Department of Music

Nicole Cosme-Clifford

How I Wrote My Prospectus

My initial plan was to find a strong thesis, gather literature, and write one section at a time throughout the fall semester of my third year. That plan quickly fell off the rails! The derailment was almost entirely due to my thesis. I wanted to write about deep learning tools in music scholarship but struggled to find a research question. In part, it was tough to figure out who my primary audience might be—computer scientists, who are excited by things like performance improvements to existing systems, or music scholars, who are excited by new humanist research questions. After a few months of deliberation, I decided on the latter, as I saw room to bring deep learning tools into humanist research. I became interested in how deep learning was being used by non-academic companies to analyze audio recordings. Specifically, things like OpenAI's Jukebox and Spotify's proprietary music processing pipeline proved that deep learning could extract features from audio recordings. These features have since shaped some of the most influential technologies of our time—generative AI and music recommendation—yet we can't say what they are or how they do this shaping. That piece of the unknown led to the final form of my prospectus. My thesis question asked how music streaming is reshaping the landscape and meaning of musical genre. I hypothesized that AI-learned audio features indirectly shape musical genre by participating in music recommendation pipelines, and I aimed to prove this by studying and building deep neural networks. The writing process moved quickly once I defined this thesis because the thesis dictated my sources and proposed project outline. My defense was rigorous but helpful. It made me realize that my original question was actually intractable (it turns out that music streaming companies are protective of their data). I shifted gears once I started working on the dissertation, and I prefer this new project to the original. In sum, my prospectus writing process looked nothing like the clear, straight road I had envisioned at the start. Instead, it was a wildly curving road riddled with offshoots, each teaching me something unique about myself, my research, and my future plans.

Advice for Prospectus Writers

1. Center your interests, knowledge, and skills. Choosing a dissertation thesis question before you've done the full dissertation research can feel a bit like finding a needle in a haystack. Every department and discipline will have unique conventions, but don't expect to have all the answers right away. Focus on the question that interests you based on your interests, knowledge, and skills. It is ok to have hypotheses and unknowns—these are the things you'll figure out as you do your dissertation research.
2. Embrace change. The thesis question and research plan you propose in your prospectus will likely change once you begin the dissertation. Don't spend too much time finding the perfect research question for your prospectus unless your department gives you a reason to. Find a good starting question and expect that your thoughts and convictions will evolve as you get deeper into your dissertation.

3. Give yourself some grace. The writing process will bring you moments of self-confidence and assuredness, as well as moments of self-doubt and uncertainty. When the good moments happen, be grateful for them. When the tough moments happen, know that you aren't alone in feeling things like frustration and doubt. Above all, remember that your research is a reflection of you, your thoughts, and your talents. You have something worthwhile to say, that only you can say. You will get through this!

I thought I might also share some questions I found helpful as I worked through my prospectus:

- Who am I in conversation with?
- What literature do I need to read? What does that literature say about my topic? Is anything missing or not quite right?
- How will I do this project (what data do I need, what methods will I use, etc.)? Is it tractable in the time frame I have?
- Where will this project take me?

The rise of "vibe": How deep learning algorithms are redefining genre in the age of music streaming

In 2022, singer Alison Krauss was a staple fixture of Spotify's most listened-to bluegrass playlists, as well as its most listened-to country playlists. This makes sense. Music streaming playlists are largely driven by listenership trends, and Krauss has earned the admiration of both bluegrass and country fans. She has won 6 Grammy awards for bluegrass and 13 for country, 14 awards from the International Bluegrass Music Association (IBMA), 8 awards from the Country Music Association (CMA), and she has been inducted into the Bluegrass Hall of Fame. Playlists, therefore, seem to paint Krauss clearly within a post-genre landscape: she is an in-between artist, a little bit country and a little bit bluegrass. What's more, her in-between-ness is constantly in flux, driven by the latest trends in listenership. But there's more to this painting than meets the eye.

All streaming artist profiles require metadata, identificatory information submitted by record-industry executives (Ingham 2019). Krauss' genre metadata (at the time of writing) tends to simply read "Country"--a one-dimensional label (used for marketing purposes) that contradicts her somewhere-in-between nature. The contradiction between Krauss' metadata and playlist data embodies a contemporary debate about the nature of genre: record-industry-assigned metadata paints genre as categorical and unchanging, while listener-based playlist data paints genre as social and fluid. Music streaming, therefore, balances the music industry's most lofty institutional actors (large PR firms, decision-makers, stakeholders, etc.) with listeners and artists, whose actions (often unknowingly) push against those actors. A similar balance has occupied music scholars for decades. In fact, music scholars and music streaming companies seem to be building the same internal argument against fixed genre categories and the institutions that enforce them; they are just using different tools to do so.[1] Different tools produce different ways of conceptualizing genre. My dissertation asks what can be gained from putting these differences in conversation.

Drawing on methods from Ted Underwood's (2019) book, *Distant Horizons*, and Nick Seaver's (2022) book, *Computing Taste*, I use algorithmic tools alongside discursive humanist frames to investigate the nature of genre in the age of music streaming. Since this task could quickly become intractable, I limit my work to bluegrass, a genre in which I participate, and which has a vibrant ontological debate. Communities of musicians and listeners that, like Alison Krauss, exist in a space between bluegrass and some other industry-fixed genre challenge prevailing definitions of bluegrass music--often unconsciously. This act of challenge is driven in a significant way by music streaming algorithms that mediate between music artists, music industry, and music listeners.

