeger, Charles. "Prescriptive and Descriptive Music-Writing." *The Musical Quarterly* 44, no. 2 (April 1958): 184-95, https://www.jstor.org/stable/740450.

Sygyt Software. *VoceVista Video*. V. 5.4.2.5435. Sygyt Software, 2017. macOS 10.13 or later.

Chapter 3:

The Sound of *Spirited Away*:

Jazz and Minimalism in the Music of Joe Hisaishi

Introduction

In 2001, the Japanese animation team Studio Ghibli released a film that Time Magazine called “one of the most celebrated animated movies of all time.”[43] The film was *Spirited Away*, the story of a young girl who discovers an abandoned theme park inhabited by gods and spirits. The soundtrack for this magical story was composed by the legendary Joe Hisaishi (b. 1950), who has composed music for nearly every one of Studio Ghibli’s animated films. The music of Joe Hisaishi occupies a unique domain in the realm of contemporary music. In terms of style, Hisaishi’s music draws its influence primarily from the Western classical tradition, but also borrows strongly from the African-American jazz tradition. In recent years, Hisaishi has vocally identified with the culture of modern minimalist composers, too, including Philip Glass and David Lang.[44]

Jazz and minimalist music have inherently different artistic motivations and were born decades apart from each other, giving each their own cultures and histories. However, Hisaishi’s music heralds a convergence of these elegant, modern genres.

---

[43] Kat Moon, “How *Spirited Away* Changed Animation Forever,” *TIME*, July 20, 2021, https://time.com/6081937/spirited-away-changed-animation-studio-ghibli/.

[44] Joe Hisaishi Official, “Joe Hisaishi presents MUSIC FUTURE Vol.5 - Joe Hisaishi Interview,” YouTube video, 3:31, October 10, 2018, https://youtu.be/MJb4qlhcxaI?t=211

His works can evoke longing, nostalgia, and the deepest of human emotions through his setting of modest melodies to intimate, surreal atmospheres. Since the 1990s, Hisaishi has earned worldwide acclaim for his unique musical style and compositional techniques.

This chapter examines the musical intersection of jazz and minimalism embodied in two compositions by Joe Hisaishi. The first section of the chapter offers a brief historical context for Studio Ghibli, including Hisaishi's early musical influences. The second half of the chapter presents analyses of two selected works – "One Summer's Day" from *Spirited Away* (2001) and "Bygone Days" from *Porco Rosso* (1992) – in order to demonstrate the convergence of jazz and minimalism in Hisaishi's music.

Studio Ghibli

Since the release of Studio Ghibli's first animated film in 1984, Joe Hisaishi has composed the soundtracks for twenty-five of the studio's productions. Among these works are celebrated fan-favorites, such as *Castle in the Sky* (1986), *My Neighbor Totoro* (1988), *Porco Rosso* (1992), *Princess Mononoke* (1997), *Spirited Away* (2001), and *Howl's Moving Castle* (2004). Many of Studio Ghibli's stories are known to include elements of magic, mysticism, dream states, and natural beauty. Overall, these films are cherished for their resonance with a kind of universal humanity, capturing familiar aspects of the human condition in a profound and inspiring way. In a New York Times article from 2022, the managing director of the Royal Philharmonic Orchestra, James Williams, offered his thoughts on the Studio Ghibli movies, stating:

> When you see those films, there's a certain humanity about the story lines, and that's absolutely reflected in Joe's music… It connects with people, regardless of their culture, and that's really powerful. What Joe has done is somehow retain that

> integrity of Japanese culture, brought in that Western tonal system and found a way for the two to retain their identities in perfect harmony.[45]

Williams observes that the Western tonal system is an integral part of Hisaishi's sound. This has certainly been a contributing factor to the worldwide appeal of Studio Ghibli movies, especially in the United States. In a 2020 interview with Kotaku, an online media platform owned by G/O Media, Hisaishi named some of his influences during his early musical development:

> I listened to all kinds of jazz during high school days. When I started learning the piano for the music college exam, I encountered Glen Gould's J.S. Bach "Invention and Symphonia." I also listened to Shostakovich, Schoenberg, Webern, Berg and Boulez. I was shocked by Terry Riley's "A Rainbow in Curved Air" and then began to listen to a lot of minimal composers like Philip Glass and Steve Reich.[46]

Hisaishi reveals through his comments how he has been influenced by classical European composers, jazz recordings, and even more modern minimalist composers. These influences and more can be heard in Hisaishi's film scoring as well as his live concerts, which often feature large symphonic orchestras.

One such concert was held in Tokyo in August 2008 at the sold-out Nippon Budokan venue, seating fourteen thousand audience members. The event was called "Joe Hisaishi in Budokan–25 Years With the Animations of Hayao Miyazaki."[47] The concert featured a total of over one thousand musicians, including the New Japan Philharmonic World

---

[45] Elisabeth Vincentelli, "The composer who turns Hayao Miyazaki's humane touch into music," *The Japan Times*, August 25, 2022, https://www.japantimes.co.jp/culture/2022/08/25/music/joe-hisaishi-studio-ghibli/.

[46] Brian Ashcraft, "Composer Joe Hisaishi Describes His Creative Process," *Kotaku*, February 4, 2020, https://kotaku.com/composer-joe-hisaishi-describes-his-creative-process-1841442868.

[47] GhibliWiki," Joe Hisaishi in Budokan," last modified June 29, 2011, http://www.nausicaa.net/wiki/Joe_Hisaishi_in_Budokan.

Dream Orchestra, eight hundred voices from the Ippan Koubo, Ritsuyuukai and Little Singers of Tokyo choirs, and one hundred and sixty marching band members from various Japanese high schools. The concert's program featured musical works from nine different Studio Ghibli films, scenes from which were projected onto a massive screen at the back of the stage. Hisaishi wrote new scores for some of the pieces which had been tailored for the large orchestra. In conducting my analysis of the composer's music, I was compelled to explore two recordings of this nature–one from the Tokyo Opera City Concert Hall and a second from the Budokan 25th Anniversary Concert.

Analysis

This chapter's analysis focuses primarily on two selected works by Joe Hisaishi. The first is "One Summer's Day," written for the 2001 film *Spirited Away*. The second is "Bygone Days," written for the 1992 film *Porco Rosso.* For "One Summer's Day," my analysis is based on a solo piano arrangement of the piece performed by Hisaishi in 2003 at the Tokyo Opera City Concert Hall.[48] For "Bygone Days," my analysis is based on a piano arrangement from the Studio Ghibli 25th anniversary concert.[49] I selected these arrangements for their context beyond the films. I contend they represent a desire from the composer to present his music in renditions that are separate from the films, rather than synchronized with them. What these two pieces have in common are shared themes of innocence, longing, nostalgia, and optimism. Moreover, they both have beautiful,

---

[48] foo9883, "08 Joe Hisaishi - One Summer's Day," YouTube video, 4:25, March 1, 2010, https://www.youtube.com/watch?v=JpXnIXP6j78.

[49] Adrián Zúñiga, "Porco Rosso - Joe Hisaishi in Budokan - Studio Ghibli 25 Years Concert," YouTube video, 4:59, March 12, 2013, https://www.youtube.com/watch?v=QsowHW0t7CA.

lyrical melodies that become enriched by lush jazz harmonies. Although Hisaishi did not begin experimenting with strictly minimalist music until later in his career, I argue that certain minimalist techniques can be found in these works. These include (1) the use and literal repetition of simple melodic phrases, (2) the development of texture through transposition of melodic material, and (3) nontraditional phrase structures and groups. Finally, I argue that Hisaishi makes frequent use of retrogressive phrase overlaps, which can be found at the closing cadences of several sections in "One Summer's Day."

Before introducing the two main analyses of this chapter, the next section discusses one example of minimalist composition and one example of jazz composition in order to highlight key characteristics of each genre.

