DL838 BA (Hons) Creative Music Production

Professional Project

Cillian Power (N00210328)

Assessing the Feasibility of Data-Driven Production Toolkits for Niche and Emerging Musical Subgenres

Supervised by Conor Brennan & Barry O'Halpin

Declaration
I hereby certify that the material, which I now submit for assessment on the programmes of study leading to the award of BA(Hons) in Creative Music Production, is entirely my own work and has not been taken from the work of others except to the extent that such work has been cited and acknowledged within the text of my own work. No portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification to this or any other institution.

I have used the IADT library referencing guidelines (available at https://iadt.libguides.com/referencing) and/or the appropriate referencing system recommended in the assignment guidelines and/or programme documentation.

I understand that the IADT regards breaches of academic integrity and plagiarism as grave and serious.

By signing this form or by submitting material for assessment online, I confirm that I have read and understood the IADT Academic Integrity Policy (available at https://iadt.ie/wp-content/uploads/2024/03/Academic-Integrity-Policy-V1-2024.pdf).

Cillian Power

Cillian Power 28-05-2024
Signature of Candidate Date

Table of Contents

Acknowledgements

I would like to thank the initial supervisor of this project, Conor Brennan, and the continuing supervisor, Barry O'Halpin.

I would like to thank my friends and family who have supported me entirely and pushed me to do as well as I could have throughout this whole process.

Abstract

This project investigates the feasibility of creating data-driven production toolkits for niche and emerging musical sub-genres. The original aim was to extract human-readable musical features such as tempo, timbre, harmonic structure, and rhythmic patterns from genre-representative tracks and translate these into a practical framework to assist producers. Such a toolkit would ideally support creative exploration by streamlining the process of adapting to unfamiliar stylistic conventions.

However, significant technical and methodological challenges emerged throughout the project. Automated analysis tools proved unreliable or incomplete for many features relevant to music production, particularly in the areas of musical transcription and timbral analysis. The opacity and instability of commercial APIs most notably Spotify's Audio Features further limited the reproducibility and interpretability of the dataset. Although low-level features were successfully extracted in some cases, the results were insufficient to construct a fully functional toolkit. Planned expert validation could not be completed within the available timeframe.

As a result, the project evolved into a critical feasibility study of data-driven genre emulation. It assesses which musical attributes can be meaningfully captured through current MIR (Music Information Retrieval) methods and identifies areas where human expertise remains essential. The dissertation presents a partial prototype informed by extracted features, a curated dataset, and a reflective account of the limitations encountered. In doing so, it highlights the gap between current MIR capabilities and the real-world needs of music producers, and argues for more interpretable, hybrid approaches that combine machine-derived metrics with practitioner insight.

Introduction

The proliferation of music across increasingly niche subgenres has created a unique challenge for producers, knowing how to quickly understand and work within the boundaries of unfamiliar stylistic conventions. Genre-specific production techniques, ranging from chord progressions and rhythmic phrasing to timbral textures and mix decisions, are often difficult to learn without significant listening, experimentation, or direct instruction. In response, this project set out to explore whether a data-driven production toolkit could help streamline this process by using computational methods to extract and formalize the defining characteristics of niche and emerging sub-genres. This project, as such, aims to ask: is it possible to create toolkits for niche and emerging sub-genres of music using a data-driven approach?

The initial aim was to develop a system capable of identifying and organizing both compositional features (such as tempo, key, phrasing, structure, and chord progressions) and sonic features (such as spectral centroid, harmonic-to-inharmonic ratio, tonal stability, and timbre). These features were to be extracted from representative genre tracks using publicly available tools and datasets, including the Spotify Audio Features API and various MIR (Music Information Retrieval) libraries. The ultimate goal was to convert these into a production toolkit; a set of templates or guidelines to support producers in creating music aligned with the target genre.

However, the project encountered multiple obstacles that gradually shifted its direction. Automated tools for music feature extraction proved inconsistent, particularly when analyzing complex or noisy source material. Proprietary systems such as Spotify's API presented issues of data opacity and reliability. The project also encountered limitations in human transcription ability, especially for harmonic and structural information that could not be extracted automatically. These technical and methodological difficulties made it clear that a complete, functioning toolkit would not be achievable within the project scope.

Rather than presenting a finished production system, this dissertation therefore shifts focus toward a critical analysis of feasibility. The research identifies specific pain points in using current MIR technologies for production-oriented tasks, outlines the shortcomings in existing feature sets when applied to real-world music-making, and reflects on the gap between academic MIR research and the needs of practitioners. It draws on adjacent literature in algorithmic composition, symbolic music generation, and playlist analytics to contextualize its findings, and argues that a hybrid approach combining machine-derived features with expert annotation offers a more promising path forward.

For the purposes of this study, musical features were divided into two categories: compositional features, which describe the underlying musical structure (such as tempo, key, mode, phrasing, and chord progressions), and sonic features, which describe the textural and timbral qualities of the sound itself (such as spectral centroid, harmonic ratio, tonal stability, and instrumental timbre). This distinction allowed the analysis to separately address both the musical "blueprint" and the acoustic "surface" of each genre.

This study concludes not with a definitive solution, but with a mapped-out boundary: a clearer understanding of where MIR currently falls short in serving creative production workflows, and what future tools must do differently if they are to bridge the space between data and musical intuition.

Literature Review

Introduction

Given the interdisciplinary nature of the project, this literature review is structured into multiple sections, each addressing a distinct field of pre-existing research that informs and supports the development of this work.

Defining Genre

Franco Fabbri's influential article, "A Theory of Musical Genres: Two Applications," provides a widely accepted definition of musical genre, describing it as "a set of musical events (real or possible) whose course is governed by a definite set of socially accepted rules." A musical event encompasses "any type of activity performed around any type of event involving sound" (Fabbri).

Building on this foundation, Peter van de Merwe, in his book, "Origins of the Popular Style: The Antecedents of Twentieth-Century Popular Music", proposed that genre should be understood as "pieces of music that share a certain style or 'basic musical language'" (Van de Merwe).

These definitions highlight the compositional and sonic factors of music as crucial to genre classification, placing elements such as melody, harmony, rhythm, and timbre at the core of the framework, as opposed to external historical, cultural, or societal factors.

This study adopts the perspective that genre is primarily determined by its audible qualities, focusing exclusively on these sonic factors.

Breaking Down Sonic Factors of Genre

Compositional factors play essential roles in genre classification. These compositional factors are outlined in this section of the literature review.

Tempo often serves as a key distinguishing characteristic of genre. As Rick Snoman explains in Dance Music Manual, “Hip-hop is generally characterized by a slower, more laid-back beat” (Snoman). Dennis DeSantis reinforces this point in Making Music: 74 Creative Strategies for Electronic Music Producers, detailing tempo ranges for various electronic genres, including, for example, 60–90 BPM for dub and 120–140 BPM for techno music (DeSantis).

Time signature also contributes to genre delineation; Snoman further notes that garage music typically employs a 4/4 time signature, while 2-step occasionally uses 3/4 (Snoman).

Harmonic structure is another important factor in genre delineation. Nolan A. Foxworth's thesis observes shared harmonic tendencies across genres, stating that "Rock, Hip-Hop/R&B, and Christian music... all seem to have common harmonic tendencies which allow for improved identification" (Foxworth).

Expanding on this, Cian O'Brien and Alexander Lerch's study, "Genre-Specific Key Profiles" reveals that blues and metal predominantly favor minor modes, while country music leans toward major modes. Although they note that "inter-genre differences are small," their analysis highlights subtle harmonic tendencies that characterize specific genres (O'Brien and Lerch).

Repetition and melodic phrasing further contribute to genre classification. Carmen Ferraro and Kjell Lemström emphasize that electronic music relies heavily on loop-based structures, noting that it "demonstrates a higher degree of repetition compared to other genres, making it distinct in terms of temporal pattern density and compactness" (Ferraro and Lemström).