Through a combination of machine learning and humanistic inquiry, then, I define the in-between spaces around bluegrass as genre-like entities (tentatively called *vibes*). In doing so, I foreground issues of access to bluegrass music for artists and listeners, explore the musical features that necessarily characterize in-between spaces around bluegrass, and raise epistemological questions about digital music discovery and music streaming algorithms' analyses of human-cultural data. Ultimately, I claim that the definition of genre is being (re)negotiated in real-time via music streaming platforms. I then show that this phenomenon invites us to broaden the definition of musical genre and assess the utility of contemporary deep learning technologies for musical genre studies.

What is Bluegrass?

[1] Streaming companies are not nearly as consciously engaged with this work as scholars are. They are driven by consumerism—algorithms based on fixed genre categories don't turn a profit as well as those based on fluid categories (Ingham 2019).

Authors sometimes describe bluegrass as an unchanging genre (Rockwell 2012; Krakauer 2018; Petrusich 2021; Fenster 1993). In its most inner circle, perhaps this is true. Traditional artists like Del McCoury and The Stanley Brothers preserve the sound world established by Bill Monroe, the man credited with the birth of bluegrass. But outside of the inner circle, where artists like Alison Krauss lie, things are constantly in flux. This outer zone, whose occupants identify with bluegrass in some way, puts pressure on the core it encircles, but historically to little avail (Rockwell 2012; Fenster 1993; Farmelo 2001; Levine 2020). Joti Rockwell's (2012) discussion of BGRASS-L, an email chain of bluegrass practitioners from the early 1990s, gets to the heart of the matter. Rockwell traces the acronym WIBA (What is Bluegrass Anyway) through discussion threads in which bluegrass practitioners and listeners share opinions on the definition of bluegrass and who can make bluegrass music. Some notable (and concerning) excerpts are given below:

> "There's a significant reason why bluegrass should be shored up and it's worth defending the word bluegrass, because bluegrass has tangled with outside elements and has been blended with outside elements and often comes out injured as a result of the interaction. Not always, but often. When a bluegrass festival starts using rock'n'roll acts, sometimes the rock'n'roll acts start to dominate with louder music and louder fans, and the bluegrass can be left out in the cold."
>
> "If it ain't got a BANJO in it, it ain't BLUEGRASS!"
>
> "The bluegrass paradigm is not a matter of performers, or originators, or anything else; it's a matter of a particular sound, and that sound is exemplified in the music of the Blue Grass Boys [Monroe's band] during the Flatt and Scruggs era."
>
> "The biggest job of blue grass is to keep out what don't belong in it."
>
> "True Bluegrass cannot be obtained by people who live in a city and have not lived in the country at all, or the mountains or at least lived in that way of life."

The above quotes voice one side of a palpable, almost dialectical friction between those who seem to sit at the self-defined center of bluegrass and those who have historically been made to orbit around that center (except, of course, at bluegrass jams, a weird kind of DMZ). This friction has marked the genre since its inception, nearly 80 years ago, and it persists today. It is historically embedded within a number of problematic institutional structures: most notably, the early-20th-century racial segregation of recorded folk music,[2] historical gender dynamics (only men were visible in this branch of recorded music until the 60s or 70s, although women were avid folk musicians in non-commercial settings), and American socio-economics (bluegrass has traditionally rejected affluence and urbanness for various reasons, although it has since found company in affluent urban neighborhoods). Chapter 3 of my dissertation will tell this history of bluegrass in more detail. However, I am still sorting out how to unpack all of its threads and put them in conversation.

While the friction in bluegrass music has persisted since the genre's inception, its primary driver has evolved. Before the streaming era, the music industry dictated how music was marketed, and thus how people could build cultural communities around it (through interactions at record shops, community music events, bluegrass festivals, etc.). In doing so, it could easily reinforce the whims of bluegrass' self-imposed center, and by extension, the institutional structures they upheld. Friction, then, might be generated when marginalized artists sought but were denied entry into industry-curated spaces (for an example, see Fenster (1993)).

Contemporary music streaming companies challenge this process. Of course, these companies practice their own kinds of gatekeeping: they may privilege popular (read: profitable) songs and artists,[3] their algorithms may encode social biases when collecting data about listeners and use these biases for recommendation decisions,[4] and they often underpay music

[2] I haven't gotten a full grasp of the scope of this issue yet, or the two that follow. I want to bring in sources like Levine (2020), Farmelo (2001), and Rihannon Giddens [Keynote for the 2017 IBMA Conference] to discuss the whitewashing of old-time music and its spillover into bluegrass. Ian also mentioned Clifton's dissertation as a reference for thinking about race and preservationism in U.S. popular music.
[3] Hesmondhalgh, et. al., "The Impact of Algorithmically Driven Recommendation Systems on Music Consumption and Production - A Literature Review," 2023.
[4] Ibid.; Parham, "The Magic and Minstrelsy of Generative AI," 2023.

artists while overpaying music labels. Music critic Alex Ross (2022) even writes that Spotify "obliterat[es] artistic identity through the operation of its notorious algorithm." But Ross may have gone a bit too far. The gatekeeping that these companies practice (while still largely problematic) isn't always as exclusive as the institutional gatekeeping of the pre-streaming era. In fact, authors Tom Mcenaney & Kaitlyn Todd (2022), Robin James (2021), and Peli Grietzer (2017) take issue with Ross' critique that music streaming 'obliterates artistic identity.' They claim that the post-genre landscape built by algorithmic music streaming has become a place where musical identities and ideologies are being (re)negotiated, not simply destroyed.

