Final Year Project

Automating Music Classification

Department of Digital Humanities

University College Cork

Declaration of Originality

[DH4003]: [Research Project]

Assignment : 1

Supervisor: [Órla Murphy]

In submitting this assignment, I confirm that all of the submitted materials are entirely my own original work, except where clearly attributed otherwise, and that it has not been submitted partly or wholly for any other educational award.

I hereby declare that:

- **this is all my own work, unless clearly indicated otherwise, with full and proper accreditation;**

- **with respect to another's work: all text, diagrams, code, or ideas, whether verbatim, paraphrased or otherwise modified or adapted, have been duly attributed to the source in a scholarly manner, whether from books, papers, lecture notes or any other student's work, whether published or unpublished, electronically or in print.**

Student Name / Number: [Kacper Kuciara], [119492072]

Date: [26-04-2024]

Table of Content

Chapter 1 - Intro

In the early 2000’s digital music became the norm almost overnight. Thanks to a rapid succession of inventions such as the mp3 file format, peer to peer file sharing and the iPod. In the year 2000 record sales fell by 33% directly as a result of the American file sharing website Napster (Hong, 2004). As music now had a convenient format in the form of mp3, was effectively free due to readily available file sharing sites and could be made portable even easier eliminating the need for a bulky tape or compact disk player, and replacing it with a small mp3 player with flash memory.

While record label executives were in crisis over the huge drop in sales as well as the potential prospect of losing the ability to monetize music sales indefinitely (Woelfel, 2001) many musicians, with the rather infamous exception of Metallica’s Lars Ulrich, were ready to embrace the change seeing it as potentially democratizing music access. Musicians themselves were also mostly more keen to embrace the digital revolution as their industry had embraced digital recording formats within the last decade themselves with great success. Digital recording directly to a computer hard disk, removing the need for expensive and finicky reels of tape also made home recording for musical artists a reality, greatly lowering the threshold required to break out onto the music scene.

As both consumers and professionals musical libraries grew, so did the need to organize and categorize these libraries. Around the early 2000s is when the field of Musical Information Retrieval began to come through with early solutions and innovations in the field of automating music classifications utilizing models of feature and data extraction from digital files and assigning them to pre-defined genres.

This project aims to further the research on methods of automating music classification based on extraction of features such as chords, modes, pitch and tempo. Another objective of this project is the development of an innovative tool to automate music classification enhanced with Spotify track detection, which if successfully recognizing a known track will verify its classification with Spotify's leading global library. The tool also aims to aid in classifying and categorizing sample libraries for working musicians to organize and speed up the process of accessing vast swaths of samples at a moment's notice with easily searchable genres.

Chapter 2 - Literature Review

Music Information Retrieval (*MIR*) is an area of research that focuses on automating the extraction of data from music. There has been a significant amount of research undertaken in the area of automating the classification and categorization of music in the last 20 years. One current method of music classification practiced in MIR is achieved via feature extraction utilizing data on the instrumentation, texture, rhythm, dynamics, pitch and melody of a track (McKay and Fujinaga, 2005). The Bodhidharma software developed by McKay extracts this data from MIDI files, and using feature weightings classifies the file as one of 9 main genres (e.g Rock) and 38 sub genres (e.g Hard Rock). The system was effective in accurately classifying the songs within the main 9 genres 90% of the time while in the sub genres the accuracy dropped to about 57% still proving a favorable result. The largest issue with this approach is that it relies on data extraction from MIDI files as it is a binary format that was designed to be easily legible by synthesizers and digital inferences in the 1980s. (Swift, 1997). Although relatively popular in professional environments, this file format is not utilized by consumers and is less useful for practical genre classification applications that we intend to propose (Holmes, 2003).

Another method utilizing feature extraction involves automating chord recognition. This can be done from music files in regular formats such as mp3 using hidden Markov models (Nasridinov and Park, 2014). Utilizing speech recognition tools researchers have converted the input signal into identifiable frequencies and then mapped it to the Pitch Class Profile where the hidden Markov model was trained to recognize each individual chord in a track (Lee and Slaney, 2006). The study revealed from a set of 20 tracks this method had an accuracy level of around 75% (Lee and Slaney, 2008).

Further methods of chord extraction include a template based chord recognition approach. In this method rather than relying on hidden Markov models, defined chord templates are applied to chrome vector calculations from the initial input track signal and run through a filtering process in order to eliminate imperfections in the signal to improve chord mapping (Oudre *et al.* 2009). Variations on this method include utilizing the Enhanced Pitch Class Profile where using the Harmonic Product Spectrum algorithm data is extracted and then applied to an existing template featuring 24 major and minor chords (Lee, 2006). EPCP has shown to be less affected by minor harmonic differences between chords than models utilizing hidden Markov models. While chord extraction methods are continually researched into, these methods still face many of the same challenges regarding accuracy due to the complex nature of audio signal retrieval and subjective nature of musical genre classification itself. While alone chord extraction may be inaccurate enough and not provide enough information to determine an adequate genre classification in more complex music, in combination with a track separation method the results could prove more usable (Rosner and Kostek, 2018).