The minimalist example I analyzed is Philip Glass's "Truman Sleeps" from the film *The Truman Show* (1998). Figure 3.1 below shows the introduction of the main theme at the start of the A section and includes mm.9-16. One idiomatic feature of this piece is the continuous eighth-note pattern of the piano's left hand. Glass uses this pattern to provide texture and supporting harmony, and therefore it is possible to hear this figure as a simple background for the main melodic theme. The theme itself is relatively simple and limited in its melodic range. The A section begins with a phrase group of two four-bar phrases, each consisting of purely stepwise motion. The melodic range across this phrase group spans just a major seventh. The first four-bar phrase is repeated almost literally by the second–the final C of the first phrase (m.11) is replaced by E^b and led to E in the second phrase (mm.15-16). The four-chord progression cycles for the entirety of the piece, offering little to no sense of tonal closure. These features are characteristic to the style of minimalist music–the use of limited melodic material combined with the uninterrupted repetition of phrase groups that undermines the sense of tonal closure.

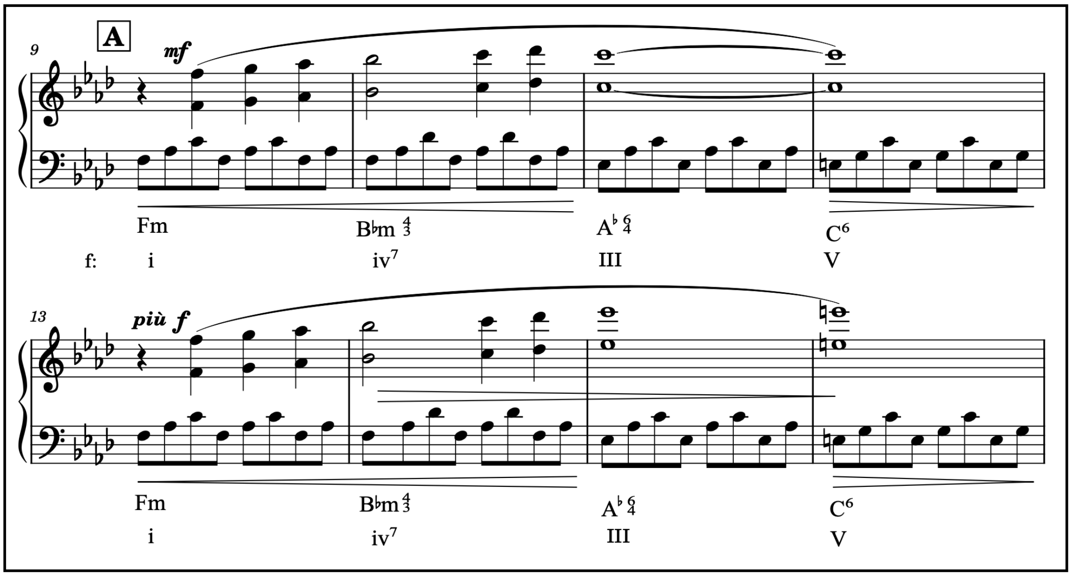

Figure 3.1 “Truman Sleeps” Introduction of Main Theme in A Section, mm.9-16, after transcription by David Jůna, edited and engraved by author.[50]

The next example, from the jazz repertoire, is Bart Howard’s “Fly Me To The Moon” (1954). This excerpt uses repetition as well, but the harmonic progression offers a slightly stronger sense of tonal closure. Figure 3.2 shows the introduction of the main theme at the start of the A section and includes mm.1-8. One characteristic feature of jazz is the use of extended or altered harmonies. In this passage, every chord is a seventh chord and some have also been altered (m.4, 7, 8). The music is full of vibrant, colorful harmonies with leading tones that pull the progression along, sometimes accelerating the harmonic rhythm (m.4, 8). The melodic theme is presented as another phrase group of two four-bar phrases. Although the two phrases give the impression of an antecedent/consequent pair,

[50] Sheet Download, “Truman Sleeps,” composed by Philip Glass, transcribed by David Jůna, https://sheetdownload.com/The-Truman-Show/Truman-Sleeps/13662, accessed February 9, 2024.

they cannot be characterized as such due to the fact that the second phrase does not end on the tonic chord (C major). Furthermore, the phrase group may also resemble a parallel period, but the first phrase (a) has a stronger cadence, ending on the tonic, and the second phrase (b) has a weaker cadence. This fact makes the label of 'parallel period' also incorrect. As such, the piece contains cadences that offer some sense of tonal closure, but they are not where one might expect them to be in a traditional tonal context.

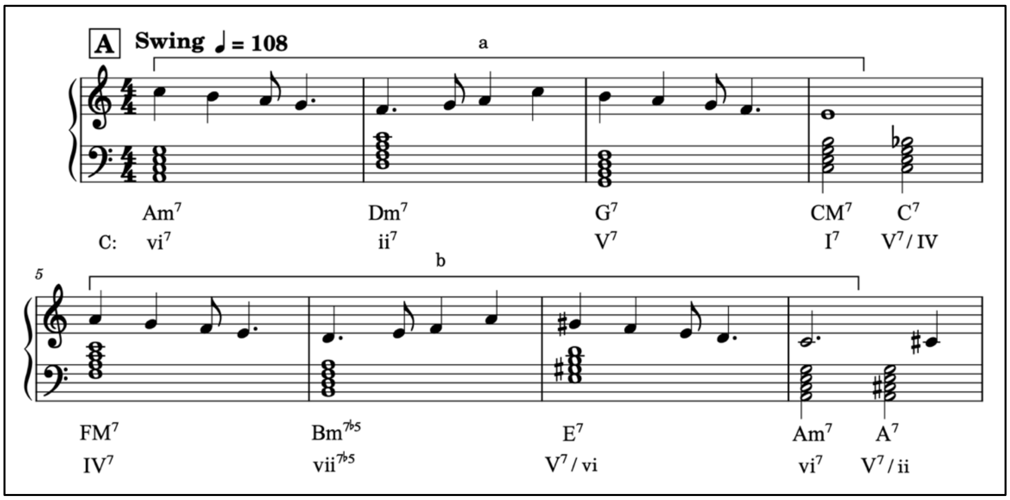

Figure 3.2 "Fly Me To The Moon" lead sheet, mm.1-8, transcription by author.

As for melodic material, this passage uses transposition rather than the kind of literal repetition seen in the previous example. Howard transposes the starting pitch of C in the first phrase (beat 1 of m.1) down to A in the second phrase (beat 1 of m.5). Here, the transposition is more traditionally tonal than those found in minimalist pieces, since all but one pitch ($G^{\#}$ in m.7) are diatonic to C major. In the analyses of Hisaishi's music to follow, I show how the composer uses many of the techniques stated above from both minimalist and jazz contexts. These techniques include the repetition and transposition of melodic material, the use of extended and altered harmonies, and the use of harmonic progressions that undermine the listener's sense of tonal closure.

One Summer's Day

The opening scene of *Spirited Away* is set to the music of "One Summer's Day," giving the audience an introduction to the story's world. The formal structure of the piece is unusual. On the one hand, there are multiple melodic themes which appear at the start of sections, giving us the impression of a strophic structure where these themes would be likely to repeat. However, the form is through-composed and only one section (C) can be heard to repeat in a closely related variation. Table 3.1 is a formal diagram of "One Summer's Day." The primary melodic themes are presented in sections A, B, and C.