Indeed, music streaming technology affords historically marginalized bluegrass artists *the possibility* of a direct route to listeners that bypasses traditional industry gatekeeping. I hypothesize that this bypass route is a new kind of ontological tool. It gives artists more agency in defining bluegrass, whether they are aware of this agency or not. Alison Krauss' music streaming data discrepancy is proof--listeners that come to know bluegrass through her music have a definition of the genre quite different than the general definition advanced in Rockwell's transcript of BGRASS-L (for one thing, women can be bluegrass musicians in Krauss' definition). That said, this isn't a perfect process; there are imbalances and problems with how it works (or doesn't work) in the real world. To understand how it came to be, and the messy details of its existence, we can reapproach the ontology of bluegrass through a decades-old discourse on the nature of musical genre.

On one side of the discourse, music industry institutions--like the Recording Academy, who presents the annual GRAMMY awards--have historically treated genre as a set of fixed categorical labels. These institutions construct and uphold genre categories according to formalized procedures. For example, *Billboard's* procedure for charting a song involves "looking at an artist's chart history" and "examining how and where the label is promoting and marketing the song." (Petrusich 2021) All of this serves a valid commercial aim: genre categories drive marketing strategies and organize catalogs of music in record stores and CD shops, which have historically shaped communities of listeners (Negus 1999; Petrusich 2021).[5]

On the other side of the discourse, humanities scholars (and inadvertently machine-learning engineers at music streaming companies) claim that fixed labels belie the true nature of genre. Literary scholar John Frow puts it concisely: genre is "a *relationship* [sic] between textual structures and the situations that occasion them." (Frow 2015) Music literature is rich with details about what this relationship looks like. For example, the second chapter of Fabian Holt's (2007) book, *Genre in Popular Music*, explains how a bluegrass-like genre called Americana emerged through the experiences of music makers and listeners. It is collective; it is enacted, not given; and it acquires meaning from social actions and context (Miller 1984). These qualities make Americana (and all genres) flexible in nature and subject to change over time. Humanities scholars and music streaming engineers agree: genre is an emergent and inherently social phenomenon. But, as I said above, this is the extent of their merger. The ideological and methodological tool sets that motivate their agreement are quite different. And neither toolset is sufficient on its own to explain how the streaming-driven connector between non-traditional bluegrass artists and listeners gives these artists and listeners more agency in defining bluegrass. Together, though, they form a wide-angle lens with high resolution.

We can see this lens take shape by looking deeper into music streaming. Here, musical identities are negotiated in real-time through a kind of digital chorology (explained later) that happens when computer algorithms mediate between music artists, music industry, and music listeners. This mediation is hyper-specific: algorithms learn mathematical relationships between musical audio content (bites of sound, metadata) and context (listenership data, social texts). Music streaming, therefore, casts a chrome coating over Frow's definition: genre is a relationship between textual structures and the situations that occasion them, *but* algorithms

[5] In recent years, fixed genre categories have also become "ground-truth" data for quantitative models of genre in the field of music information retrieval (MIR).

now play a major role in analyzing and shaping that relationship. Prevailing theories of genre in music scholarship, which leverage ethnographic methods and intra-humanities conversations but lean away from heavy discussions of musical content, seem ill-matched to the tech-heavy and content-dependent view of genre that emerges from music streaming (Holt 2007; Moore 2001; Sparling 2008; Drott 2013; Krakauer 2018; Rockwell 2012). In the same breath, prevailing ideas about genre in music streaming lack engagement with the critical frameworks that allow humanists to see algorithmic mediation as a means by which non-traditional bluegrass artists and listeners renegotiate the definition of bluegrass in real time.

If we carefully blend the two approaches, we get the wide-angle, high-resolution lens I mentioned earlier. A recent example is found in anthropologist Nick Seaver's (2022) study of music streaming algorithms' role in curating musical taste. He blends analyses of human musical listening behaviors with technical explanations of the mechanical forces that enable these behaviors. My dissertation builds on Seaver's work and pushes it toward musical genre studies. In chapter 1, I reproduce and deconstruct the algorithmic methods used in music streaming to expose the immovable link they create between musical content and musical context. This also contributes to the emerging field of algorithmic interpretability. In chapter 2, I draw on musical genre studies, ethnography, auditory culture, and algorithmic epistemology to explore how the relationships and sonic characteristics identified by music streaming algorithms participate in the formation of cultural music communities and could constitute avenues to identity building within unconscious digital spaces. In short, algorithmic methods allow us detailed access to the digital universe of relationships inside of music streaming. In chapter 3, these methods tell us that Alison Krauss' streaming data disjoint emerges from the unique way that streaming algorithms link music artists to communities of listeners. Leveraging this link, humanist frameworks give artists like Krauss, and their communities of listeners, new voices (intentional or not) within the conversation about what bluegrass is and who can make it.

In bringing machine learning and humanistic inquiry together, though, we reveal an even larger ontological phenomenon: the nature of musical genre has been altered by streaming algorithms. Bluegrass turns out to be just a microcosm. This invites us to broaden prevailing definitions of musical genre to account for the influence of music streaming, a call I take up in chapter 4.

Toward a Broadening of Musical Genre

Some authors have taken music streaming's unique pushback against fixed genre labels as a symptom of the death of genre. Perhaps under the strictest and most objective definition of genre, these authors are right. But others claim that genre isn't dead; its ontology has just shifted (Drott 2013; Petrusich 2021; Dredge 2020). That is, genre means something different now, in the streaming era, than it did in the pre-streaming era. Contemporary listeners have more options for music discovery than did pre-streaming listeners.[6] Rather than sifting through bins of physical recordings sporting *a priori* genre labels, for example, listeners may discover music through streaming services that--via machine learning algorithms--suggest songs based on a number of factors. As I began to explain above, this shift in music discovery is not only a source of new tension in the conversation about bluegrass ontology, but also the source of a purported shift in the ontology of genre.