Rhythm is another important factor in genre delineation, shaping the energy and flow of music. Danielsen et al., in "Shaping Rhythm: Timing and Sound in Five Groove-Based Genres", explore how micro-rhythmic variations, including subtle shifts in timing and dynamics, contribute to genre-specific grooves. For example, while EDM employs precise, loop-based rhythmic patterns that reinforce its driving energy, jazz relies on flexible timing and swing (Danielsen et al.).

In addition to compositional factors, sonic factors, such as textural density, timbre and Layering and textural density assist in the delineation between genres.

Timbral characteristics play an important role in the rapid identification of music genres. Gjerdingen and Perrott, in their study "Scanning the Dial: The Rapid Recognition of Music Genres," emphasize how genre recognition is often quick, occurring within seconds of hearing a song. This rapid identification is largely attributed to timbral cues, described as encompassing "all the spectral and rapid time-domain variability in the acoustic signal" (Gjerdingen and Perrott). These cues are highly indicative of specific genres; for instance, the distinct guitar tone of Country music or the low-frequency emphasis in Hip-Hop tracks serve as defining auditory markers. The study highlights that even short segments, lasting less than a second, are often sufficient for listeners to correctly identify genres, highlighting the importance of timbral information in shaping genre delineation(Gjerdingen and Perrott 98). By analyzing acoustic signals, the authors demonstrate that timbre is not just an aesthetic sonic factor but a fundamental dimension of genre classification.

Extending on the usage of timbre in delineation of genre, music technology such as novel instruments and audio effects can be a defining factor in delineating the sound of a genre. In Buckley's paper, "Technology and the Music Industry: A Study of Effects and Production Techniques Across Genres", the author notes that "distortion amplifies the aggressive tone of electric guitars in rock and metal, while synthesizers and digital sampling are central to the experimental nature of electronic and progressive rock genres. Technologies like Auto-Tune have become emblematic of modern pop, contributing to its polished and futuristic aesthetic." (Buckley). These production techniques illustrate how music technology differentiates and informs genre-specific sounds.

Layering and textural density are significant sonic factors that distinguish musical genres. In electronic dance music (EDM), for instance, layering is used to manage dynamic shifts and build energy. Ferraro and Lemström explain that “there is a temporal alignment between the Sub-Band Flux changes and the perceived textural layers’ onsets and offsets in electronic dance music (EDM),” showing how the interaction of layers creates rhythmic and emotional textures unique to the genre (Ferraro and Lemström). Similarly, in digital pop music, layering extends beyond traditional methods to prioritize "the inter relational control and shaping of the envelope," as the author notes in their study (Morreale et al.). These distinct approaches to layering highlight how different genres leverage the technique to achieve their characteristic soundscapes. Whether emphasizing precise rhythmic alignment or exploring creative envelope manipulation, layering remains a key factor in crafting genre-specific sonic factors.

<u>Timbre</u>

As established above, timbre plays a critical role in genre recognition, often acting as a defining characteristic that allows listeners to distinguish between musical styles even when other elements like tempo or harmony are similar.

However, despite its significance, the deconstruction and faithful recreation of timbre present major challenges. Unlike more easily quantifiable musical attributes, timbre is shaped by a complex interplay of spectral, temporal, and dynamic factors that are difficult to isolate and measure systematically. Numerous studies have emphasized the subjective nature of timbre perception and the current technological limitations in capturing its full nuance, making it one of the most challenging aspects of genre emulation through data-driven methods.

Jiang et al. (2020) present an ambitious attempt to model timbre perception by combining subjective listener evaluations with objective feature extraction from a controlled dataset of musical instrument recordings. Using statistical learning techniques such as linear regression and neural networks, they demonstrate that certain perceptual qualities of timbre, such as brightness or texture can be partially predicted from acoustic features. However, while the study is informative in revealing the multi-dimensional and subjective nature of timbre, it ultimately highlights the difficulty of translating perceptual models into actionable production guidance. The framework remains descriptive rather than prescriptive, offering no direct pathway for reconstructing or synthesizing specific timbres, and thereby reinforcing the broader challenge faced by data-driven approaches to genre emulation.

In *The Perceptual Representation of Timbre*, McAdams (2019) highlights the complex, multidimensional nature of timbre as a perceptual rather than purely physical property. The study shows that timbre perception arises from a fusion of spectral, temporal, and spectrotemporal features, shaped both by acoustic factors and listener experience. Using methods like multidimensional scaling (MDS) of dissimilarity ratings, the paper identifies recurring perceptual dimensions, such as spectral centroid (brightness), attack time, and spectral variation, but also acknowledges significant variability depending on the sound set and listener biases. While the work is foundational for understanding how listeners perceive timbre and for building descriptive models, it remains largely non-actionable for producers or MIR practitioners: it maps perceptual spaces without offering deterministic or prescriptive models for *constructing* or *reproducing* genre-specific timbres in practice. Thus, while informative, it reinforces the view that recreating timbre systematically remains an unresolved challenge.

Despite existing research into the perceptual dimensions of timbre, the reliable systematization and recreation of timbre remains an open problem. Current methods offer descriptive frameworks but fall short of providing producers or researchers with actionable, reproducible tools for generating genre-specific sounds.

Genre's Societal Factors

While sonic factors such as tempo, timbre, and harmonic structure are central to defining musical genres, societal factors play an integral role in their categorization (Fabbri). Genre is not only a product of sonic factors but also a product of collective interpretation, the history of those creating and listening to the music, and cultural context, shaped by the perceptions of listeners.

This underscores the importance of querying user-maintained music databases, such as RateYourMusic.com, which capture the nuanced, crowd-sourced understanding of genre affiliation. Such platforms provide insights into the wide range of genres a song may belong to, revealing how societal and cultural dynamics interact with sonic factors in the formation and evolution of genres.

Machine Analysis

The creation of a production toolkit for niche and emerging subgenres of music additionally requires a comprehensive understanding of the interplay between sonic factors and audience perception.
Key insights for this endeavor can be drawn from research on musical memorability, the dynamics of popular culture, and the psychological responses to musical elements.

In "What Makes Songs Catchy", the author examines how structural and melodic patterns contribute to a song's memorability. The study emphasizes that "hooks, repetition, and structural clarity" are crucial in capturing listener attention and enhancing appeal (Margulis). These features often define subgenre-specific characteristics, making them essential for the toolkit's usefulness.

In "What Makes Popular Culture Popular", the author discusses the idea of "optimal differentiation", proposing that successful songs balance familiarity and novelty. The authors argue that this balance allows products to "stand out enough to be noticed but not so much as to alienate" (Berger, Jonah, Bradlow). This insight is particularly relevant for emerging subgenres, where innovation within genre conventions often drives appeal.
For the production toolkit, this idea offers a framework for guiding producers to produce tracks that introduce distinct and novel elements while still being cohesive to the rest of a genre's norms.

In the article, The authors discuss The Echo Nest as a leading music intelligence provider, acquired by Spotify in 2014, which serves as the backbone for its analytics and recommendation systems. They highlight its extensive dataset, which includes over 30 million songs and 3 million artists, capturing both objective musical features (e.g., tempo, key, and mode) and qualitative data derived from analysis.

In addition, The authors found significant genre-based tendencies in the values measured by Echo Nest (now Spotify)'s musical features, such as acousticness, energy, and danceability. By analyzing within-genre averages for these attributes, they demonstrated that genres exhibit distinct patterns that reflect their sonic factors. They state, "Genres differ systematically in terms of their quantitative acoustic features," with each genre presenting a unique combination of values that influence how listeners perceive them (Berger, Jonah, Bradlow).

For instance, high energy and danceability are often associated with electronic and pop music, whereas acousticness tends to characterize folk and classical genres. The authors further note, "Cosine similarity measures show that songs within a genre cluster more closely together based on these features than songs across genres," illustrating that these quantitative traits are not only indicative of a genre but also help define the boundaries between them.

These findings highlight the importance of measurable sonic factors in distinguishing genres, reinforcing their applicability in tools like recommendation systems or a production toolkit tailored to a specific genre. These analyses provide support for using parameters like "danceability", "energy", and "acousticness" as important elements in identifying and replicating the sound profiles of niche and emerging subgenres.