The track separation method proposed by Kostek and Rosner (2018) utilized Non Negative Matrix Factorization to extract five individual tracks out of 8000 different tracks analyzed from the Synat database reaching a 72% successful classification rate compared to past results of 60% (Tzanetakis *et al*. 2002) and 57.8% (Burred, 2014), showing the biggest improvement in more complex genres such as Metal, Rock and Latin Music. This method applied with a reliable chord extraction model such as could the template based chord recognition model proposed by Oudre *et al.* (2009) prove to be a more reliable method of genre classification.

Thanks to the generational leaps occurring within the sphere of Artificial Intelligence and Natural Language Processing in the past few years, there is also a considerable amount of ongoing research into methods of incorporating deep neural network methods into the field of MIR. GPT 3 has proved to be relatively adept at extracting information from and making estimates as to the genre of a given sequence of notes however losing its footing when it came to more complicated notes sequences that could be interpreted in different ways musically (Krol *et al.* 2022). When presented with a text - based music notation system, GPT did relatively well with identifying features such as key and timing but again failed with identifying more complex features such as dynamics (Krol *et al.* 2022). Another deep neural network approach has been proposed by Patil *et al.* (2023) using a mathematical model focusing mostly identifying unique rhythmic genres found in Indian and Hindustani music using their respective datasets, Hindustani Music Rhythm and Indian Music Genre alongside the widely used GTZAN music data set. The method employed by Patil *et al.* (2023) came in with minimal accuracy ranges of 94%, 93% and 96% for each respective data set, clearly indicating deep neural network methods will be the focus of MIR research for the foreseeable future. This method is also unique as it incorporates non-Western musical theory and research, not only increasing the performance by training with more diverse information but also making the research more useful and inclusive for other cultures.

As music is a thoroughly emotional art form with different emotional responses being conjured from and associated with various genres (Sloboda, 1991), it only makes sense that efforts to automate classification take emotional responses into account as another potential feature to aid classification accuracy. Ogino and Yamashita (2015) created models for emotional retrieval from lyrical data of tracks by choosing verbs and adjectives and utilizing language processing and adaptive boosting were able to assign each track with a set of emotions which could to the user.
One area where further research may be done in helping to utilize GPT for music classification could be in recognizing lyrical content of the music to add another feature set that can be extracted and analyzed as text processing has been something GPT has shown to be remarkably capable in (Manning, 2022). Research on emotional models utilizing intelligent algorithms is also sparse and could be

fleshed out as the limited results of research show promise in this area of research (Liu 2023).

Overall MIR research, specifically in the area of automating classification, rely heavily on feature extraction from musical tracks utilizing methods such as hidden Markov models, pre-defined chord templates as well as machine learning and deep neural network models. The biggest challenges facing the prospect of automating classification of music remain in the area of accuracy as well as human methods of classification of music (Mitri *et al.* 2004). Future research should focus on further integrating emerging artificial intelligence technologies as early results appear promising (Patil *et al.* 2023). More focus should also be placed in the area of classification utilizing emotional features due to the large amount of previous research available on the topic from the areas of psychology and sociology (Sloboda, 1991); (Lena and Peterson, 2008); (Kim *et al.* 2010).

Chapter 3 - Tools and Methodologies

3.1 - Research Methodology

The objective of the study is to further research into potential methods of automating music classification into pre-existing genres by furthering research into areas of MIR such as emotional and mood classification as well as feature and chord recognition.

The research approach for this project relied on a mixed usage of quantitative and qualitative research methods due to the nature of the field of Music Information Retrieval as this area of study requires both approaches. Quantitative research methods such as exploring data from previously done research such as studies on human responses to different qualities displayed in music such as variations between major and minor modes (Hevner, 1935) or values such as pitch and tempo (Hevner, 1937) as these more empirical methods are required in establishing boundaries between factual based and feeling based results required to develop a consistent and repeatable system of classification (Chau *et al.* 2016). However elements of qualitative research are also necessary in regards to genre classification as musical genres naturally evolve and progress within society (Lena and Peterson, 2008), with definitions and classifications of what types of features or qualities qualify a track into what genre changing drastically in shorter and shorter time spans thanks to the immediate access of the internet and social media being able to influence trends and standards quicker than ever before (Zhong, 2022). As the aims of this project are to further research in both, more objective and subjective arenas of the MIR research mixed research methods made the most sense to employ.