Table 3.1 "One Summer's Day" Formal Diagram.[51]

	Intro	A	B	C	D	Intro'	E	C'	Total
Duration	4	17	8	14	4	4	4	14	69
mm. (-)	1-4	5-21	22-29	30-43	44-47	48-51	52-55	56-69	
Tonic	C	C	C	Am	C	C	C	Am	

The very first notes of the introduction are open, rolled piano chords in an upper-middle register, beginning with a beautiful F major 9 chord. These opening harmonies all have a bright and clear sonority. Figure 3.3 shows the introduction of the piece and includes mm.1-6. Asterisks in the figure indicate quartal or extended harmonies. Hisaishi uses stacked, open intervals to create an airy atmosphere with lots of space. Following the first F major 9 chord, the next three chords are quartal harmonies, commonly found in jazz music. I argue that the two slash chords in mm.2-3, while not purely quartal, have a quartal character because of the stacked perfect fourths in the treble clef.

[51] MuseScore, "One Summer's Day," composed by Joe Hisaishi, transcribed by Torby Brand, https://musescore.com/torbybrand/one-summer-s-day-spirited-away, accessed February 9, 2024.

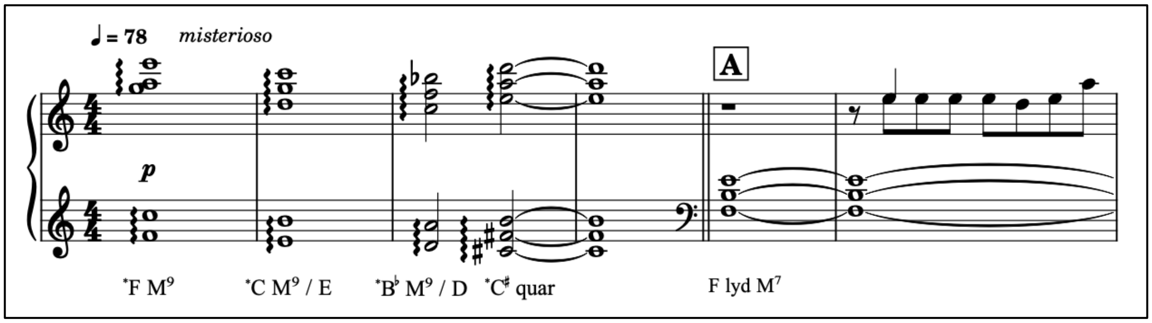

Figure 3.3 “One Summer’s Day” Intro, mm.1-6, transcription by Torby Brand, edited by author. Asterisks indicate quartal or extended harmonies.[52]

Quartal harmonies are used to obscure the harmonic direction of a musical passage since the anticipated resolution of these chords is ambiguous. In this passage, Hisaishi begins to establish one kind of expectation with a diatonic stepwise descent in the first three chords. However, he then interrupts this sequence with a purely quartal C# chord in m.3, which also accelerates the harmonic rhythm of the chord progression. With the arrival of the C# quartal chord, the tonal center is left behind and the feeling of home becomes ambiguous. This is especially due to the fact that the entire piece has only one authentic cadence in C major (m.16), which does not appear at the end of a formal section.

The main theme of the A section is a sentence structure. Figure 3.4 below displays this theme within the entire A section and includes mm.5-21. Asterisks in the figure indicate quartal, extended, or altered harmonies. The theme begins in measure 6 on successive articulations of the note E, a major 7th above the bass F. It is possible to hear the E in m.6 as a leading tone, but instead of resolving to F, the melody continues E-D-E-A-E-D in mm.2-3, lightly hovering around F and hinting at the resolution without ever arriving there. The contour of subphrase a’ is almost identical to that of subphrase a, only

[52] Accessible at https://musescore.com/torbybrand/one-summer-s-day-spirited-away

transposed down by step. I argue this recycling of material is an idiomatic technique of minimalist composition and the same technique appears in our next analysis.

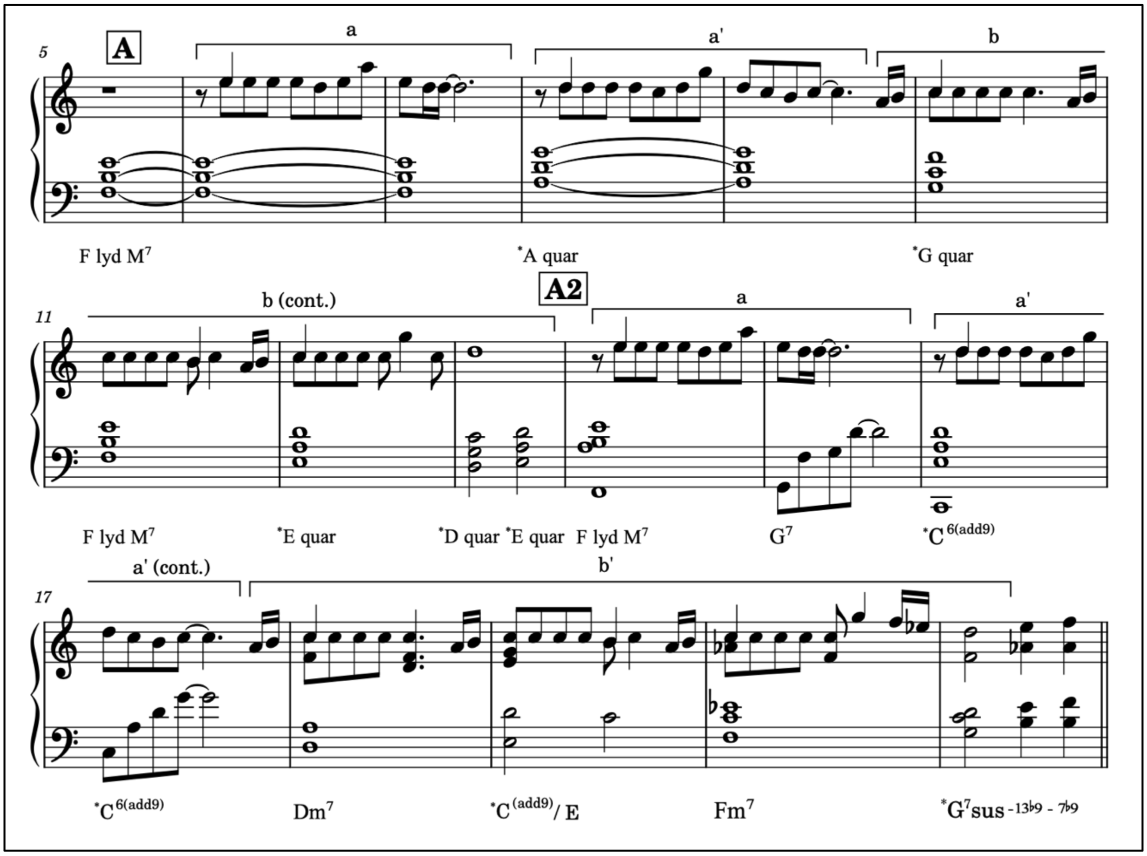

Figure 3.4 "One Summer's Day" A Section, mm.5-21, transcription by Torby Brand, edited by author. Asterisks indicate quartal, extended, or altered harmonies.[53]

The eight-bar phrase of the main theme is extended by one measure at the beginning of its first statement, contributing as well to the suspension of time. The sentence structure is repeated in what closely resembles a parallel interrupted period. However, the first sentence does not end with a half cadence, nor does the second sentence end with an authentic cadence. Instead, the first sentence overlaps into the second through modal harmonic motion (m.13). Thus, a phrase overlap occurs at the end of the first sentence.

[53] Ibid.

The second sentence ends with an altered dominant G^7sus chord, which does not resolve to the tonic C major. Instead, the B section begins in the following measure with an $F^{6(add9)}$ chord (shown in Figure 3.5). Here I present an original term to describe this type of harmonic motion: *retrogressive phrase overlap*. This term can be defined as a phrase overlap where a half cadence is followed by a predominant chord at the start of the next phrase. The retrogressive phrase overlap disrupts the listener's sense of tonal closure, since the music seems to restart on IV after the half cadence instead of resolving. Without a strong sense of tonal closure in the A section, the listener remains suspended in time as the overlap propels them into the B section. Throughout the piece, Hisaishi writes variations on the A section's melodic theme, but the A section never returns.