Together, these studies provide a multidimensional understanding of the factors contributing to musical appeal and genre differentiation. By obtaining insights from these works, the production toolkit can integrate evidence-based features that support creative exploration while maintaining genre cohesion, ensuring its relevance and usability for producers in niche musical contexts.

MIR Software

Music information retrieval software is essential to the creation of music toolkits.

Over the past two decades, Music Information Retrieval (MIR) has evolved significantly, enabling new approaches to musical analysis, recommendation systems, symbolic generation, and even compositional assistance. However, much of this progress has been directed toward music consumption, classification, and symbolic modeling, with less emphasis on production-facing workflows. The application of MIR in producer-centric workflows, particularly for the construction of data-driven production toolkits in niche and emerging genres remains severely underdeveloped. This gap becomes especially apparent when reviewing adjacent studies that have successfully applied MIR in more consumption- or composition-oriented contexts.

Engart (2019) demonstrated the creative use of MIR in algorithmic composition, using features like spectral centroid and MFCCs not to reproduce known musical forms, but to sculpt gestural arcs and structural templates for experimental fixed-media works. While his outcomes were creative rather than replicative, his approach shares key methodological DNA with this study, namely a reliance on MIR features to inform and guide creative output. Crucially, Engart retained a human-in-the-loop model, acknowledging that MIR could assist but not replace compositional intent. His use of MIR as a search and structuring tool, rather than a blueprint, reflects a pragmatic compromise that contrasts sharply with the more generative ambitions of this dissertation.

Cella (2021) pushes this reflection further in Music Information Retrieval and Contemporary Classical Music: A Successful Failure. Over 15 years, Cella explored the expressive use of MIR and machine learning in composition, only to conclude that existing tools remain poorly suited for advanced artistic control particularly in styles where timbre and gesture are central. Like this project, Cella frames his work as a productive failure, identifying a critical mismatch between available MIR tools and the nuanced demands of contemporary music making. His call for hybrid systems that blend logical structure with machine learning echoes this study's eventual pivot away from a purely data-driven toolkit toward a reflective critique of MIR's production blind spots.

Meanwhile, large-scale MIR research continues to thrive in consumption-oriented domains. Gabbolini and Bridge (2024) provide a comprehensive survey of MIR research applied to playlist creation and enhancement. Their work highlights how features like tempo, timbre, and metadata are widely used in systems optimized for streaming platforms. These efforts are driven by personalization, coherence, and user engagement metrics, goals that contrast sharply with those of the music producer. Similarly, Gorgoglione et al. (2023) applied MIR and Big Data analytics to study innovation in the music of the Red Hot Chili Peppers. Their use of features like spectral centroid and danceability to measure stylistic change demonstrates the usefulness of MIR in structured, artist-centric analysis where the dataset is large, coherent, and well-archived.

By contrast, symbolic music generation has taken a different path. Le et al. (2024) detail how Natural Language Processing (NLP) models, particularly Transformers, have been applied to symbolic data such as MIDI files. In this study, music is tokenized and treated like language, enabling melody generation, harmonization, and style transfer. These models benefit from the structured and consistent nature of symbolic music representation, bypassing many of the difficulties associated with audio-based MIR.

While these studies represent significant progress, they all operate in domains with either structured data that is not always helpful to humans (symbolic formats) or music consumer interactions (playlists, mainstream discographies). Very few works have attempted to apply MIR to aid producers to work within given genres, where structural clarity, interpretability, and genre-specific production conventions are not easily extracted from audio features.

Taken together, these studies show that MIR has enabled significant advances in consumption, classification, and symbolic generation, but that its tools and assumptions are poorly suited for producer-facing applications, especially in genres where intuition, timbre, and structure resist quantification. This project, like that of Cella, becomes a diagnostic case study in what MIR cannot yet do: support the creation of practical, automated toolkits that are generative, reproducible, and musically meaningful to producers without substantial human adaptation. It is here that future work must intervene, not by abandoning MIR, but by reorienting it toward the needs of music makers, not just music consumers.

Methodology

Overview

The initial ambition of this project was to develop production toolkits for niche and emerging sub-genres of music through a data-driven methodology. This was attempted by way of machine-extracted sonic features from audio tracks belonging to a target genre. The underlying principle was to streamline the process for producers to familiarize themselves with the characteristics of unfamiliar musical styles by providing quantifiable metrics, thereby reducing the reliance on manual analysis and intuitive guesswork.

This involved gathering human-readable and quantifiable data points related to both compositional elements, such as tempo, key, chords, and structure, and sonic factors, including timbre, spectral content, and tonal stability.

Over time, the methodology evolved into a mixed approach, combining quantitative analysis, prototype creation, and reflective evaluation due to limitations in tool reliability, data quality, and practical constraints.

The intended outcome was a toolkit that would enable producers to create music conforming to a specific genre, validated through feedback from expert listeners.

However, the research process encountered a series of significant and unforeseen challenges, spanning from difficulties in acquiring the necessary data to limitations in translating this data into a practical toolkit, as well as the disruptive impact of external factors.

This section serves to introduce these limitations and to establish the framework for a subsequent detailed discussion on how these challenges, rather than hindering the academic value of the work, can be reframed as critical insights into the complexities of data-driven music technology research.

Data Collection

Dataset Selection

To examine the feasibility of constructing data-driven production toolkits for niche musical styles, a representative dataset of tracks was compiled for two emerging sub-genres: Asian Rock (a derivative of Pluggnb) and Neo-Grime (also referred to as UK Wave). Track selection was based on both popularity and critical acclaim to ensure cultural relevance and stylistic representativeness.

Tracks were sourced from RateYourMusic.com, a large, community-curated music database that allows users to tag, rate, and review songs. For each genre, the top 12 tracks were selected based on user-generated ratings and the quantity and quality of written reviews under the relevant genre tag. This approach prioritized tracks that had not only been frequently listened to but also widely discussed and endorsed within their respective online music communities. The resulting dataset comprised 24 tracks in total, 12 per genre, which served as the foundation for all subsequent analysis.

Genre Context

The two genres selected for analysis reflect highly specific yet increasingly visible stylistic developments within underground and internet-native music communities:

The first genre picked is called Asian Rock. This musical genre originated from PluggnB, itself a sub-variant of Plugg, a genre rooted in American trap music. The genre can be "characterized by its distinct production style which features dense snare and percussion rolls in combination with airy bells, synth plucks, and soaring synthetic guitar leads" (RateYourMusic)

Neo-Grime is an evolution of UK Bass music, Grime, and Wave, merging grime's percussive aggression with the atmospheric depth of Wave. The genre can be "characterized by lush, intricate, and icey synthwork derived from wave, which is often coupled with nocturnal and futuristic visuals. These synths are combined with major grime elements, like synth square waves typical of 2010s instrumental grime production, and with syncopated, triplet rhythms borrowed from grime".(RateYourMusic)

These two genres were chosen due to their distinct musical lineages and lack of mainstream popularity. Additionally, they present difficulty in reducing to data-derived features that can capture subtle genre markers.

Digital copies of the tracks were primarily obtained from digital storefronts such as Qobuz or Bandcamp, in order to preserve audio fidelity and mix characteristics that may be altered by a streaming services' unknown audio normalisation or audio downsampling procedure.

One exception to this is Red Eyez by Lexxi, a song in the Neo-Grime dataset, which was unable to be found on digital storefronts at the time this study was conducted. As a result, a version found on YouTube was analysed instead. While this audio file was certainly lower fidelity, it was reasoned that this song could still be accurately assessed for compositional features.

One song in the Asian Rock dataset, “Jeans Soaked” by Lil Shine, includes a 2 minute long intro consisting of a sampled recording of classical music. For the purposes of this study, this was cut for the MIR and compositional analysis as this intro is not seen in any of the other songs, and is believed by the researcher to serve more as an interlude within the context of the album the song comes from, than an incorporated part of the song.