Ethical considerations regarding the usage of copyrighted material in the form of licensed music were made and as the project meets the guidelines of the European Copyright Directive permitting to “use copyrighted material for education, research and preservation of cultural heritage: the exceptions allowing these uses have been modernised and adapted to the technological changes, to allow uses online and across borders.” no further actions were required (European Commision, 2021).

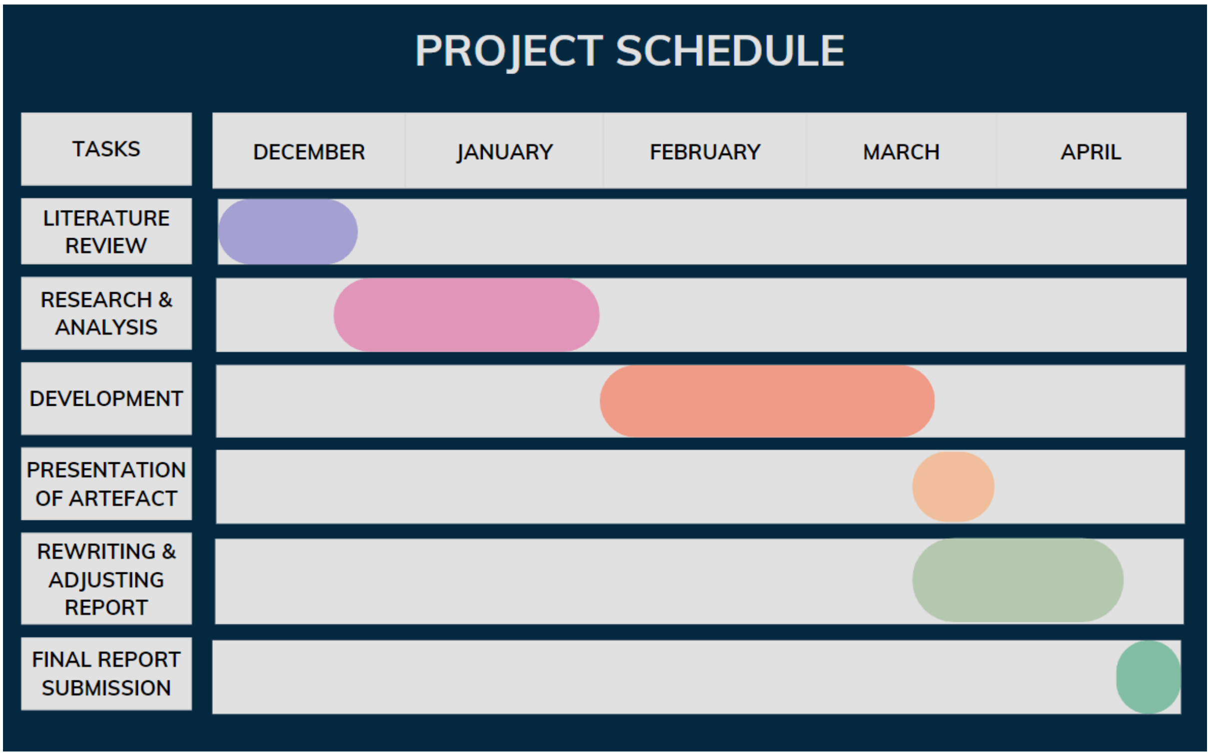
3.2 - Planning

The planning process began with establishing the requirements, objectives and goals of the project following in Agile methodology. As the scope of the project was to further research and create a working artefact in the sphere of automating music classification the following objectives emerged:

- Implement an enhanced feature extraction model
- Analyze extracted features i.e Chords, Mode, Pitch, Dynamics etc.

- Establish necessary parameters for the genre classification function
- Employ verification system if matching track found in Spotify database
- Test the system thoroughly
- Design a graphical user interface

Once the scope and requirements were understood the project was also split into individual sections which could be utilized as milestones along the schedule. A timeline was established for each respective section of the project based on estimates of previous work of a similar nature such as the literature review section and previous research undertakings (*pictured below*).



Potential risks emerged in the form of a potential lack of the required technical knowledge of python systems in order to fully realize the vision of the project; however with the aim of relieving that possibility, Python revision courses were added to the initial schedule.

3.3 - Tools and Software

When the time to choose a programming language to design the artefact in the clear choice was Python despite the field of Music Information Retrieval having utilized C++ libraries in the past. Python is favored amongst data science and data analyst users as it is built with data analytics tools and repetitive tasks in mind. Python's simple syntax as well as the vast amount of libraries available, thanks to large community support, required for Music Information Retrieval made it the most straightforward choice.

The process of chord extraction was handled by the *Chord-Extractor 0.1.2* Python package. This package was chosen as it is based off of the the NNLS Chroma and Chordino Vamp plugin library utilized by companies such as Last.fm and Chordify to for enhanced classification and chord recognition for musicians.