The genre of the pre-streaming era is now running alongside a streaming-age phenomenon that some journalists call *vibe* (Mcenaney & Todd 2022; James 2021; Grietzer 2017; “Vibe, Mood, Energy” 2022; Petrusich 2021). Vibe is a broad term with an inherently murky definition. For example, Peli Grietzer’s dissertation in comparative literature (2017) defines ‘vibe’ as ‘a logic of formal affinity,’ a set of immanent aesthetics that coheres ‘a collection of objects or phenomena.’ Grietzer also associates and often conflates ‘vibe’ with

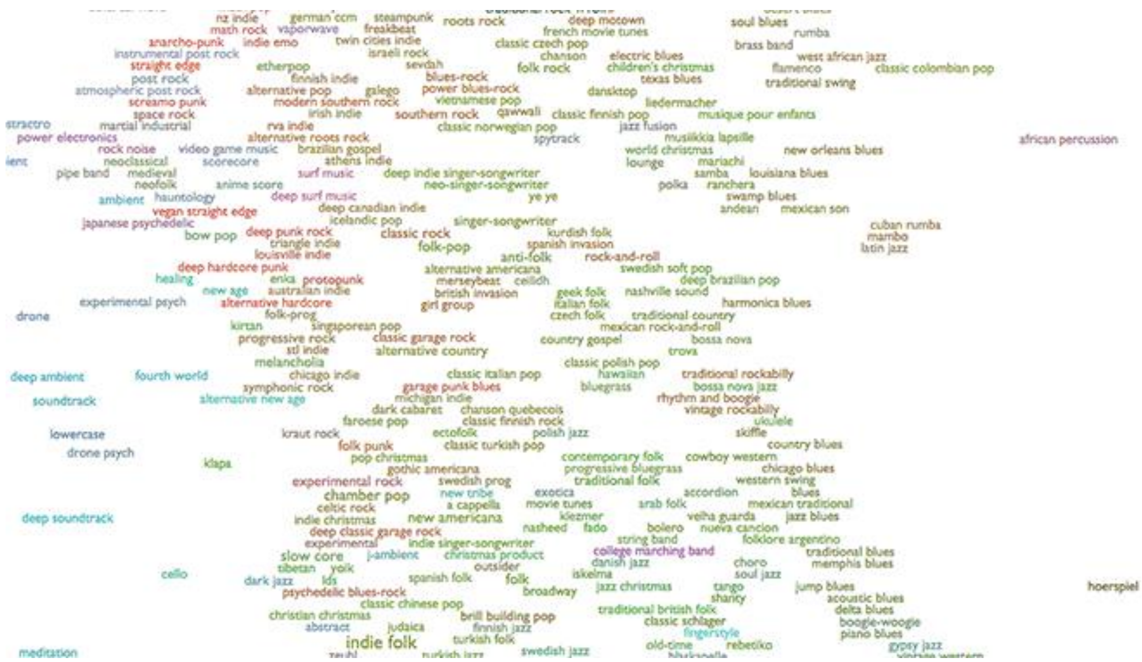
[6] This access-to-everything framework has earned serious and valid criticism (ex: Ross (2022)), but this is not my focus.

terms like style, genre, and *je ne sais quois*. The definition I will propose is a bit less murky. I will give a precise technological definition later, but for now it suffices to say that vibes are collections of music based on things like social relationships, personal listening trends, moods, or occasions (ex: playlists called 'Pollen' or 'Indie Bluegrass'). They arise within music streaming services that build listener networks. For example, journalist Amanda Petrusich (2021) quotes a Spotify executive's take on music recommendation: “We're not arbiters of taste...We’re here to try to connect our audience with different types of music, regardless of genre...If you look at playlists like ‘Pollen’ or ‘Warm,’ they really aren’t about specific genres. It’s more about having all of these songs woven together to satisfy a particular user.” Like the genre of the pre-streaming era, therefore, vibes are emergent, socially motivated, and subject to internal change. But unlike genre, vibes are defined by musical content and have a kind of impermanence.

This proposed definition of ‘vibe’ challenges our existing definitions of musical genre. It shifts us away from the world of albums and well-established genre labels (like bluegrass), and toward a world of playlists and ambiguously-named collections of sounds that seem genre-like. As I noted above, this shift is largely happening via machine learning algorithms that facilitate music discovery through a recommendation process. Spotify is more transparent than Apple Music about how this process works, so I will focus my attention there. Mostly, I will focus on a deep learning algorithm called a convolutional neural network (CNN). Seaver (2022) suggests that CNNs seem to be the source of vibe (although Seaver never uses the word or crystallizes the concept—he uses terms like “subgenre,” “activity,” and “cluster” to refer to the phenomenon I call vibe). I hypothesize that Seaver’s suggestion follows from the algorithm’s architecture and its unique way of processing musical audio.

A CNN's goal, in music streaming, appears to be to cluster a service’s users into groups based on song listenership and then predict which group(s) are most likely to play a song that none has yet heard. Here's how I imagine this works: If given a collection of musical audio recordings and some identifying information about listeners who have recently played those recordings, a CNN would extract features (bites of audio, metadata) from the recordings and learn a mathematical relationship between clusters of features and clusters of listeners.[7] When given a new song, then, the algorithm would scan the song for features that it knows and use those features to predict which clusters of listeners are most likely to play the song.