Source Tools and Workflow Changes

Initially, the choice of songs included in the dataset was intended to be both analysed and partially sourced using the Spotify Web API, particularly its Audio Features and Audio Analysis endpoints. These endpoints provided access to a variety of track-level descriptors such as tempo, key, energy, danceability, valence, and timbral vectors derived from internal MIR models. These values were to be used for pre-selection, analysis, and eventual comparison against custom-extracted features from local tools.

However, on November 27, 2024, Spotify unexpectedly deprecated its Web API services for both Audio Features and Audio Analysis. This major change disrupted the planned workflow, removing access to previously available descriptors and necessitating a pivot in both data sourcing and analysis methodology.

As a result, track selection was done entirely using RateYourMusic rather than incorporating Spotify metadata. MIR features then had to be extracted using local and open-source tools. This was done by writing a python script that used Essentia and Librosa to extract various sonic features from the downloaded songs. In addition to this, any planned comparisons between or refinements including Spotify-derived Audio Features and manually extracted descriptors were removed from the study scope.

Tools Used

The data analysis phase of this project relied on a combination of automated MIR tools, platform-specific APIs, and manual annotation, with the initial assumption that current technologies could reliably extract compositional and sonic features from audio recordings. However, the tools presented a mix of strengths and limitations. While the researcher found that tempo and key detection software performed with good accuracy, tools designed to transcribe chords and extract harmonic or timbral nuance often produced inconsistent results, especially when analyzing timbrally or musically complex material, such as the musical genres chosen as case study genres.

These challenges reflect broader systemic issues in empirical music research, where data-handling practices are still evolving, and technological infrastructures remain poorly suited to extracting the deeper features that shape genre identity and creative decision-making.

MIR Tools

Several open-source tools were used for extracting features from audio files.

Sonic Visualiser was used primarily for visual inspection of waveforms, spectrograms, harmonic content, and tempo estimation. Its plugin-based architecture (e.g., Vamp plugins) allowed modular analysis but lacked automation and reproducibility at scale.
Essentia is a comprehensive C++/Python library for audio feature extraction, used in this study for measuring low-level features such as spectral centroid, pitch, inharmonicity, and harmonic ratio. While robust, the figures produced were ambiguous as to how they were calculated.

Librosa is a Python-based toolkit that provides spectral analysis, MFCC extraction, chroma features, and structural segmentation. It enabled some degree of feature mapping but required custom scripts and manual validation.

Despite their modularity and open access, these tools were not capable of extracting higher-level musical parameters such as chord progression, phrasing, or mix structure, with sufficient reliability. In particular, chord recognition algorithms frequently fail when confronted with genre-typical harmonic ambiguity or non-standard tunings.

Spotify Web API

Initially, a large portion of the analytical pipeline relied on Spotify's Audio Features and Audio Analysis endpoints. Each of the tracks' reported tempo, key, and mode values were assessed. In addition, the tracks were assessed for their Spotify proprietary audio features such as "energy", "valence", "danceability", "acousticness", "instrumentalness", "speechiness", and "liveness" values.

These features were intended to guide track selection and provide coarse descriptors for track selection. These descriptors, however, offer broad sonic summaries, often optimized for recommendation systems rather than production insight. There was also limited transparency in how metrics like "energy" or "danceability" were computed, making them difficult to interpret or trust in a production context.

This misalignment between available data and practical production relevance revealed a deeper issue: the gap between MIR designed for classification and the needs of creators, who require information about how something was created, and not just what it sounds like.

Manual Annotation

Due to the aforementioned limitations, manual listening and annotation became necessary. These included verifying or correcting automated key and tempo estimations, attempting to transcribe chord progressions by ear when automated tools failed, and estimating musical phrasing. Common types of timbre were also noted, but not systematised.

This hybrid approach was time-intensive and limited in scale, but it highlighted the continued importance of human musical expertise, especially when dealing with emerging genres that lie outside the training scope of most MIR models.

Feature Extraction

The feature extraction phase of this project sought to identify both compositional and sonic features that define the two case study genres. This involved a combination of automated analysis, critical listening, and manual annotation, with the goal of identifying recurring characteristics that could form the basis of a production toolkit.

While low-level features such as tempo, song length, and mode showed some consistency across tracks, deeper stylistic markers proved elusive. Many of the most distinctive features, particularly those relating to timbre, harmonic ambiguity, and unconventional cadences could not be captured reliably through automated means and required manual interpretation.

Compositional Features

Automated tools were used to extract basic compositional data, followed by manual verification where needed. Tempo was manually extracted, along with phrasing and song length information. Key and mode were identified via both Spotify's (now deprecated) API and local tools like Essentia and Antares Auto-Key. Basic major/minor mode was usually identifiable, but tracks with modal interchange or ambiguous tonality returned unreliable results. Manual adjustment was sometimes necessary.

Chord estimation and transcription was attempted with Sonic Visualiser and chord detection vamp plugins such as Chordino, as well as the Ableton Audio-to-MIDI feature. However, results were inconsistent, especially in tracks with non-diatonic harmony, extensive effects processing, or unconventional voicing. As such, this data was treated as tentative at best.

Sonic Features

To analyze the sonic features of the tracks in the dataset, a custom Python script was developed using established audio analysis libraries, including Librosa and Essentia. The goal was to extract features that could contribute to understanding the overall sound character of different genres.

First, the songs were organised into numbered lists, divided by genre, each including a 10 second long sample of white noise for control purposes. Using a python script, the songs were each assessed for tonal stability, spectral centroid, and harmonic ratio across the length of the whole song.

Harmonic ratio measures how much of the track's energy is harmonic (i.e., pitched and tonal) compared to noisy or percussive elements. Tracks with higher harmonic ratios tend to sound more melodic and stable, while tracks with lower ratios may sound more noisy or chaotic. The harmonic energy and percussive energy were separated, and their proportions were calculated. The resulting ratio (originally a value between 0 and 1) was scaled onto a 0–100 range to make it more interpretable, where 0 indicates predominantly noisy sounds (calibrated against white noise sample) and 100 indicates highly harmonic content.

Tonal stability assesses how consistently a track holds its pitch over time. Using pitch detection methods, the script analyzed small segments of audio to estimate the fundamental frequency. The degree of pitch fluctuation (measured as variance) was used as an indicator of tonal instability. This raw pitch variance, originally measured in Hertz squared, was similarly scaled to fit within a 0–100 range, where higher values represent greater pitch stability.

To improve the usefulness of spectral centroid measurements, an additional Python script was developed. In the original full-spectrum analysis, deep bass frequencies often dominated the centroid calculation, resulting in misleadingly low values that primarily reflected sub-bass content rather than the overall brightness of the track. To address this, the script applied a high-pass filter at 100 Hz to each track and re-calculated the spectral centroid on the filtered signal. This secondary measurement provided a more informative indication of midrange and high-frequency content, offering a better reflection of the perceived brightness important for genre characterization.

Spectral centroid was extracted using "librosa.feature.spectral_centroid", and used as a proxy for perceived brightness or timbral sharpness of a mix. Higher spectral centroid values, measured in Hertz as being where the audio was the highest in volume, generally correspond to brighter, sharper sounds, while lower values are associated with darker tonal qualities.

These features were chosen because they offer a low-level, broadly interpretable description of the textural and tonal characteristics of music without requiring full musical transcription.

However, it is important to note that while these measurements provided some useful insights, they capture only a small part of what makes a track identifiable within a genre. As discussed later, deeper qualities such as timbre and expressive performance details remained outside the reach of this automated extraction process.

All of these sonic feature analyses aimed towards finding some way of informing the recreation of mix and instrument timbre. Timbre resulted in being the most important yet least systematisable feature. Attempts were made to approximate timbre for recreation based on critical listening and the abovementioned MIR analysis.

Despite these efforts, timbre could not be meaningfully reduced to a quantifiable feature. Its subjective and context-dependent nature resisted standardization and further emphasized the limitations of purely data-driven approaches in production-oriented research.