The Spotify Web API was employed in order to function as a verification system in case of detecting a matching track, artist or album ID, as the initial artefact design aimed to be able to automate genre classification for existing libraries as well as for user generated content such as professional sample libraries.

The Tkinter package was chosen to build the GUI for the artefact as it is the standard Python GUI Toolkit and is the most commonly used resulting in it being widely documented.

The GTZAN dataset was chosen as it is widely utilized in field of MIR and will provide an accurate comparison to other research

3.4 - Implementation

Attempting to extract 4 genres from the GTZAN dataset, the chord extractor runs on 100 tracks present in the genre dataset which gets us a set of chord frequencies. The data is then normalized utilizing min-max normalization and graphed in order to visualize what chords are most used in what genre. Then we get the chord frequency of our sample tracks being The Animals - House of the Rising Sun, Superheaven - Youngest Daughter and Billy Joel - Piano Man. That set of chords is then graphed and compared to the GTZAN dataset in order to achieve a genre classification for each track. The tracks are also ID matched with Spotify's API in order to validate the key that each track should be in.

Chapter 4 - Results Analysis

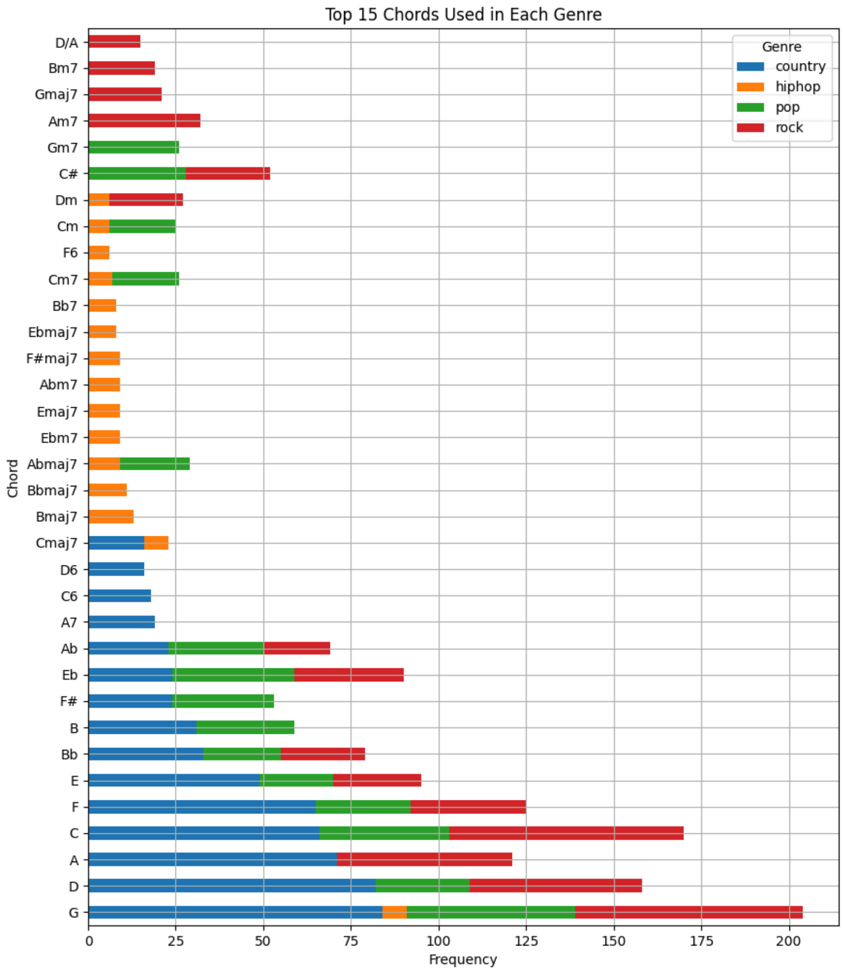


Figure 1

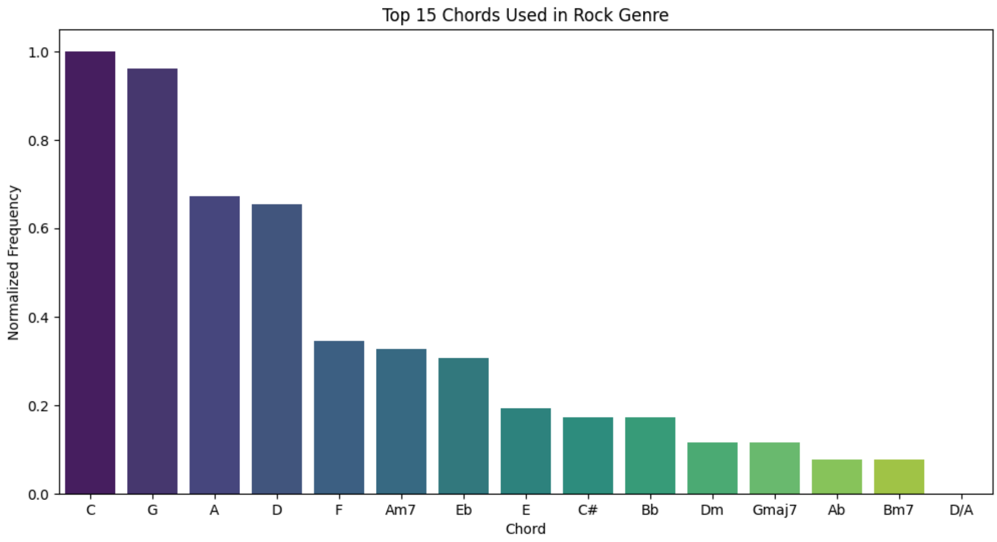


Figure 2

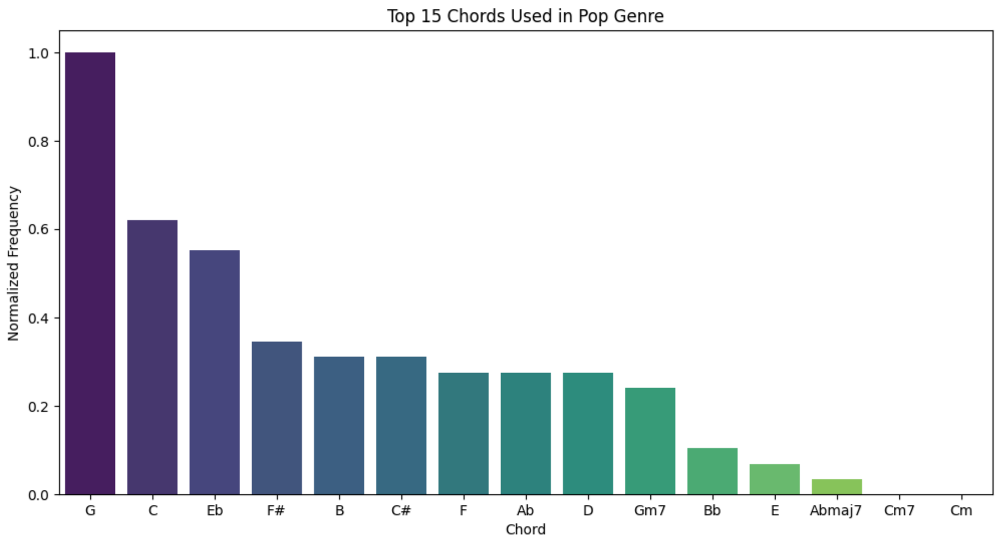


Figure 3

Figure 4

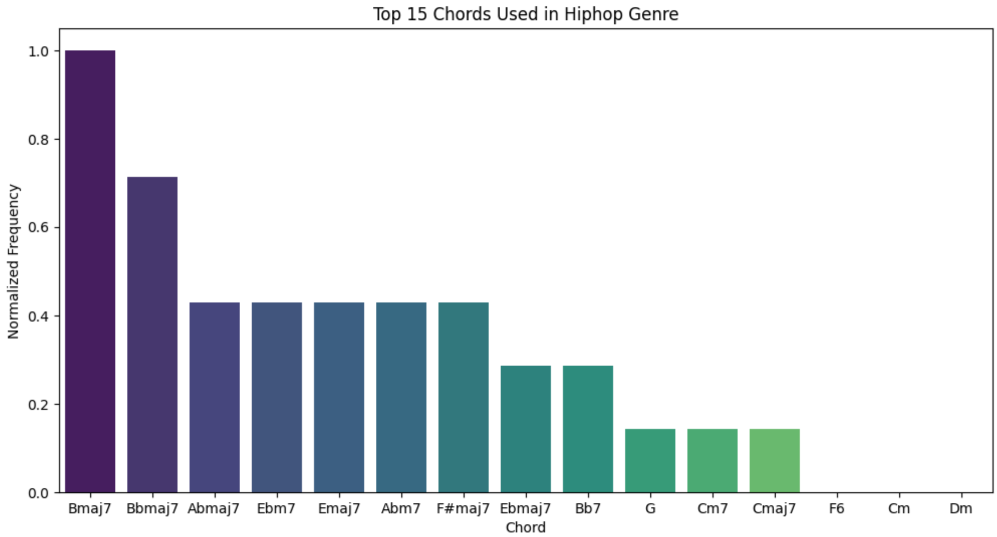


Figure 5

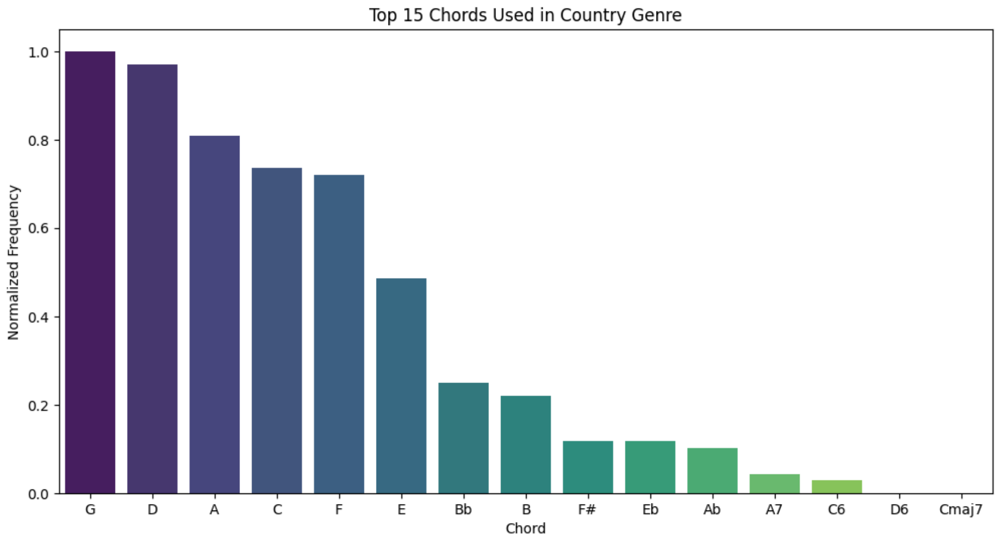


Figure 6

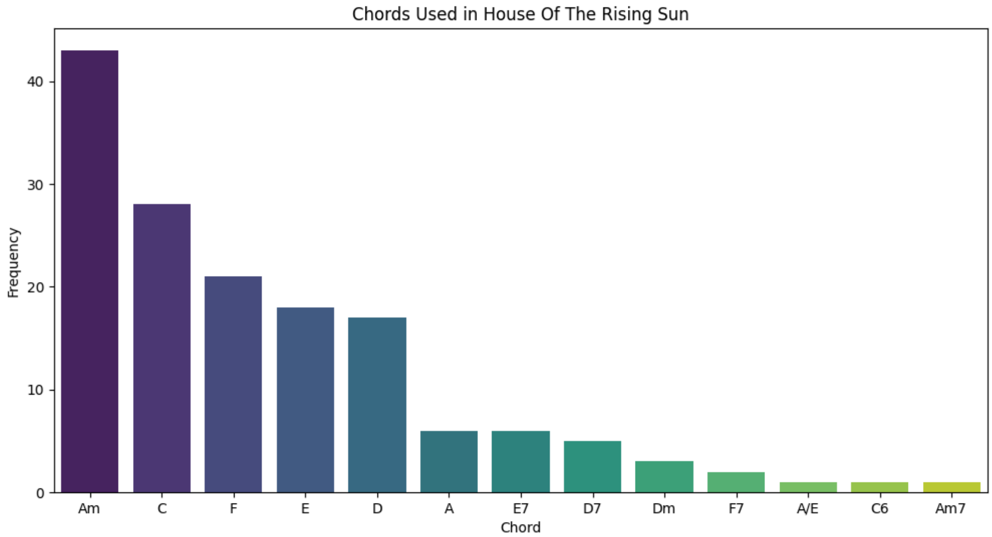
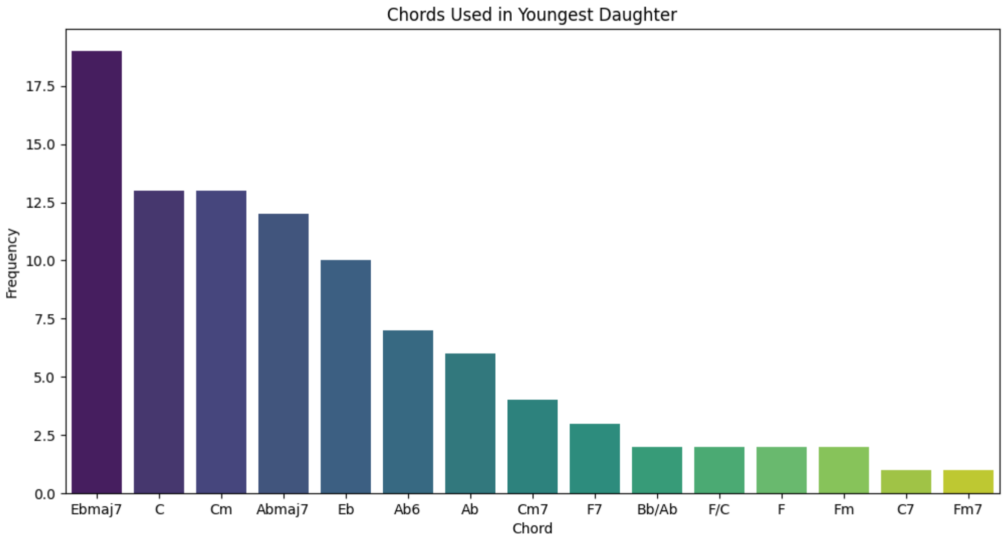