Figure 3.5 below shows the entire B section and includes mm.21-29. A red arrow indicates the retrogressive phrase overlap which occurs between the end of the A section and the start of the B section (mm.21-22). The B section contains two phrase groups, each consisting of two two-bar subphrases. The section begins on an $F^{6(add9)}$ chord and the first subphrase cadences with (C: (ii-V) / IV) [Gm^7-C^7] in m.23. The tonic harmony of "One Summer's Day" is, therefore, rather ambiguous since Hisaishi writes only one authentic cadence in C major (m.16), which does not occur at the end of a formal section. The primary tonal collection of the piece is C major, but I argue that most of the melodic themes suggest F lydian major 7 as the "home" tonality. This is further suggested by the piece's final cadence, discussed in more detail later in the chapter. The final cadence is another retrogressive phrase overlap that concludes the piece on an F lydian major 7 chord. The B section does, however, end with a half cadence on a G^9sus chord (m.29) which resolves to A minor in a deceptive resolution at the start of the C section.

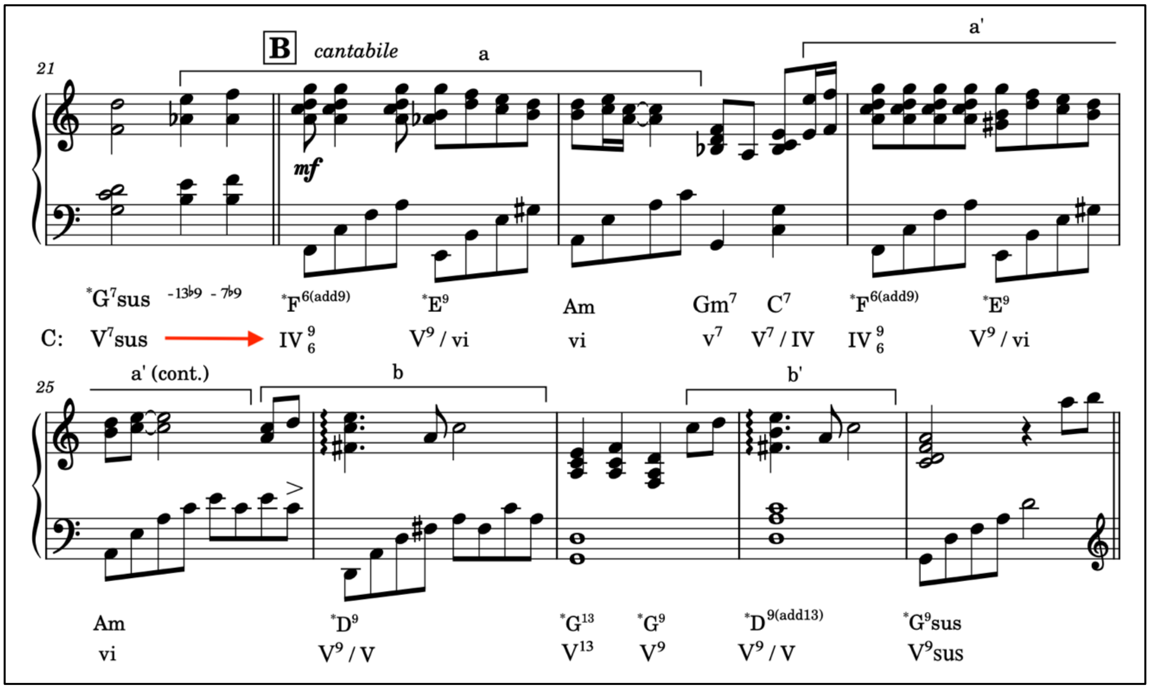

Figure 3.5 “One Summer’s Day” B Section, mm.21-29, transcription by Torby Brand, edited by author. Red arrow indicates retrogressive phrase overlap.[54]

The prominent use of extended and altered chord voicings throughout this piece provides strong evidence for the jazz influence in Hisaishi’s compositional style. As for minimalist techniques, each of the four subphrases in the B section are limited to a range of a perfect fifth and move mostly in stepwise motion.

The end of the C section contains another retrogressive phrase overlap, shown below in Figure 3.6, which includes mm.39-43. In this passage, Hisaishi prepares a traditional half cadence with (C: ii^{7b5}-V) [Dm^{7b5}-G], but he does not resolve the cadence to the tonic C major. Again, he follows the cadence with a predominant chord, this time an F lydian major 7 chord, which we heard earlier in the A section. A red arrow in mm.40-41 of Figure 3.6 indicates this retrogressive phrase overlap.

[54] Ibid.

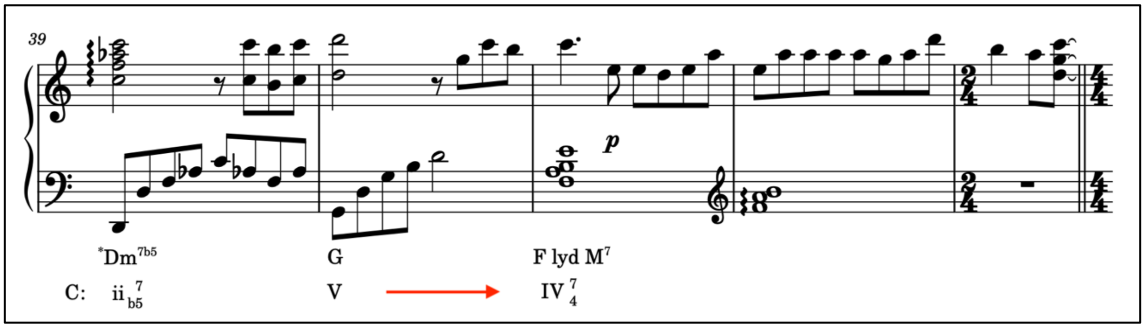

Figure 3.6 “One Summer’s Day” end of C Section, mm.39-43, transcription by Torby Brand, edited by author. Red arrow indicates Retrogressive Phrase Overlap. [55]

In the final measures of the piece, Hisaishi prepares the same half cadence (C: ii^{7b5}-V) [Dm^{7b5}-G] and once again follows this with an F lydian major 7 chord. Figure 3.7 below shows this retrogressive phrase overlap in the final measures of the piece, indicated by a red arrow, and includes mm.64-69. After the cadence, Hisaishi prolongs the final F lydian major 7 harmony with ascending arpeggios to conclude the piece.

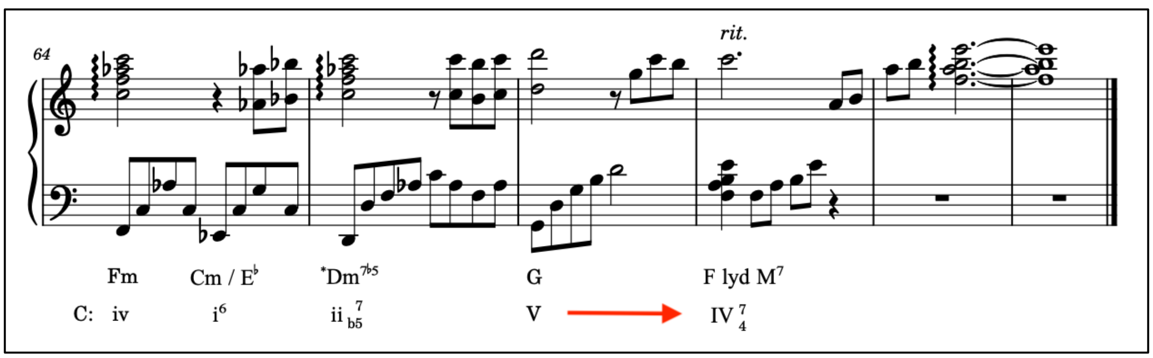

Figure 3.7 “One Summer’s Day” final measures, mm.64-69, transcription by Torby Brand, edited by author. Red arrow indicates retrogressive phrase overlap.[56]

Lastly, the piece’s D section provides more evidence for the minimalist techniques employed in Hisaishi’s music. Figure 3.8 below shows this passage, including mm.44-52.