Example 1: "Every Noise at Once" Visualization of 1,387 Genres



Source: https://mymodernmet.com/every-noise-at-once-interactive-music-map/

It is important to note that CNNs have many layers of feature clusters (this puts the *deep* in deep learning algorithm). Vibes should emerge in the very last layers, which likely have millions of clusters. Each of these clusters should represent some collection of recordings, and thus, a vibe (Seaver 2022, Choi, Fazekas, and Sandler 2016; Pons et al. 2017; Chen and Wang 2017; Feng, Liu, and Yao 2017). What's more, the vibe clusters are implicitly organized by similarity, and this creates a kind of virtual coordinate space within the algorithm. We can visualize such a space (for example, the GIF above shows Spotify employee Glenn McDonald's

---

[7] This process has mixed reviews, but it works well enough for Spotify's many users to abandon older modes of music listening.

“map of musical genres,” called “Every Noise At Once”). Armed with its virtual space of related vibe clusters, then, the hypothetical CNN would make recommendation decisions by mapping these clusters of sounds to clusters of listeners.

If this mapping procedure is accurate, it tells us something unique about vibe, which is not reflected in McDonald’s visualization. The visualization shows *named* clusters (playlists and subgenres), but vibes can be any mapping of listeners and songs formed in the last layers of the company's algorithm, whether that mapping is realized in name or not.[8] This not only gives ‘vibe’ a very specific definition, but it also means that there are likely latent (unrealized) vibes that form as a byproduct of the algorithmic music recommendation process. What's more, this process is reproducible. The convolutional neural network, therefore, is both the mechanism through which vibe seems to emerge, and also an ideal tool for studying vibe from a music-theoretical perspective. If we could use the CNN as a tool, we could ask nuanced questions about vibe. For example, what sonic features define listenership clusters?[9]

My research suggests that the sonic features learned by convolutional neural networks are audio artifacts unknown to music theory and are hard to study (due to their small size).[10] Yet, they are statistically regular (they have predictive power). CNNs were developed to connect local features in images (ex: lines, curves, etc.) to image categories (ex: cat or dog). Their predisposition to local features means they struggle to identify long-range structures in musical audio (Choi, Fazekas, and Sandler 2016; Pons et al. 2017; Chen and Wang 2017; Feng, Liu, and Yao 2017). This makes them inherently bad at learning the kinds of musical semantics that music theorists are accustomed to because their field of vision is not long enough to capture long-range chord progressions, repeated melodies, or large-scale musical form. Instead, I

[8] In fact, McDonald claims that he needed to come up with lots of arbitrary cluster names like “shiver pop” before making this visualization (Seaver 2022, 131).
[9] Mcenaney & Todd (2022) claims that metadata is the only contributor, but Seaver (2022) claims that audio features hold significant weight. I agree with the latter.
[10] These features, audio bites, can be played back for human listening. But they are generally difficult for humans to perceive due to their short length.

believe that they capture the sonic fingerprints of small, fleeting moments in audio recordings--the sound of a guitar or the overtones that happen when two voices blend for one note. Therefore, the clusters of features in the last layers of a hypothetical CNN for music streaming, which represent what I call vibe, are theoretically just summaries of the sonic fingerprints in the audio recordings favored by different groups of listeners (Seaver 2022). If this is true, it might help us understand the Alison Krauss example: she fits somewhere-in-between bluegrass and country partly because the sonic fingerprints in her recordings are linked to listener experiences and actions in both genres.

That said, this idea raises just as many epistemological questions as it does music-theoretical ones. In my hypothetical scenario, users and songs begin as data points in a remote database system. Their listening habits--encoded as pulses of light--whiz through electrical wires until they trigger unique strings of on-off switches deep inside a supercomputer. With enough of this whizzing and triggering (the training of a neural network), the supercomputer learns which habits are like others, and it extrapolates this information in a consequential way. Two users who listen to the same songs must have aesthetic similarities. Perhaps they also have ideological similarities. A connection has been made (and the seeds of vibe planted). However, this connection wasn't made in person; it was made by an algorithm acting as a social proxy. What do such music streaming algorithms know about human listener behavior, and what do we know about algorithmic behavior? Sociologist Susanne Krasmann (2020) writes that recommendation algorithms only know superficial things about human behavior–they can identify regular patterns of human action, but they cannot give meaning to those patterns. This task has been taken up, instead, by various human actors. Who are the human actors deriving meaning from streaming algorithm data? What kinds of meanings do they propose?

Likewise, our knowledge of streaming algorithms is limited. If CNNs in music streaming do produce realized and unrealized vibes, which live in a digital algorithmic space, then these algorithms create genre-like forms that were non-existent or repressed in the pre-streaming world. These forms depend on the idea that algorithms learn the particulars of an individual and simultaneously assert that those particulars are not from the individual but from a collective that the individual participates in. What are the vibes that never get realized? What particulars do they capture? Where do they live in the algorithm's digital spatial landscape? How is distance measured in that landscape? If distance were measured differently, might these unrealized vibes be realized? Scholars like Seaver (2021) and Underwood & So (2021) have begun to approach these questions within the domain of music streaming, but there is still plenty of work to be done.

Summing it all up

The streaming-based process of negotiating vibe between artists, industry, and listeners forges a connection through which Krauss offers listeners her definition of bluegrass. This is essentially a positive feedback loop: listeners discover non-traditional bluegrass music through unconscious connections made inside the algorithmic space drawn by some neural network; artists with enough revenue and streaming momentum, then, produce new music to feed the tastes of these listeners. This process subconsciously pushes back against bluegrass' historically exclusionary center by proposing new definitions of what bluegrass music sounds like and who can participate in it (as artists or listeners). Bluegrass is no longer defined only by traditional institutional gatekeepers. It is also being defined by music streaming companies and the people who work for them, and between music artists and listeners--both traditional and non-traditional. Sometimes by people who are unaware of or ambivalent to bluegrass' history.