Figure 7



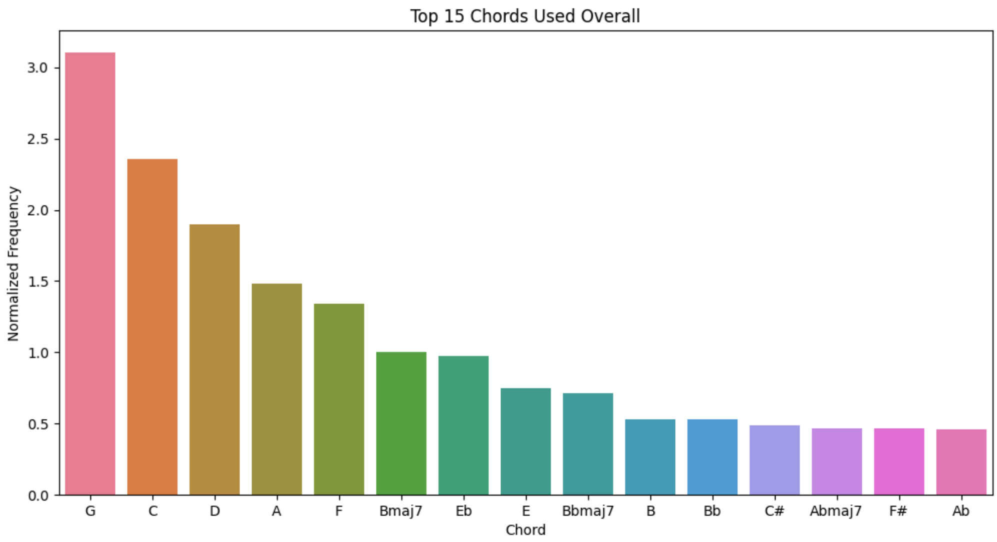


Figure 8

The chord graphs show when compared to our GTZAN dataset, our sample tracks The Animals - House of the Rising Sun and Superheaven - Youngest Daughter both share the chord C major as the second most prominent chord which features as the second most prominent chord in the pop genre GTZAN subset as well as being the second most used chord accounting for all 4 genres. Overall both of our sample tracks fall closest in chord frequency to the Rock and Pop genre with major crossover with the Country genre and less so with the Hip Hop genre subset.
By this method we could semi reliably determine that the track Superheaven - Youngest Daughter could be classified as a Rock song while The Animals - House of the Rising Sun could be classified as a Country song utilizing our limited genre sets.

Chapter 5 - Challenges

During the development cycle of the artefact many unexpected challenges arose. One challenge was regarding the Chord-Extractor package. As the Chord-Extractor package is no longer receiving any updates it is only supported up to Python version 3.9 and only runs up to Python version 3.9. As a result of this multiple hardware resets were required across 2 Mac platforms as the database refused to run on any version of Python higher than 3.9. For unknown reasons, Python 3.12 could not be uninstalled from either Mac computer and a clean new version of Mac OS had to be reinstalled in order to function. Unfortunately by mistake another instance of Python 3.12 was installed on the computer and required another system reinstall as none of the proposed fixes worked. This caused considerable delays in development. Once the chord-extractor function got running it turned out to unfortunately be quite unreliable as on certain occasions it would show the same chord sets for all sample tracks, *Figure 9.* depicting it working as expected and *Figure 10.* showing it applying the same chord set to all tracks.

```
python3 main.py

SONG: House Of The Rising Sun

Chords extracted from the song:
{'A': 6, 'C': 28, 'D': 17, 'F': 21, 'E': 18, 'A/E': 1, 'C6': 1, 'Am': 43, 'D7': 5, 'F7': 2, 'E7': 6, 'Dm': 3, 'Am7': 1}

Chord analysis suggests: The song is in a MAJOR key
Spotify API track analysis suggests: The song is in a MINOR key
The tempo of the song is 117.2 bpm
The time signature of the song is 3/4


SONG: Youngest Daughter

Chords extracted from the song:
{'C': 13, 'Ebmaj7': 19, 'Ab': 6, 'Cm7': 4, 'Eb': 10, 'Abmaj7': 12, 'F7': 3, 'Bb/Ab': 2, 'Cm': 13, 'C7': 1, 'Ab6': 7, 'F/C':
```