[55] Ibid.

[56] Ibid.

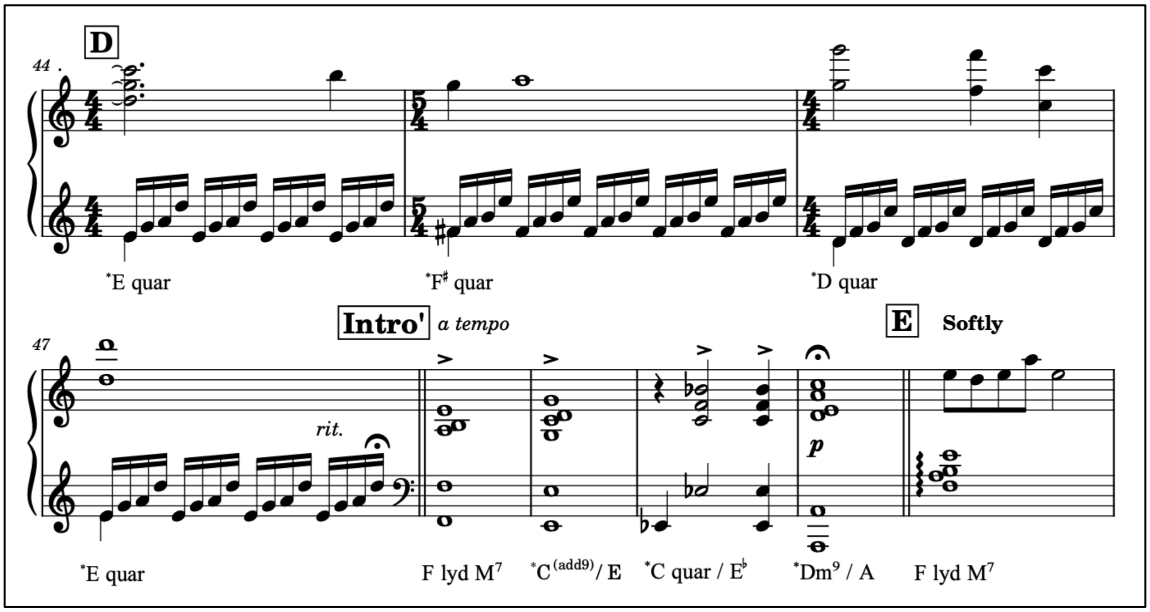

Figure 3.8 “One Summer’s Day” D Section, mm.44-52, transcription by Torby Brand, edited by author. Downward stems indicate transpositions of repeated quartal figure.[57]

In this excerpt, Hisaishi changes the overall texture dramatically–the piano's left hand plays a repeating sixteenth-note accompaniment figure on a new quartal chord voicing. Between each measure, Hisaishi develops the texture through successive transpositions of the repeating figure by step. These transpositions are examples of modal harmonic motion, another minimalist technique, since every voice moves in parallel motion and the quartal harmonic quality stays the same in each measure. The next analysis provides further evidence of Hisaishi’s minimalist and jazz composition techniques.

Bygone Days

This chapter’s second analysis focuses on “Bygone Days,” a piece from the film *Porco Rosso.* As mentioned at the start of this chapter, this analysis focuses on a live version of “Bygone Days” performed at the 25th Anniversary Studio Ghibli concert.

[57] Ibid.

The transcription I refer to (engraved by MuseScore user Hugolrx) captures Hisaishi's piano part and does not include the rest of the orchestra. The formal structure of "Bygone Days" most closely resembles a strophic form and, in that sense, is more traditional than the form of "One Summer's Day." Table 3.2 is a formal diagram of "Bygone Days."

Table 3.2 "Bygone Days" Formal Diagram.[58]

	Intro	|: A	B	C	A' :|	A	C	Coda	Total
Duration	9	10	4	7	6	10	7	8	88
mm. (-)	1-9	10-19	20-23	24-30	31-36	37-46	47-53	54-61	
Tonic	B^b	B^b	Cm	B^b	B^b	B^b	B^b	E^b	

The piece begins in B^b major with a wistful, ethereal melody on the subdominant. The melody quickly evolves with sixteenth-note ascending thirds, followed by a falling chromatic line in rapid triplets. Figure 3.9 below shows the piece's full introduction and includes mm.1-10. The passage is full of extended and altered harmonies, including a particularly spicy $D^{7\#9}$ chord (m.7). The melody's triplet figures, chromaticism (mm.4-5), and grace note embellishments (m.2) are all indicative of the piece's jazz influence and character. The tonal center of the piece, not unlike "One Summer's Day," is also somewhat ambiguous. The primary tonal collection of the piece is B^b major, but I argue that most of the melodic themes suggest E^b major as the "home" tonality. The piece begins and ends on an E^b major 7 chord. Moreover, most of the main melodic themes

[58] MuseScore, "Bygone Days," composed by Joe Hisaishi, transcribed by Hugolrx, https://musescore.com/user/10864221/scores/6632348, accessed February 9, 2024.

begin on an E^b major chord, and the A section's two phrases both end with B^b dominant chords–the first resolves to E^b major at the start of the next phrase, shown in Figure 3.10.

Figure 3.9 "Bygone Days" Intro, mm.1-6, transcription by Hugolrx, edited by author. Asterisks indicate extended or altered harmonies.[59]

The piece's primary melodic theme is presented in the A section and the other main themes all closely relate to this first theme. Figure 3.10 below shows the entire A section and includes mm.9-19. The section contains a phrase group which consists of two nearly identical five-bar phrases. Each five-bar phrase has a two-bar subphrase and a three-bar

[59] Accessible at https://musescore.com/user/10864221/scores/6632348.

subphrase. In a similar style to that of "One Summer's Day," the melody's first subphrase dances along in a mostly stepwise motion, and the contour of the second subphrase is almost identical, only transposed down by step and extended by one bar. Hisaishi uses many seventh chords in this passage, including both I M^7 and V^7 / IV in m. 14, and an altered dominant chord (F^+) in m.13, which is a common jazz alteration. The composer also writes E^b pedal points in m.11 and m.16, indicated by black arrows in Figure 3.10.