Reflections

Ultimately, only a subset of features, like tempo, phrasing length, mode, and track length exhibited consistent enough patterns to inform actionable rules for a production toolkit. More nuanced aspects, such as accidentals, cadence types, and timbre, were identifiable through listening but not extractable via automation. This mismatch between what can be measured and what defines a genre underlines the core problem of this project: the expressive dimensions of music often escape quantification, particularly in styles that rely on subtle inflection, sonic texture, and aesthetic convention more than formal structure.

Toolkit Construction Attempt

The final phase of the project aimed to synthesize the insights gathered through feature extraction into a practical production toolkit for each genre. This toolkit was intended to act as a reference framework for music producers, enabling them to reproduce or adapt to unfamiliar styles with minimal trial and error. By translating extracted features into actionable production parameters, the toolkit was envisioned as a bridge between empirical analysis and creative application.

Translating Features into Toolkit Components

The initial concept for the toolkit involved creating a genre blueprint consisting of an excel reference sheet outlining the typical values or characteristics for each extracted parameter, such as tempo range, key/mode tendencies, and phrasing length. Additionally, synth patch re-creations made in Serum that emulated genre-signature timbres identified during listening and analysis were intended to be created as part of the toolkit.

Each component was derived from the extracted dataset, with the goal of offering clear entry points for producers to replicate stylistic markers without as extensive listening or genre immersion.

Obstacles and Failure Points

While several low-level features were successfully compiled and included in the prototype toolkit, key limitations emerged that ultimately restricted the toolkit's completeness and usability.

Firstly, timbre descriptions were too vague. Although timbre was identified as central to genre identity, it proved difficult to describe in actionable or reproducible terms. Attempts to recreate genre-signature sounds using Serum were inconsistent, and no standard taxonomy existed for describing or categorizing the necessary timbral nuances. Efforts to standardize this into usable presets were therefore incomplete and highly subjective.

Mix and effects guidance was incomplete and unimplemented. The toolkit lacked sufficient depth in areas such as EQ profiles, spatial effects, compression chains, and loudness norms, elements that strongly influence the feel of a track but are not easily extracted or summarized from raw audio. Without a detailed mix analysis pipeline (which proved too time-consuming to implement), this section of the toolkit remained underdeveloped.

Due to the poor performance of chord recognition tools and the complex, often ambiguous harmony used in the genres, chordal and progression guidance could not be generalized. While anecdotal observations were made (e.g., use of accidentals or unexpected cadences), they could not be formalized into a system producers could follow.

Prototype / Micro-Toolkit

As a result of these limitations, the final output was delivered as a prototype or micro-toolkit, containing only a basic reference sheet for each genre with tempo range, track length, phrasing conventions, and modal tendencies. Various low level sonic features such as harmonic ratio and whole-song spectral centroid were included, but of very limited usefulness.

This partial toolkit serves more as a proof-of-concept than a finished product. It demonstrates which elements of a genre can be described and automated using current MIR tools, and more importantly, which cannot offer practical insight into the boundaries of data-driven production support.

In doing so, it reinforces that human listening, interpretation, and creative intuition remain irreplaceable components of genre emulation, and that MIR-based approaches will require substantial refinement, particularly in the domains of timbre representation, harmonic description, and production context before they can offer reliable assistance to producers working in emergent styles.

Music Production Experiment

Following the feature extraction phase, a music production experiment was conducted in an attempt to create example tracks using the preliminary toolkit. However, due to the lack of detailed and reliable data, the production process ultimately relied heavily on further rounds of critical listening and trial-and-error experimentation. Several attempts were made to produce genre-representative tracks, including additional efforts informed by external resources such as YouTube tutorials on producing within the target styles. Despite these efforts, none of the resulting tracks reached a level of genre fidelity or production quality sufficient for expert listener evaluation. As such, the validation phase planned for the project could not be completed. description of tracks made using (or attempting to use) the toolkit.

Expert Listener Validation

Intended Purpose and Methodology

Validation by subject matter experts was initially planned as a key component of this project. The intent was to evaluate the perceived authenticity of music produced using the prototype toolkit by consulting both expert listeners and non-expert audiences. Their feedback would help assess whether the toolkit was successful in capturing the defining elements of the target genres and whether the generated track "sounded" like it belonged to the target genre.

Expert validation was to be conducted in two phases.

For expert listener feedback, a track produced using the toolkit would be submitted to one or more genre specialists, identified through their curation and review activity on platforms such as editors of genre pages on the website RateYourMusic.com, whose contributions reflect deep familiarity with stylistic boundaries, were chosen as ideal candidates. Their listening histories and written reviews were manually reviewed to ensure alignment with the target genre.

Each expert would have been invited to participate in a feedback session (written or informal interview), where they could respond to open-ended prompts such as “Does this track conform to genre expectations?”, “Are there genre-specific elements that are missing or misrepresented?”, and whether particular features of the song felt authentic or out of place.

A follow-up survey would be conducted among general, non-expert listeners, presenting them with the toolkit-generated track alongside a curated set of real genre examples. Participants would have been asked “Which track(s) do you feel belong to the same genre?”, “Does the example track sound out of place?”, and “What elements stood out as similar or different across the tracks?”. The purpose of this stage was to test broader audience perception and triangulate expert feedback with intuitive genre recognition from unfamiliar listeners.

Although the validation stage was not completed, its planned structure highlights the importance of cultural and perceptual input in data-driven music production research. The hypothetical feedback process would have provided essential contextual depth to the quantitative feature extraction, helping distinguish which musical markers are meaningful to human listeners and which are merely surface-level descriptors.

The omission of this validation component ultimately reinforces a key finding: that data can suggest patterns, but genre recognition and musical authenticity remain deeply human processes that require qualitative engagement.

Analysis

Stated Results

The feature extraction process across 24 tracks yielded consistent results for basic compositional elements, such as tempo, phrasing length, track duration, and mode. However, more complex musical and sonic features, including chord progressions, timbral profiles, and production techniques could not be extracted or formalized with sufficient accuracy. A partial prototype toolkit was developed, consisting of a PDF reference sheet.

Planned expert validation was not completed due to time constraints, and the deprecation of Spotify's Web API significantly disrupted the automated data acquisition process.

Interpretation of Results

The results suggest that data-driven production toolkits are only partially feasible using current tools and methodologies. While some structural and tempo-related traits of a genre can be reliably captured, the more genre-defining characteristics such as timbre, harmonic nuance, and mix style remain outside the scope of current MIR automation. Additionally, the project's challenges with API reliance, tool transparency, and human transcription limitations reflect broader systemic issues in empirical music production research.

As methodological challenges accumulated, it became clear that a fully data-driven, production-ready toolkit was not achievable within the constraints of this project. Instead, the project pivoted toward evaluating the feasibility of such toolkits more generally, using the partial outputs and limitations encountered as case study material.

What Worked

Some aspects of the project were successful. The extraction of low-level sonic features such as spectral centroid, harmonic ratio and tonal stability using Librosa and Essentia functioned reliably. The extraction of simple compositional features with automated tools also worked reliably. The use of RateYourMusic as a resource to build a curated dataset of genre-representative tracks proved effective. Preliminary efforts to structure genre-specific reference sheets based on recurring compositional patterns provided a useful foundation for toolkit development. Additionally, the project successfully identified key methodological risks that impact data-driven creative research, including API deprecation, tool fragility, and the challenges posed by tacit, non-quantifiable musical knowledge.

What Did Not Work

Following the deprecation of Spotify’s API, the study was forced to shift entirely to local analysis using MIR libraries such as Essentia and Librosa. This change also necessitated the abandonment of several planned API-based comparisons and statistical mapping tasks, redirecting research time toward manual annotation and feature extraction. Efforts to model timbre proved too complex to systematise, and no viable analytical framework was found for capturing or reproducing a genre’s signature timbres. Chord recognition and harmonic analysis tools produced low-accuracy results, particularly in tracks exhibiting modal ambiguity or significant audio processing. The planned expert validation phase could not be conducted, limiting the project’s ability to assess the perceptual accuracy of the toolkit outputs. Additionally, the resulting toolkit remained incomplete, lacking essential production-layer components such as mixing guidance, instrument timbre replication, and detailed stylistic recommendations necessary for full genre emulation.