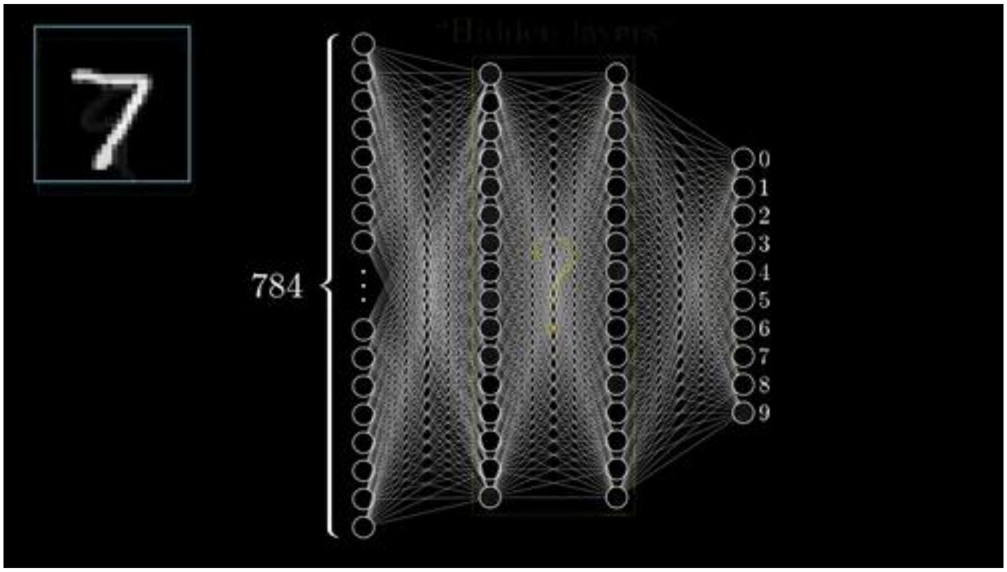
On a broader scale, musical genre can no longer be defined without consideration of the phenomenon I have called vibe and the mechanism that produces it. For now, it suffices to say that vibe complexifies and challenges the prevailing scholarly consensus about musical genre. The resulting discourse on genre is, therefore, fuzzy at best. A musical genre that includes vibe must be social and fluid, distinct from the fixed categorizations often found in quantitative academia and the record industry, but also mindful of contemporary technology and its role in building musical identities and communities of listeners through a music recommendation process. This discourse frames my proposed dissertation at the highest level. Adopting a methodological approach that blends humanist and technological tools and frameworks, I ask nuanced questions about the nature of both bluegrass music and musical genre in the age of music streaming.

In chapter 1, I use listener data and corresponding audio data[11] to build a mock convolutional neural network for music recommendation (likely on a small scale). There are countless ways to use this algorithm, but I want to make an intervention into the world of algorithmic interpretability. This is a burgeoning and increasingly necessary field—I'll raise the existential horror induced by ChatGPT as an example—but interpretability work is notoriously difficult, due to deep neural networks' complex architectures (explained below).

The GIF below describes a generic neural network for image classification (this will turn out to be relevant to music in a minute). The network's goal is to predict which number is shown in a given image. The image of the number 7 in the top left corner of the GIF has 784 pixels, each with a grayscale value between 0 (black) and 1 (white). Rather than keeping them in a grid arrangement, we can arrange the pixels into a one-dimensional list, preserving their grayscale values. This 784-entry list becomes the input to the neural network in the GIF: every circle (called a node) in its leftmost layer is a number representing the grayscale value of one pixel in the image. The second layer (the subsequent column of circles) complicates things by invoking the concept of generalization-–the essence of deep learning's success.

Example 1: Deep Neural Network Architecture

[11] Spotify has an API where I might find data. Another recent source of audio data for deep learning is Google's MusicCaps dataset (released January 2023).



Source: [Best Neural Network GIFs | Gfycat]

The neural net could look at the 784 pixels in its first layer and say, “this must be a 7,” but this process is inefficient. For it to work, the network would have to effectively memorize the pixel arrangement from this one image. Consequently, it would poorly classify images of 7’s with slightly different pixel arrangements. Generalization is a more efficient alternative to memorization. It says that all handwritten 7’s must share some concise set of visual features that reveals their collective numeric identity. The neural network’s task is to find that set, and there are several ways it can do this. In the GIF above, for example, the second layer is made smaller than the first, which effectively forces the network to constrain its decision-making criteria. It says, “I cannot use all 784 pixels, so I must group similar ones, save some abstract representation of the group, and retain only those representations that are valuable to my classification decision.” A typical neural network does this reduction process for each of its layers. Therefore, each layer is some kind of abstract representation of the previous layer.

Things get complicated when we consider that the process above happens for every image in the dataset, with the input layer resetting each time. Therefore, circular nodes may carry information about all numbers in the dataset, not just the number 7. The neural net then learns which nodes are helpful for classifying which numbers. It does this by assigning numerical weights (shown as lines between nodes) that quantify how helpful a node will be. We can see this happen in the GIF: helpful nodes appear as white circles in layers 2 and 3, connected to the previous layer by yellow lines, and a single white circle representing the label "7" appears in the last layer. A neural network easily understands the details of this process. But since nodes are non-unique, layers become progressively more abstract, and datasets are large, it is difficult for humans to figure out which nodes the algorithm uses to classify a particular image and what visual features those nodes represent (the GIF above is a demonstration of theory rather than practice).