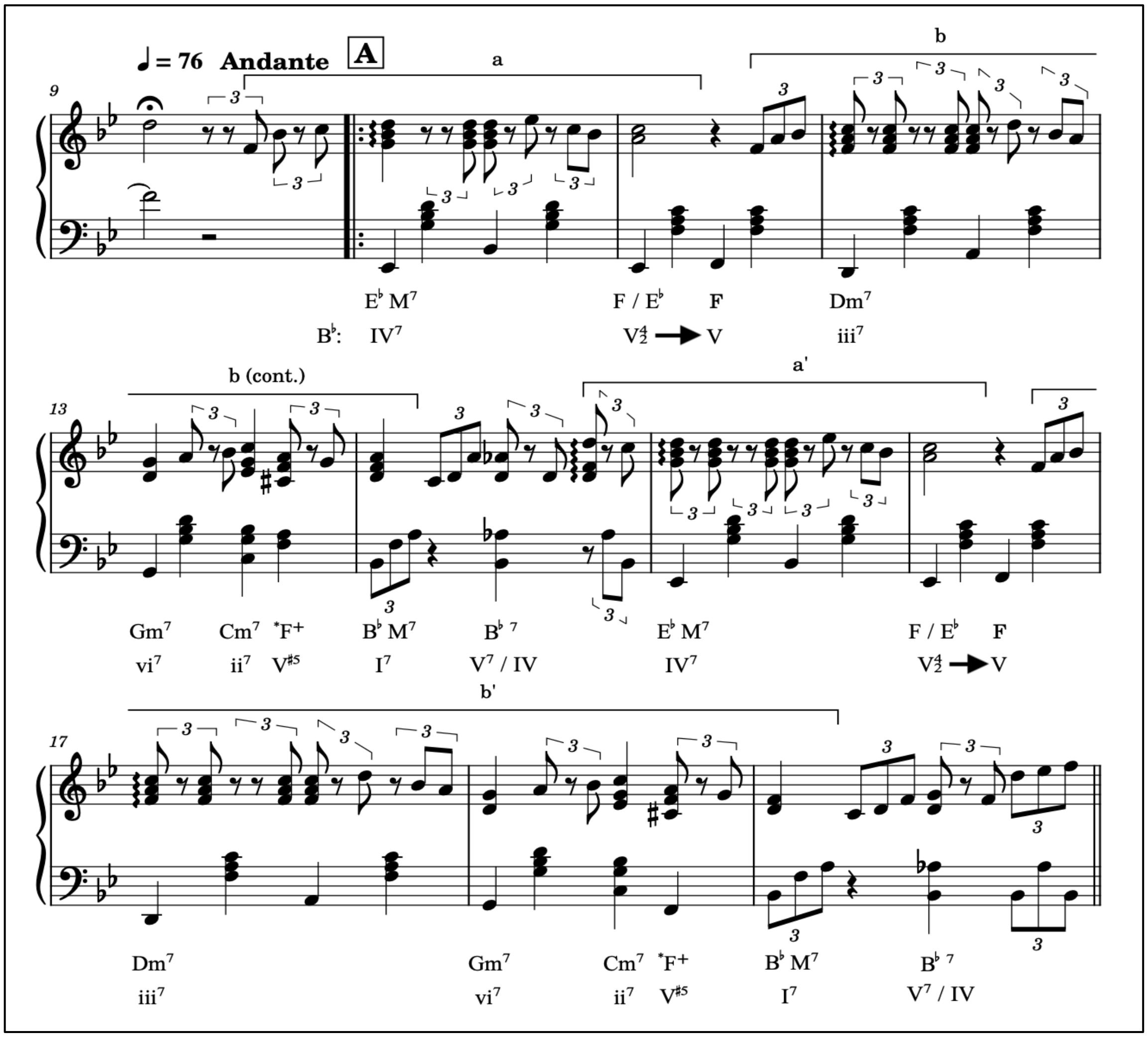

Figure 3.10 "Bygone Days" A Section, mm.9-19, transcription by Hugolrx, edited by author. Black arrows indicate pedal points and release.[60]

[60] Ibid.

Pedal points are a compelling jazz technique which have a two-fold effect. First, they create an unusual, colorful harmony and a feeling of suspension. Then, the suspension turns into propulsion when the pedal point is released. In this passage, the releases occur on beat 3 of m.11 and m.16.

The C section presents a powerful, unexpected cadence using quartal harmonies and modal harmonic motion. Figure 3.11 below shows the entire C section, including mm.24-30. The section begins with four measures alternating between E^b major 7 and B^b major.

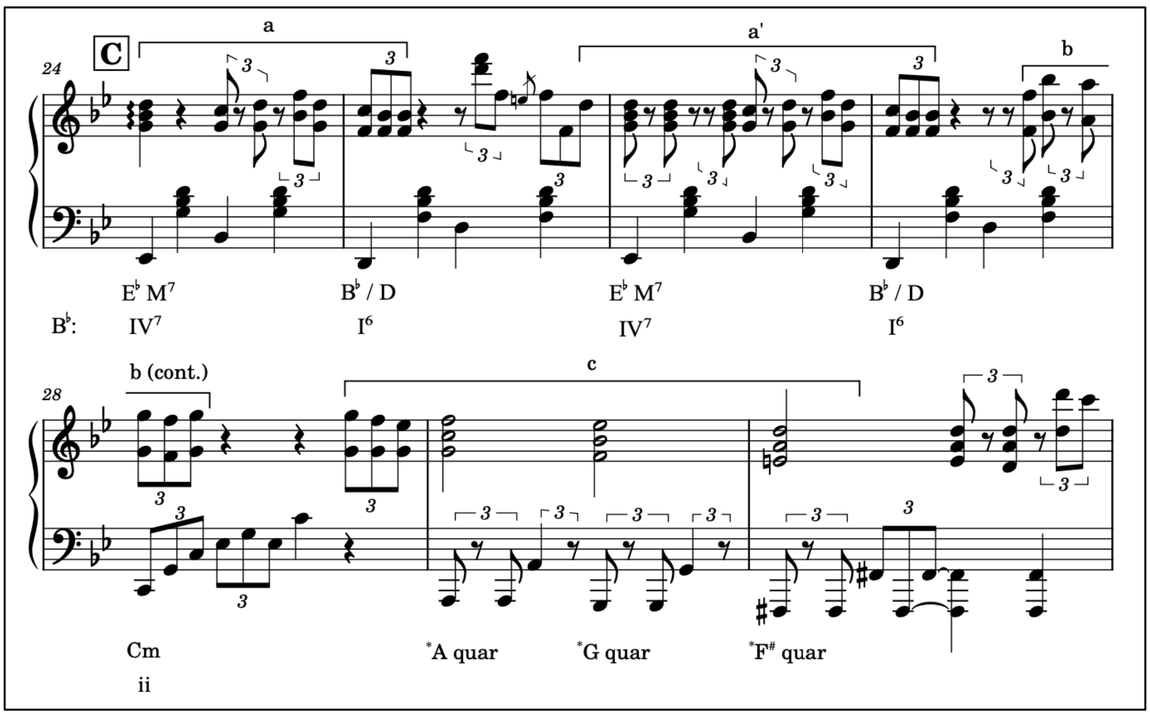

Figure 3.11 "Bygone Days" C Section, mm.24-30, transcription by Hugolrx, edited by author. Asterisks indicate quartal harmonies.[61]

Then, beginning in m.29, Hisaishi writes a stepwise descending bassline with stacked perfect fourths above the bass notes (mm.29-30). The cadential $F^\#$ quartal chord in m.30 closely resembles a D^{7b9} chord that occurs at the end of the B section (see Appendix B).

[61] Ibid.

Following the C section, the A' section contains several of the piece's most jazzy and compelling extended harmonies. Figure 3.12 shows the A' section, including mm.31-36.