Discussion

This chapter reflects critically on the methodological choices, results, and limitations encountered during the project. It also interprets the findings in the context of the original research aims, to explore whether a data-driven production toolkit could meaningfully support producers in replicating emerging musical sub-genres.

Interpretation of results

Substantial effort was dedicated to gathering, analyzing, and structuring musical feature data from two niche genres: Asian Rock and Neo-Grime. Despite careful dataset selection and the use of multiple Music Information Retrieval (MIR) tools, several limitations constrained the project's outcomes.

Primarily, the data extracted was not sufficiently precise or detailed to support comprehensive toolkit construction. Low-level features such as tempo, phrasing length, and mode were extracted reliably. However, critical genre-defining aspects like timbre nuance, harmonic ambiguity, and production aesthetics, were unable to be assessed by available tools.

Technical descriptors such as spectral centroid or harmonic ratio, while technically descriptive, lacked intuitive interpretability for a music producer aiming to recreate genre-specific sounds.

Furthermore, the researcher's own musicianship limitations, particularly in transcribing chord progressions and melodic structures, compounded the challenges where automated tools fell short.

It should be noted that not all observed features are of equal importance across genres. Features with a wide range of variance demonstrate that these features may be able to be freely changed in a prospective work without affecting genre recognisability. Features with a narrow range of variance demonstrate that these features may not be able to move from the median too far and remain recognisable as the same genre.

When data was gathered successfully, it failed to independently guide authentic genre replication. Attempting production of tracks based solely on extracted features revealed substantial gaps; stylistic coherence still relied heavily on external resources, including critical listening, genre familiarity, and practical experimentation, and supporting tutorial material from the music-listening community and existing tutorials.

This is interpreted to mean that key genre-defining elements were either unrepresented or too loosely described in the collected data.

These limitations underscore the central finding of the project: that current MIR-driven approaches, when applied in isolation, are inadequate for fully capturing and analysing musical style in a production context. Human-centered strategies that are difficult to quantify, such as listening, interpretation, and iterative experimentation, all remain essential components to any music production workflow, including one that may be more data-driven. While MIR can assist in mapping broad genre tendencies, the expressive and aesthetic decisions that define a genre remain outside the reach of existing automated systems.

This finding aligns with previous critical work in the field, such as that by Cella (2021), who identified the gap between machine-extracted features and meaningful artistic control, and by Engart (2019), who emphasized the need for human mediation even when algorithmic processes are used in compositional workflows.

Significance of Findings

Although this project did not result in a complete and functioning data-driven production toolkit, the process revealed significant insights about the practical, technical, and conceptual limitations of using MIR tools in music production workflows, particularly for niche or emerging genres.

One of the most important findings was the disconnect between available MIR features and what producers actually need to create stylistically accurate music in a given genre. Basic compositional and sonic features such as tempo, key, and spectral centroid were easy to extract and useful for basic characterization, but many critical, genre-specific sonic factors such as instrument timbre, articulation and overall mix profile proved much harder to quantify in any way that could inform production decisions. Similarly, genre-specific compositional factors were unable to be assessed in a meaningful way.

This project’s prototype development process demonstrated that automated analysis alone is currently insufficient for capturing the tacit knowledge embedded in genre practice. Without a strong level of musicianship or embedded domain expertise, interpreting extracted features and turning them into actionable production strategies becomes extremely challenging.

The difficulty in creating a toolkit was compounded by the instability and opacity of external data sources, such as Spotify's Audio Features API, which unexpectedly changed during the project timeline. The deprecation of tools and lack of transparency in how features are calculated weakened their usefulness as reproducible production parameters. These practical challenges illustrate the fragility of relying on non-transparent systems for research infrastructure.

This project serves to highlight the underexplored space between MIR and music production. While much MIR research has focused on applications surrounding the consumption and categorisation of music, such as playlisting, tagging, and recommendations, little work has been done to develop tools that directly assist human music producers in understanding and replicating genre conventions based on audio data. This clearly marks a gap in academic research and software development around the integration of MIR technology and music production.

In this sense, the findings of this study are significant not as a proof-of-concept for a finished system, but as a critical feasibility study that assesses the conditions under which data-driven production toolkits might be possible, and the barriers that currently prevent their success. These insights form the foundation for a new line of inquiry into what future music production tools could look like.

Comparison with previous studies

This project aligns most closely with the practice-led investigations by Engart (2019) and Cella (2021), both of whom explored the creative potential and limitations of MIR for music making.

Engart used MIR as a search and structuring tool for algorithmic composition, employing features like spectral centroid and MFCCs to inform gestural evolution and sound organization. While his aim was not to replicate a genre, his process of working with large datasets and algorithmically organizing them for creative output parallels the early intentions of this study. Importantly, Engart maintained human oversight and acknowledged the interpretive limitations of MIR features.

Cella's work presents perhaps the most philosophically aligned precedent. After 15 years using MIR and ML tools in contemporary classical composition, he concluded that most MIR systems are poorly equipped to deal with timbral, structural, or non-Western musical logic. His paper, A Successful Failure, reframes technological shortcoming as insight, arguing that truly effective music-making tools must blend logic, learning, and intuition. This echoes the trajectory of this project: an initial attempt at automation that gave way to a diagnostic understanding of MIR’s current production limits.

Unlike Engart and Cella, this project was framed from a producer’s perspective and focused on genre emulation and stylistic support, not experimental composition. This difference in audience and application brings new challenges particularly in trying to formalize the tacit, often intuitive knowledge that producers rely on. The failure to construct a reliable, data-driven production toolkit becomes a critical contribution: a practical demonstration of where MIR research must expand to serve the needs of music makers, not just music consumers or analysts.

Limitations and Implications

The limitations of this study are extensive in the respect that it did not achieve what it set out to do by making a successful toolkit, but it does lay out some foundational aspects of what is needed for future attempts at data-driven music production

Methodological Limitations

In this study, not all musically relevant data was observed. This is partially because audio feature extraction software/MIR is not currently advanced enough to reliably extract all features from certain genres accurately. This leaves a lot of the compositional analysis to be done manually, leaving ample room for error depending on individual musical skill.

Due to time constraints and the deprecation of the Spotify API, which significantly delayed data processing and diverted attention to methodological restructuring, this validation process could not be implemented. As a result, no formal expert feedback or audience testing was conducted within the project timeframe.

An additional limitation encountered during the project was scope creep related to programming tasks. Significant time was diverted into refining custom scripts for audio feature extraction and experimenting with different analysis methods, often beyond what was necessary for the immediate goals of the study. This reduced the time available for other critical phases, such as production and validation, and contributed to the incompleteness of the final outputs.

Compositional Analysis

A critical reflection on the research process revealed that the researcher's own level of musicianship may have significantly impacted the accuracy of the transcribed compositional data. The process of manually transcribing musical information, such as chord progressions and melodic lines, requires a high degree of musical skill and aural acuity that the researcher may not have been able to reach. This slowed down project progression immensely as this had to be repeated several times, and has likely also impacted the reliability and accuracy of the data, and therefore the resulting toolkit.

This limitation is important to acknowledge as it underscores the inherent challenges of conducting comprehensive music analysis without a strong musical background, or sufficiently accurate and robust audio feature extraction tools.
It is the hope of the researcher that MIR software becomes more advanced in future to enable the creation of truly comprehensive toolkits without requiring as extensive of a musical background that is currently required.

Sonic Analysis

The difficulty in comprehensively assessing and reproducing sonic factors, particularly timbre, which is an abstract and multifaceted property of sound, presents a significant challenge in data-driven music production. Much of the unique "sound" of a genre is deeply rooted in subtle timbral characteristics that proved to be prohibitively difficult to quantify and translate into actionable production advice.

Description of instrument timbre proved particularly challenging. The researcher's best attempt involved trying to recreate sounds typical of the target genre using virtual instrument presets in Serum, a widely used and provably versatile wavetable software synthesizer.