The CNN that I will build is slightly more complex than the neural net in the GIF, but it is similar and also comes from image processing. A digital audio file is like an image in that it is a 1D string of numbers representing the amplitude change of a musical sound wave.[12] We can imagine, therefore, that each node in the first layer of the diagram above represents a small slice (or sample) of audio with a single amplitude value, rather than a pixel. I have many thoughts about the drawbacks of falsely equating pixels with audio samples, but I won't get into them here. I will only say that I believe sample-based CNNs cannot generalize nearly as well as pixel-based ones. For example, a node in the second layer of the network above might collect all pixels that fade from white to black–the rightmost edge of the numbers 7, 4, or 1–and reduce them to some abstraction that retains most of their integrity. It cannot do the same for audio samples: each sample is potentially unique, and at best it is most similar to other samples from the same song, album, or artist.

The likelihood of robust generalization gets lower as things get broader (which seems counterintuitive). For example, I would expect a CNN to generalize moderately well across songs from the same album but poorly across artists. This explains my earlier hypothesis about

[12] "Amplitude" is the air pressure when a sound wave hits a recording device. It is generally unique to the timbre of the instrument being played (ex: a violin and a viola playing the same A have differently shaped pressure waves), to the combination of instruments being played, and to the notes being played.

vibe: that CNNs are predisposed to create it. Sample-based CNNs will attempt to generalize in the most mathematically efficient way they can—it is built into their architecture. This means their node clusters will contain audio snippets with similar amplitude records, and thus, similar sound profiles. Indeed, Seaver (2022) shows that early-layer nodes cluster similar-sounding samples (like instrument sounds from songs on the same album) and late-layer nodes group these clusters, eventually producing what I call vibe. My broad hypothesis is that the generalization process, when coupled with the fact that CNNs can only see music as a string of amplitude values, predisposes these networks to build vibes—collections of songs that are mathematically and sonically alike.

Though some research has been done on this topic, it isn't enough for me to say if my hypothesis is correct. In a typical neural network, there are thousands, if not tens of thousands, of clusters. This has historically been a barrier to interpretability research—most nodes and layers are unseen and the features they represent are unknown. I am still working out a strategy for sifting through all of these, but once I do, I will ask why the features discovered by the CNN's various layers might lead to genre-like clusters of songs. Is such a model robust, and what does it leave out? Does it learn to connect features between layers in musically semantic ways (making semantics part of the recommendation decision), or is recommendation based purely on collections of sound, in which any semantic features are merely byproducts of the act of bringing these sound collections together. As I suggested above, my past research with CNNs tells me that the latter case is most likely: vibes are almost certainly sound-driven, and the recordings within a vibe likely have the same kinds of sounds.[13] What musical features define vibes, how many are there, which ones are sparse (few connections to other nodes) and which are dense (many connections to other nodes), which are completely unrealized, and what socio-technical conditions lead to sparseness, denseness, or unrealized-ness?

[13] This is consistent with some complaints about Spotify's so-called "vibe" playlists (which to many sound like a perpetual wash of sound after enough time).

To push things further, I will employ the help of another deep learning algorithm called Jukebox.[14] This is not a *discriminative* algorithm like the CNN (which seeks to learn some mapping function between input and output, largely for predictive purposes) but a *generative* one (which seeks to learn the mathematical distribution of its inputs so that it may sample from this distribution to produce novel outputs). Jukebox spits out "novel" recordings of pop styles that generally reek of machine artifice, although the algorithm is an impressive feat of engineering. It seems to sample existing recordings and recombine the samples in non-syntactical ways, also grouping like sounds with like. I could, therefore, ask it to generate music in the style of some artist, and then feed those recordings into the CNN. If the CNN matches the recordings to listener clusters just as easily as it matches recordings from the artists on which the samples are based, then I might have more evidence to say that the CNN is purely sound-based and not cognizant of musical semantics. If the matching is significantly worse, then perhaps the CNN does learn semantics between its layers (adding a wrinkle to my hypothesis above, and lots of others), and researchers have not been able to see this because they have not looked at inter-layer connections in detail.

In chapter 2, I consider the epistemological repercussions of my analyses in chapter 1--what do music streaming algorithms know about us, and what do we know about them? Who are the human actors deriving meaning from streaming algorithm data? What is the topography of the landscape of listeners drawn by a music streaming CNN? That is, how are listening communities being clustered, what are the metrics that draw boundaries between clusters, and how are distances between clusters measured? How is distance between recordings, artists, and listeners computed, what relationships do these distance metrics privilege, and what relationships do they disadvantage? What other options for distance metrics exist?

[14] Dhariwal, et. al. 2020

In the remaining chapters, I draw on the work of Ted Underwood (2019), a literary scholar who asserts that machine learning algorithms provide a unique way of exploring thorny, large-scale humanist questions about genre. I take the technical work in chapters 1 and 2 as a preamble to two large-scale humanist questions, the subjects of chapters 3 and 4, respectively. In chapter 3, I focus on bluegrass, a genre whose history has been marked by a debate about its ontological status. Here, I propose the idea that bluegrass-adjacent vibes, like those built by Alison Krauss and similar artists, reinvigorate an ontological debate that has always been. I give an overview of bluegrass, its historically exclusionary practices, and the ontological debate that has emerged from those practices. I then ask how algorithms give non-traditional artists (I focus on artists who identify as women, artists associated with the vibe called Indie bluegrass, and artists associated with the vibe called Latingrass) some kind of voice in this debate--what does that voice look like, who are the artists involved, and how does music streaming enable them to (re)negotiate what bluegrass is and who can make it? In chapter 4, I zoom out further, to address the definition of genre. I discuss prevailing theories of musical genre and propose a broadened, contemporary definition that includes streaming age phenomena like vibe. In doing so, I hope to show vibe not as an isolated incident, but as a phenomenon that precipitates from a long evolution of genre.