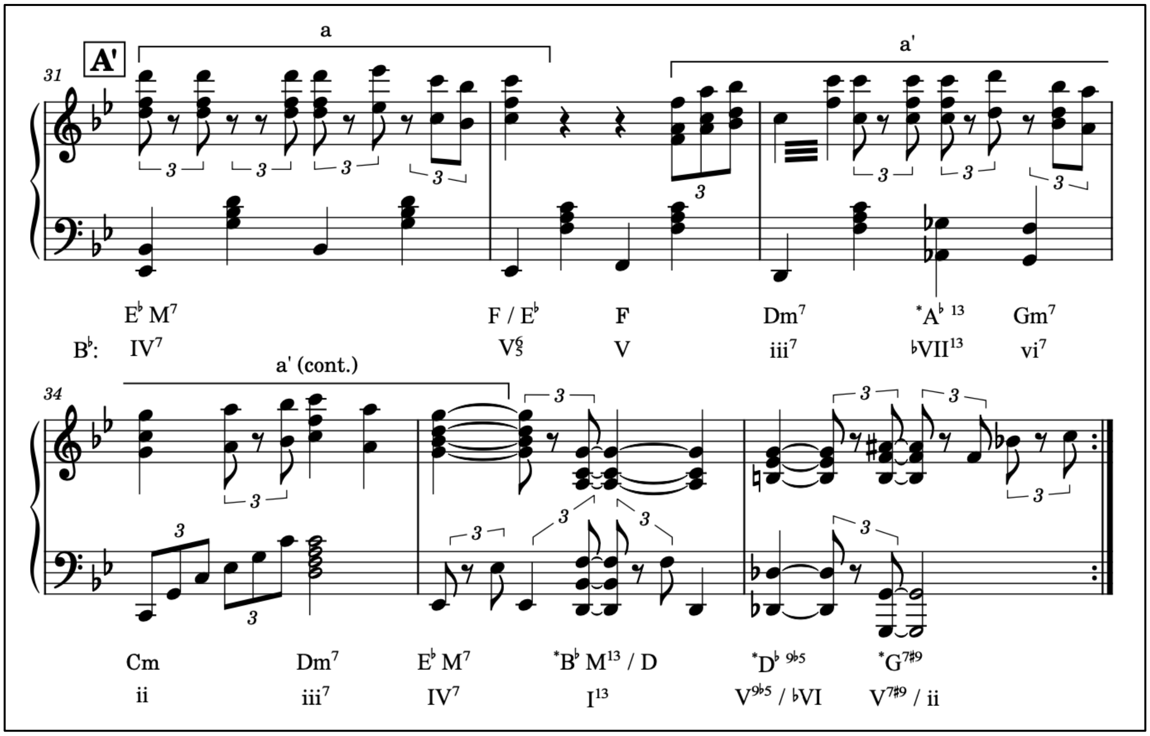

Figure 3.12 "Bygone Days" A' Section, mm.31-36, transcription by Hugolrx, edited by author. Asterisks indicate extended or altered harmonies.[62]

The section begins with the same melodic theme of the A section, now expanded to voice the melody in octaves. The harmonic progression becomes modal in m.33 as the piano's left hand moves down stepwise in parallel sevenths. The voicing of the B^b major 13 chord in m.35 is spread out in wide intervals and arguably even has a quintal character. Hisaishi uses even more extended harmonies to write an incredible cadence at the end of the section. In m.36, he writes a colorful $D^{b\ 9b5}$ chord which he voices as an E^b augmented triad over the D^b root. The final chord in the cadence is a particularly spicy $G^{7\#9}$ chord,

[62] Ibid.

which resolves to E^{b} major 7 at the start of the coda in the following measure. Interestingly, Hisaishi writes a similar cadence in “One Summer’s Day” in the middle of the C section, where a $C^{\# 9b5}$ chord is followed by a G^{9} chord, although the cadence resolves to A minor instead of E^{b} major 7 (see Appendix A). Overall, this passage highlights the composer’s brilliant palette of tone colors and rich jazz harmonies.

Finally, the coda arrives after one last refrain of the piece’s A section. Figure 3.13 below shows the entire coda, including mm.54-61. The coda is formally interesting because it closely resembles the C section–the first four measures (mm.54-57) alternate between the subdominant and tonic chords. However, the coda now tonicizes E^{b} major instead of B^{b} major, thus alternating between A^{b} major 7 and E^{b} major.

Figure 3.13 “Bygone Days” Coda, mm.54-61, transcription by Hugolrx, edited by author. Red circle indicates a tritone substitution.[63]

[63] Ibid.

Moreover, the coda's melody begins on G, whereas all previous melodies began on the note D, reflecting a transposition through the tonicization.

The final cadence of the piece is a quintessential example of another classic jazz technique–the tritone substitution. The tritone substitution functions similarly to a standard ii-V^7-I progression, where the V^7 chord is replaced with $^bII^7$. The V^7 and $^bII^7$ chords have root notes separated by a tritone–however, the $^bII^7$ chord functions in a similar way harmonically since both chords share the same tritone interval within them. In the key of E^b major, the dominant $B^{b\,7}$ chord contains the tritone of D-A^b. This tritone is also found in E^7, the $^bII^7$ chord, except enharmonically spelled as D-$G^\#$. In m.59, Hisaishi writes a tritone substitution with an E^{13} chord, indicated by a red circle in Figure 3.13. In this case, he leaves out the chordal seventh D, which would normally resolve to E^b, and instead doubles the root E. The $G^\#$ in the E^{13} chord does, however, resolve by half step to G natural in m.60. The piece ends on the newly tonicized harmony of E^b major 9 in triplet arpeggios. The arpeggios evolve in a hemiolic alternating pattern through the piano's upper register until arriving at the final melody note, D, a familiar major seventh above the bass.

Conclusion

This chapter analyzed two piano arrangements of works by Joe Hisaishi to highlight the jazz and minimalist compositional techniques found in each work. The first piece, "One Summer's Day," can be characterized by its through-composed formal structure that effectively recycles musical materials throughout, and its rather ambiguous tonal center between C major and F lydian major 7. Hisaishi employs quartal harmonies, extended and chromatically altered harmonies, all commonly found in jazz music. His treatment of phrase rhythm is especially interesting–he uses sentence structures in ways that follow Western classical tradition, but he also writes irregular parallel periods with cadences that disrupt the sense of tonal closure. The composer also makes consistent use of retrogressive phrase overlaps, an original term presented in this chapter, where a half cadence is followed by a predominant chord at the start of the next phrase. This technique undermines the sense of tonal closure and suspends the feeling of time. The second piece, "Bygone Days," can be characterized by its even larger palette of colorful jazz harmonies and its strophic formal structure with several repeated sections. This piece includes other jazz techniques, such as triplet figures and grace note embellishments. As an additional note, I argue that the high prevalence of triplet figures likens the meter of the piece to a classic swing groove. Hisaishi uses minimalist techniques in this piece as well, including modal harmonic motion, frequent repetition and transposition of melodic materials, and nontraditional phrase structures that disrupt the sense of tonal closure.

Jazz and minimalist music have had distinct cultures and histories for most of the twentieth century, though the music of Joe Hisaishi signals a harmonious convergence of these two contemporary genres. Even more, Hisaishi was born and raised in Japan, yet he elegantly blends his own style with old and new Western art traditions. This fact is likely evidence for a broader globalization of music that will continue to evolve in the future. The music of Joe Hisaishi has already stirred the hearts of millions of listeners and will hopefully resonate with future generations around the world.

References

Ashcraft, Brian. “Composer Joe Hisaishi Describes His Creative Process,” *Kotaku*, February 4, 2020, https://kotaku.com/composer-joe-hisaishi-describes-his-creative-process-1841442868.

foo9883, “08 Joe Hisaishi - One Summer's Day,” YouTube video, 4:25, March 1, 2010, https://www.youtube.com/watch?v=JpXnIXP6j78.

GhibliWiki,“ Joe Hisaishi in Budokan,” last modified June 29, 2011, http://www.nausicaa.net/wiki/Joe_Hisaishi_in_Budokan.

Glass, Philip. “Truman Sleeps.” See Sheet Download.

Hisaishi, Joe. “Bygone Days.” See MuseScore.

Hisaishi, Joe. “One Summer’s Day.” See MuseScore.

Joe Hisaishi Official, “Joe Hisaishi presents MUSIC FUTURE Vol.5 - Joe Hisaishi Interview,” YouTube video, 14:15, October 10, 2018, https://www.youtube.com/watch?v=MJb4qlhcxaI.

Moon, Kat. “How *Spirited Away* Changed Animation Forever,” *TIME*, July 20, 2021, https://time.com/6081937/spirited-away-changed-animation-studio-ghibli/.

MuseScore. “Bygone Days.” Composed by Joe Hisaishi. Transcribed by Hugolrx. https://musescore.com/user/10864221/scores/6632348. Accessed February 16, 2024.

MuseScore. “One Summer’s Day.” Composed by Joe Hisaishi. Transcribed by Torby Brand. https://musescore.com/torbybrand/one-summer-s-day-spirited-away. Accessed February 16, 2024.