Toolkit Creation

Once the requisite amount of data had been collected, the subsequent challenge lay in translating these metrics into a comprehensive and practically useful production toolkit. This process revealed significant gaps in the research regarding the nuanced aspects of sonic features and musical composition factors that individually contribute to the identity of specific music sub-genres.

The initial compositional analysis focused on factors such as the length of phrasing, tempo, key, mode, overall song length, chord progressions, sections, and structural patterns. However, through experimentation and further reflection, it became evident that crucial compositional elements were not adequately assessed.

These included cadences, typical melodic intervals, the use of passing notes and accidentals, the presence of microtonality, the relationship between bass lines and top lines accounting for the underlying or implied chords, and rhythmic complexities. The sheer number of potentially relevant musical parameters (each genre having its own set of relevant parameters) underscores the inherent complexity of music analysis. The omission of these factors from the initial assessment highlights the difficulty of creating a truly comprehensive toolkit based solely on a limited set of easily quantifiable compositional features.

Toolkit Utilisation

A crucial turning point in the research came with the experimentation phase, where the researcher attempted to create music within the target genre without explicitly utilizing the data-derived toolkit, instead, relying on pure listenership and familiarity with the genre.

This practical exercise revealed a significant number of missing elements that were not effectively captured by the initial data analysis and, more importantly, that the researcher lacked the means to accurately quantify.

However, this process highlighted the difficulty of articulating and replicating the subtle characteristics of timbre through purely technical means. Furthermore, conducting detailed mix analysis for the entire dataset proved to be prohibitively time-consuming and would have necessitated an unacceptable reduction in the already limited dataset size.

This experience underscored the inherent limitations of relying solely on quantifiable data to capture the more intuitive and experience-based aspects of music production, where tacit knowledge plays a crucial role.

Time Management and Research Focus Issues

Finally, the researcher acknowledges significant issues with time management and research focus. A considerable amount of time was spent pursuing tangential research avenues, including attempting to reverse-engineer or approximate the algorithms behind Spotify's "Audio Features" and trying to understand how music generation AI programs analyze and recreate music (in hopes of applying these methods to present data in a human-readable way).

While these were intellectually stimulating pursuits, they ultimately detracted from the core tasks of gathering data and producing music using the intended toolkit. This reflects a common challenge in research, particularly in interdisciplinary fields, where the breadth of potential inquiry can lead to scope creep and a diversion from the primary research objectives.

One major limitation of this study was the inability to produce finished musical examples based on the prototype toolkit. The incompleteness and imprecision of the extracted data, particularly regarding timbre and compositional nuances, meant that no sufficiently genre-accurate tracks were completed to a standard appropriate for expert validation. As a result, the intended feedback phase with expert listeners could not be realized, further restricting the empirical assessment of the toolkit's effectiveness and highlighting the gap between theoretical analysis and practical music production outcomes.

Future Research

Research into the continued development of Music Information Retrieval (MIR) software would benefit the creation of data-driven production toolkits immensely. Music recognition software should aim to identify and transcribe melodies and chords with greater accuracy. Although existing tools can accomplish this to an extent, there remains a significant amount of noise and inaccuracy in the data produced, particularly when analyzing complex or heavily processed musical content.

The second area concerns future researchers who attempt to create production toolkits or streamline music production practices for unfamiliar genres. One potentially promising direction would be to seek to quantify and measure a broader range of compositional factors that were not evaluated in this study. From a melodic and harmonic perspective, future inquiries could include the identification of common interval patterns, the characterization of typical melodic shapes (e.g., rising versus falling phrases), the analysis of the use of accidentals and passing notes, the cataloging of typical cadences, and the exploration of modal mixes or changes. Researchers should also examine the interaction between top lines and underlying chords, determine the behavior of basslines (such as whether they double the root, follow independent motion, or imply alternate harmonies), and assess the presence of microtonality where applicable.

Rhythmic aspects could also be systematically studied, including typical rhythmic subdivision patterns, common syncopation types, preferred drum groove styles, and the use or absence of tempo changes.

Structurally, researchers could focus on identifying standard song structures (such as ABCABC, ABBA, ABAB, etc.) and detailing the typical length and content of each section, for instance, whether there is a sparse introductory section, a rhythmically altered outro, or other potential traits for a given dataset.

In addition to compositional features, unevaluated sonic factors also present opportunities for deeper analysis. Future studies could attempt to systematically document instrument timbre, per-section or whole-song spectral content, stylistic ornamentation (such as vibrato, legato gliding, etc.), the timbre of sampled sounds like drums, loudness levels between sections of a song (including PLR, integrated loudness, and short-term loudness), and the general use of production effects such as reverb, compression, saturation, distortion, delay, and panning. Further attention to and systematisation of general mixing tendencies across genres could significantly enhance the utility of future toolkits.

A key recommendation for future researchers is the development of timbre-specific analytical frameworks. Timbre emerged during this project as the most critical yet least systematizable dimension of genre identity.

Future research would benefit from the prioritization of the design of new models or taxonomies specifically focused on timbre, using perceptual categories rather than relying solely on spectral descriptors. Additionally, exploring machine learning approaches trained on timbre-labeled datasets could better predict and describe genre-signature sounds. Incorporating synthesis, for example using virtual instruments, as part of timbre analysis, not just measuring outputs but focusing more on attempting reconstructions may offer new insights into genre-centric timbre.

Furthermore, future researchers should avoid over-reliance on commercial MIR APIs, particularly when these systems are opaque or unable to explain their analytical models. While focusing on emerging genres is important due to their creative relevance, building larger datasets would likely help establish more patterns, validate observations, and bring necessary nuance to feature mapping, making the translation of data into practical toolkit guidance clearer and more actionable.

Future research may also consider assessing whether symbolic music generation could assist in dataset creation, or whether symbolic generation is useful in this context, acknowledging the absence of accurate automated audio-to-music transcription for the target genres. Symbolic approaches could supplement MIR-derived features by offering structured, editable representations of musical content, but their relevance needs critical evaluation on a genre-by-genre basis.

To maximize practical adoption among producers, future toolkits should present extracted data in clear, human-readable formats such as simplified guides or checklists. Ideally, a fully developed toolkit would include preset packs or sample libraries derived from the extracted genre features.

Toolkit designs should prioritize interpretable features that align directly with real-world production tasks such as sound design, arrangement, and mixing decisions. Toolkits should avoid presenting data that, while technically descriptive, offers little practical utility. For example, stating that a sound "has a harmonic-inharmonic ratio 74.8% higher than white noise" may be mathematically true, but it is unlikely to meaningfully guide a producer's creative decisions.

Future studies should also build structured validation directly into their methodology to assess the effectiveness of the developed prototype toolkits. Expert validation would begin with careful criteria for selecting experts: these could include editors or active contributors to user-maintained music databases, such as RateYourMusic genre pages, and individuals with a verified listening history and substantial review volume for the target genre, ideally with self-reported familiarity with the genre's history, key artists, and stylistic aesthetics.

Ideally, there should be at least one subject matter expert per genre, with two or more experts preferable to allow for triangulation of perspectives. The feedback format could be structured to allow flexible, freeform commentary alongside targeted questions. Potential questions for experts to be asked could include: "Does this track conform to your understanding of [genre]?", "Which aspects feel authentic or inauthentic?", "What musical or production elements are missing?", and "Would you expect to find this track on a playlist or release in this genre?"

The evaluation method would involve qualitative analysis of expert feedback to identify recurring themes. Researchers could also compare expert impressions with broader listener survey patterns, using these insights to isolate which toolkit components contributed most to perceived authenticity. An additional refinement might involve presenting multiple musical examples derived from the toolkit to expert listeners, such as one track that adheres very strictly to toolkit parameters and one that takes deliberate liberties, and gathering commentary on the perceived differences between them. Such an approach could reveal highly valuable qualitative data about the limits and strengths of data-driven toolkit design.