Citations

Chen, Ning, and Shijun Wang. 2017. “High-Level Music Descriptor Extraction Algorithm Based on Combination of Multi-Channel CNNs and LSTM.”

Choi, Keunwoo, György Fazekas, and M. Sandler. 2016. “Explaining Deep Convolutional Neural Networks on Music Classification.”

Dhariwal, Prafulla, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. "Jukebox: A generative model for music." *arXiv preprint arXiv:2005.00341* (2020).

Dredge, Stuart. 2020. “Lorem, Pollen and Oyster: How Spotify’s Genreless Playlists Are ‘Driven by Culture.’” September 23, 2020. https://musically.com/2020/09/23/lorem-pollen-and-oyster-how-spotifys-genreless-playlists-are-driven-by-culture/.

Drott, Eric. 2013. “The End(s) of Genre.” *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance* 57 (1): 1–45.

Farmelo, Allen. 2001. “Another History of Bluegrass: The Segregation of Popular Music in the United States, 1820–1900.” *Popular Music and Society* 25 (1–2): 179–203.

Feng, Lin, Shenlan Liu, and Jianing Yao. 2017. “Music Genre Classification with Paralleling Recurrent Convolutional Neural Network.”

Fenster, Mark. 1993. “Alison Krauss and the Contemporary-Traditional Conflict in Modern Bluegrass.” In *All That Glitters: Country Music in America, Edited by George H. Lewis, Popular Press - Popular Culture*, edited by George H. Lewis, 317–28. Bowling Green, Ohio: Bowling Green State University Popular Press.

Frow, John. 2015. *Genre*. London: Routledge, Taylor & Francis Group.

Grietzer, Peli. 2017. “A Theory of Vibe.” *Glass Bead*. https://www.glass-bead.org/wp-content/uploads/GB_Site-1_Peli-Grietzer_Eng.pdf.

Holt, Fabian. 2007. *Genre in Popular Music*. Chicago, UNITED STATES: University of Chicago Press.

Ingham, Tim. 2019. “The Record Industry Gave Everything Away to Spotify. Could It Change Its Mind?” Rolling Stone. September 26, 2019. https://www.rollingstone.com/music/music-features/record-industry-gave-everything-away-to-spotify-more-like-netflx-890417/.

James, Robin. 2021. “Is A Vibe the Same Thing As A Style?” Its Her Factory Newsletter. December 11, 2021. https://itsherfactory.substack.com/p/is-a-vibe-the-same-thing-as-a-style?s=r.

Krakauer, Benjamin. 2018. “A ‘Traditional’ Music Scene and Its Fringes: Experimental Bluegrass of 1970s New York City.” *American Music* 36 (2): 163–93.

Krasmann, Susanne. 2020. “The Logic of the Surface: On the Epistemology of Algorithms in Times of Big Data.” *Information, Communication and Society* 23 (14): 2096–2109.

Levine, Chloe E. W. 2020. “How Blue Is Bluegrass?” *Harvard Political Review*, April 8, 2020. https://harvardpolitics.com/how-blue-is-bluegrass/.

Miller, Carolyn R. 1984. “Genre as Social Action.” *The Quarterly Journal of Speech* 70 (2): 151–67.

Moore, A. 2001. “Categorical Conventions in Music Discourse: Style and Genre.” *Music & Letters* 82 (3): 432–42.

Negus, Keith. 1999. *Music Genres and Corporate Cultures*. London: Routledge.

Petrusich, Amanda. 2021. “Genre Is Disappearing. What Comes Next?” *The New Yorker*, March 4, 2021. https://www.newyorker.com/magazine/2021/03/15/genre-is-disappearing-what-comes-next.

Pons, Jordi, Olga Slizovskaia, Rong Gong, Emilia Gomez, and Xavier Serra. 2017. “Timbre Analysis of Music Audio Signals with Convolutional Neural Networks.” In *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE. https://doi.org/10.23919/eusipco.2017.8081710.

Mcenaney, Tom, and Kaitlyn Todd. 2022. “ARE SPOTIFY’S VIBES THE END OF SEGREGATED LISTENING? (THAT’S NOT WHAT THE DATA SAYS.).” Public Books. October 6, 2022. https://www.publicbooks.org/spotify-vibes-algorithms/.

Rockwell, Joti. 2012. “What Is Bluegrass Anyway? Category Formation, Debate and the Framing of Musical Genre.” *Popular Music* 31 (3): 363–81.

Ross, Alex. 2022. “Reasons to Abandon Spotify That Have Nothing to Do with Joe Rogan.” *The New Yorker*, February 2, 2022. https://www.newyorker.com/culture/cultural-comment/imagine-a-world-without-spotify.

Seaver, Nick. 2021. “Everything Lies in a Space: Cultural Data and Spatial Reality.” *The Journal of the Royal Anthropological Institute* 27 (S1): 43–61.

———. 2022. *Computing Taste : Algorithms and the Makers of Music Recommendation*. Chicago: The University of Chicago Press.

Sparling, Heather. 2008. “Categorically Speaking: Towards a Theory of (Musical) Genre in Cape Breton Gaelic Culture.” *Ethnomusicology* 52 (3): 401–25.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Oxford University Press.

Underwood, Ted, and Richard Jean So. 2021. “Can We Map Culture?” *Journal of Cultural Analytics* 6 (3). https://doi.org/10.22148/001c.24911.

“Vibe, Mood, Energy.” 2022. The Drift. January 19, 2022. https://www.thedriftmag.com/vibe-mood-energy/.