Sheet Download. “Truman Sleeps.” Composed by Philip Glass. Transcribed by David Jůna. https://sheetdownload.com/The-Truman-Show/Truman-Sleeps/13662. Accessed February 16, 2024.

Vincentelli, Elisabeth. “The composer who turns Hayao Miyazaki’s humane touch into music,” *The Japan Times*, August 25, 2022, https://www.japantimes.co.jp/culture/2022/08/25/music/joe-hisaishi-studio-ghibli/.

Zúñiga, Adrián. “Porco Rosso - Joe Hisaishi in Budokan - Studio Ghibli 25 Years Concert,” YouTube video, 4:59, March 12, 2013, https://www.youtube.com/watch?v=QsowHW0t7CA.

Appendix A

Full transcription of Joe Hisaishi's "One Summer's Day." Piano arrangement from 2003 concert at the Tokyo Opera City Concert Hall. Transcription uploaded to MuseScore by user Torby Brand and can be accessed via web link in References. All figures show Torby Brand's transcription with additional markings made by the author of this chapter. These markings include all chord symbols, all roman numeral analysis, and select accidentals which have been reinterpreted.

One Summer's Day

(Spirited Away)

Arranged
for piano

By Joe Hisaishi
Arr: Torby Brand


a'
b
Am Gm7 C7 *F6(add9) *E9 Am *D9
vi v7 V7 / IV IV 9 6 V9 / vi vi V9 / V
b'
*G13 *G9 *D9(add13) *G9sus
V13 V9 V9 / V V9sus
Am Em6 4 F C6 4 Cm6 Dm7
mf
C# 9b5 G9 Am Em6 4 F C6 4 Cm6 Dm7 B♭m6 5 C7
Fm Cm6 *Dm7b5 G F lyd M7
p
D
C: ii 7 b5 V IV 7 4
*E quar *F# quar


46
Intro'
a tempo
rit.
*D quar
*E quar
F lyd M7
*C (add9) / E
*C quar / E♭
*Dm9 / A
E
Softly
52
F lyd M7
Em7
D9
B9sus
E7
Am
Em6/4
57
F
C6/4
Cm6
Dm7
C♯ 9b5
G9
Am
Em6/4
61
F
C6/4
Cm6
Dm7
B♭m6/5
C7
Fm
Cm / E♭
65
rit.
C: iv
i6
*Dm7♭5
G
F lyd M7
ii 7/b5
V
IV 7/4

Appendix B

Full transcription of Joe Hisaishi's "Bygone Days." Piano arrangement from 2008 Studio Ghibli 25th Anniversary Concert in Tokyo. Transcription uploaded to MuseScore by user Hugolrx and can be accessed via web link in References. All figures show Hugolrx's transcription with additional markings made by the author of this chapter. These markings include all chord symbols, all roman numeral analysis, the insertion of two formal sections (A and C) before the coda, and alternative interpretations of select harmonies found in mm.29-30 and mm.35-36.

The Bygone Days
(Kaerazaru Hibi)

Joe Hisaishi

a'
b'
15
E♭M7
IV7
F / E♭
F
V4/2 → V
Dm7
iii7
Gm7
vi7
Cm7
ii7
*F+
V♯5
19
B
cresc.
B♭M7
I7
B♭7
V7 / IV
Cm
Cm♮7
Cm7
Dm
Dm♮7
Dm7
E♭M7
E♭7
E♭(add6)
C
a
a'
23
mp
D7♭9
B♭:
E♭M7
IV7
B♭ / D
I6
E♭M7
IV7
27
b
c
B♭ / D
I6
Cm
ii
*A quar
*G quar


A'
a
*F# quar
E♭ M7
B♭: IV7
F / E♭ F
V6/5 V
a'
Dm7 *A♭ 13 Gm7
iii7 ♭VII13 vi7
Cm Dm7 E♭ M7 *B♭ M13 / D
ii iii7 IV7 I13
A
*D♭ 9♭5 *G7#9
V9♭5 / ♭VI V7#9 / ii
E♭ M7
F / E♭ F(add9)
Dm7 F / A
Gm7 Cm7 F+
B♭ M7 B♭ 7
E♭ M7
F / E♭ F(add9)


C
Dm7 F / A Gm7 Cm7 F+ B♭M7 B♭7 EbM7
Bb / D EbM7 Bb / D
Adagio ♩= 70
Coda
a
a'
A♭M7 *E♭(add9) / G A♭M *E♭(add9) / G
E♭: IV7 I(add9) IV7 I(add9)
b
rit........
*Fm9 *E13 *E♭M9
ii9 ♭II13 I9

Conclusion

This thesis has presented research in the domains of music theory and music technology. Across the three chapters, each topic investigated a different aspect of harmony and the modern ways in which humans communicate in terms of harmony.

The first chapter surveyed the latest research projects in AI-generated music to examine the technical and artistic limitations of machine learning when tasked with generating music. The music generated by MIDI networks tends to exhibit a distinct robotic quality, which sounds dispassionate and less organic than human-made music. The waveform networks can produce an output closer to a professionally mixed record, but the technology is not yet able to convince listeners that computers were absent from the production process of the song. I propose readily available solutions for both kinds of neural network. First, the solos, riffs, and melodies generated by MIDI networks would become more compelling if performed by humans. This would likely reconcile the lack of musicality or musicianship heard in the performances by digital instruments. Second, the waveform networks could be improved if there was an initial data-processing phase of *source separation* that occurred before training. In this way, each instrument of a song would be treated individually, and then later mixed into a final record, more closely resembling the traditional process of studio recording. Until machine learning algorithms enter the next generation, AI music may continue to be eclipsed by AI visual art.

The second chapter explored the history of written musical notation in order to most effectively transcribe compositions of electronic dance music. First, I examined two case studies of transcription systems which Charles Seeger presents in his essay; a graph from Milton Metfessel and a graph from Olav Gurvin. I then introduced the 2010 global hit song “Levels” by Avicii. By combining a traditional piano roll notation with an audio spectrogram, my transcription attempts to simultaneously display the pitch of the gliding synthesizer along with other arpeggiated notes. The result is a musical graph which conveys the precise pitch of the synthesizer at any beat in musical time.

The final chapter presented formal analyses of two musical works by the Japanese composer Joe Hisaishi from Studio Ghibli. I discuss how Hisaishi’s music borrows strongly from the Western classical tradition, but also from African-American jazz music and minimalist composition techniques. The two pieces I analyze present simple, modest melodies which become enriched by intimate, surreal atmospheres. The supporting harmony of the melodies often have a characteristic jazz quality to them through Hisaishi’s use of extended and altered harmonies. In terms of minimalist techniques, the composer effectively recycles and transposes short melodic fragments to create the fabric of his compositions. Finally, I argue that Hisaishi makes frequent use of dominant retrograde progressions which contribute to the feelings of suspended time and of wandering far away from home.

With further research, I’d like to explore applications of artificial intelligence in the music theory classroom. For example, students could be asked to improvise for their classmates on their primary instrument in a style or genre of their choice. Artificial intelligence already has the functionality to not only accompany any music making with additional instrumentation, but also transcribe improvised solos in real-time to be projected onto a screen for the class to follow along. Perhaps certain extended techniques like string glissandi and saxophone growls could be transcribed using a new kind of hybrid notation, offering more detail and information about pitch and harmonic content in this context and others. Finally, the music of Joe Hisaishi heralds a convergence of modern genres from around the world, but his music is not alone in this regard. I'd be interested to research the work of jazz saxophonist Kamasi Washington, who combines classical repertoire like Debussy's "Clair de Lune" with jazz instrumentation and vocals. I look forward to hearing even more hybrid genres yet to be born from the globalization of music and the ineffable creativity of modern composers.