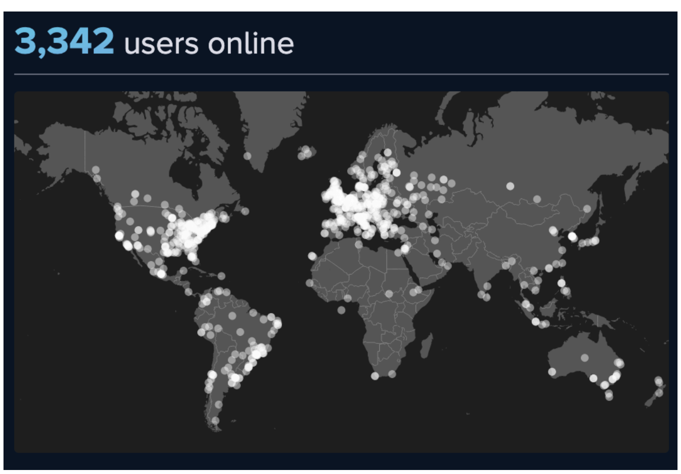
Finally, expanding validation efforts to include broader non-expert listener surveys could uncover perceptions that a small expert population might miss. Rather than treating the initial toolkit as a finished product, iterative prototyping and refinement cycles based on expert and listener feedback should be integrated into future studies. This mixed approach combining automated extraction with expert annotation and qualitative feedback, may offer the most viable path forward for developing practically useful, data-driven production toolkits for emerging genres.

Anticipated Limitations for Future Research

As well as the abovementioned methodological limitations, future research into this area should address the anticipated limitations of this avenue of research.

It may not be possible to assess genres that are too small and niche, as an insufficient amount of existing music and a listenership that was too small would hinder meaningful analysis and agreement on genre categorization. How small the threshold is remains to be determined.

If future studies were to rely on getting a dataset of songs from the RateYourMusic database, highly regional genres, or genres popular only within certain groups of people may present difficulties from insufficient data due to RateYourMusic's user distribution. While RateYourMusic doesn't directly publish web traffic statistics, the RateYourMusic forums have a list of all registered users online, and a map view to see where users roughly are. Most users are situated across Europe and the USA, with a few groups of users in South America and Oceania.



RYM users online, 27 April 2025, 15:11

Machine analysis, while currently effective in identifying quantifiable parameters such as BPM or frequency content, may struggle to capture more nuanced subjective elements like "feel" or contextual nuances tied to subgenres. (timbre and rhythm)

Additionally, the development of a production toolkit faces complexity in balancing room for creativity with respect to targeting machine-attainable parameters. This could be alleviated by cohering to a proposed optimal differentiation of musical features, like the paper mentioned in the literature review by Berger, Jonah, and Bradlow.

Narrowing the study's focus exclusively to sonic factors risks neglecting cultural, social, and historical factors that are each important to a subgenre's identity, leading to incomplete representation of a genre.

While preliminary research suggests that genre recognition relies heavily on timbre, lyrical content and overall themes and artist aesthetics cannot be excluded from genre. Studies based on lyrics-based genre classification, such as "Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network." by Alexandros Tsaptsinos reinforces this fact of the nature of musical genre. While genre recognisability should not be highly impacted by this decision, genre authenticity to listeners may be.

Furthermore, as emerging subgenres evolve rapidly, parameters identified during analysis may quickly become outdated as genres shift or integrate new influences.

Conclusion

This project set out to explore whether a data-driven approach could support the creation of production toolkits for niche and emerging musical sub-genres. By extracting quantifiable musical features from representative tracks, the goal was to provide producers with a structured, genre-specific framework that could reduce the time and effort needed to adapt to unfamiliar styles.

While some success was achieved in identifying consistent low-level features, such as tempo, phrasing, and mode, critical aspects of genre identity, such as timbre, harmonic nuance, and mix design, remained elusive. Automated tools struggled with reliability, particularly in the transcription of chords and the interpretation of complex sonic textures. The unexpected deprecation of Spotify's API further disrupted the project's workflow and highlighted the fragility of relying on proprietary data sources.

Rather than a complete production toolkit, this dissertation presents a reflective case study in the feasibility and limitations of current Music Information Retrieval (MIR) methods in creative production contexts. It demonstrates that while MIR can support structural mapping, genre emulation remains deeply tied to human interpretation, tacit knowledge, and cultural context.

The resulting prototype serves as a partial step forward, and the challenges encountered offer valuable guidance for future researchers. True progress in this field will likely depend on hybrid methods combining automation with expert insight to bridge the gap between measurable sound and meaningful music.

Bibliography

"Asian Rock." *Rate Your Music*, https://rateyourmusic.com/genre/asian-rock/. Accessed 28 Apr. 2025.

Buckley, Emily. Technology and the Music Industry: A Study of Effects and Production Techniques Across Genres. Research Paper, 2022.

Berger, Jonah, and Eric T. Bradlow. "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music." Psychological Science, vol. 29, no. 3, 2018, pp. 361–372. DOI:10.1177/0956797617736980.

DeSantis, Dennis. Making Music: 74 Creative Strategies for Electronic Music Producers. Ableton AG, 2015.

Elvers, Paul. "Is There Any Documentation about Spotify's Audio Features?" *ResearchGate*, 28 Sept. 2016, https://www.researchgate.net/post/Is-there-any-documentation-about-Spotifys-audio-features.

Fabbri, Franco. "A Theory of Musical Genres: Two Applications." Popular Music Perspectives, edited by D. Horn and P. Tagg, International Association for the Study of Popular Music, 1981, pp. 52–81.

Ferraro, Carmen, and Kjell Lemström. "Large-Scale Genre Classification Using Repeating Patterns." Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 2019.

Foxworth, Nolan A. Musical Genre Identification and Differentiation of Rock, R&B/Hip-Hop, and Christian Songs Through Harmonic Analysis. Honors Thesis, 2017.

Gjerdingen, Robert O., and David Perrott. "Scanning the Dial: The Rapid Recognition of Music Genres." Journal of New Music Research, vol. 37, no. 2, 2008, pp. 93–100.

Gouyon, Fabien, Bob L. Sturm, João Lobato Oliveira, Nuno Hespanhol, and Thibault Langlois. "On Evaluation Validity in Music Autotagging." *arXiv preprint* arXiv:1410.0001 (2014).

Hern, Alex. "Spotify Acquires Music Data Firm The Echo Nest." *The Guardian*, 6 Mar. 2014, https://www.theguardian.com/technology/2014/mar/06/spotify-echo-nest-streaming-music-deal.

Margulis, Elizabeth Hellmuth. "What Makes Songs Catchy?" Frontiers in Psychology, vol. 4, no. 1, 2013, pp. 1–13. DOI:10.3389/fpsyg.2013.00666.

MIDiA Research. "Global Music Subscription Market Sees Growth Led by Emerging Markets: MIDiA Research Q3 2023 Report." MIDiA Research, 2023, www.midiaresearch.com/blog/global-music-subscription-market-sees-growth-led-by-emerging-markets-midia-research-q3-2023-report.

Morreale, Fabio, et al. "Pop Materialising: Layers and Topological Space in Digital Pop Music." Organised Sound, vol. 25, no. 3, 2020, pp. 1–15.

"Neo-Grime." *Rate Your Music*, https://rateyourmusic.com/genre/neo-grime/. Accessed 28 Apr. 2025.

O'Brien, Cian, and Alexander Lerch. "Genre-Specific Key Profiles." Music Perception, vol. 33, no. 3, 2015, pp. 243–52.

Snoman, Rick. Dance Music Manual: Tools, Toys, and Techniques. Focal Press, 2019.

Spotify. "Quarterly Financial Results." Spotify Investors, 2024, investors.spotify.com/financials/default.aspx#quarterly-results.

Stober, Sebastian. "Toward Studying Music Cognition with Information Retrieval Techniques: Lessons Learned from the OpenMIIR Initiative." *Frontiers in Psychology*, vol. 8, 2017, article 1255, doi:10.3389/fpsyg.2017.01255.

Tsaptsinos, Alexandros. "Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network." *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, 2017, Suzhou, China. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Van de Merwe, Peter. Origins of the Popular Style: The Antecedents of Twentieth-Century Popular Music. Clarendon Press, 1999.