Figure 9.

```
python3 main.py

SONG: House Of The Rising Sun

Chords extracted from the song:
{'C': 13, 'Ebmaj7': 19, 'Ab': 6, 'Cm7': 4, 'Eb': 10, 'Abmaj7': 12, 'F7': 3, 'Bb/Ab': 2, 'Cm': 13, 'C7':

Chord analysis suggests: The song is in a MINOR key
Spotify API track analysis suggests: The song is in a MINOR key
The tempo of the song is 117.2 bpm
The time signature of the song is 3/4


SONG: Piano Man

Chords extracted from the song:
{'C': 13, 'Ebmaj7': 19, 'Ab': 6, 'Cm7': 4, 'Eb': 10, 'Abmaj7': 12, 'F7': 3, 'Bb/Ab': 2, 'Cm': 13, 'C7':

Chord analysis suggests: The song is in a MINOR key
Spotify API track analysis suggests: The song is in a MAJOR key
The tempo of the song is 177.734 bpm
The time signature of the song is 3/4


SONG: Youngest Daughter

Chords extracted from the song:
{'C': 13, 'Ebmaj7': 19, 'Ab': 6, 'Cm7': 4, 'Eb': 10, 'Abmaj7': 12, 'F7': 3, 'Bb/Ab': 2, 'Cm': 13, 'C7':
```

Figure 10.

The scope of the project eventually also became a problem as many features or functions refused to run correctly including GUI elements which due to time constraints ultimately had to be removed from the final build as no matter what occurred they caused errors.

Another area in which the scope of the project became a problem was in regards to the initially proposed sorting functionality as that was eventually scrapped as well due to being too complicated to work around with while basic functions such as chord recognition were still not operating correctly.

If the project was to be redone the scope would be made smaller and focused on more reliable models such as the Pitch Class Model utilized in Lee and Slaneys research (2006).

Chapter 6 - Summary and Conclusions

Thanks to the initial release of the iPod in 2001 and the advent of the age of streaming in the 2010's, music has certifiably moved into the digital realm, leaving behind the era of physical media such as tapes and CD's. This newfound ease of access has caused consumer libraries to increase both in variety of music and size (Sharma *et al.* 2023). These changes have brought about the possibility and increased the need to move past standard methods of manually assigning genres and categories to tracks done by experts. Research into Automating music classification is currently still a large part of the field of Music Information Retrieval. There are many models of feature extraction such as being researched such as the Pitch Class Profile, utilizing hidden Markov models, (Lee and Slaney, 2006) pre defined chord templates method (Oudre *et al.* 2009) as well emerging deep neural network approaches (Patil *et al.* 2023) due to the large strides made in artificial intelligence in recent years. Further areas of research for the field of Music Information Retrieval could include the use of artificial intelligence for processing lyrical content of tracks in order to add another potential feature to improve classification accuracy.

Research also shows that current methods of automating music classification are still not accurate enough to reliably and consistently identify tracks correctly. Tools such as the Chord-Extractor Python package are still not accurate enough to be reliably utilized in a project with a large scope. Spotify serves well as a validator due to their extensive library of tracks and their accessible Web API and could be utilized in larger projects as a secondary validation track if the sample being analyzed is found within its library.

In conclusion, the project has shown that while chord recognition is possible and useful in the process of automating music classification, in its current state it is still an imperfect solution due to accuracy issues. The scope and scale of the project proved to be too grand in comparison to the amount of technical issues encountered and the skill ceiling required to fully utilize the depth of Musical Information Retrieval research, as a result has fallen short of reaching its full potential although with further refinement could be transformed into a plugin into a Digital Audio Workstation for detection.

Bibliography

- Sharma, D, Taran, S, and Anukul, P. (2023) ‘ A fusion way of feature extraction for automatic categorization of music genres’, *Multimedia Tools and Applications*, 82(1)
- McKay, C, and Fujinaga, I. (2005) ‘ Automatic music classification and the importance of instrument identification’ , *Proceedings of the Conference on Interdisciplinary Musicology (CIM05*),
- Nasridinov, A, and Park, Y. (2014) ‘ A Study on Music Genre Recognition and Classification Techniques’ , *International Journal of Multimedia and Ubiquitous Engineering*, 9. 33-42
- Lee, K, and Slaney, M. (2006) ‘ Automatic Chord Recognition from Audio Using a HMM with Supervised Learning’ , 133-137
- Lee, K, and Slaney, M. (2008) Acoustic Chord Transcription and Key Extraction From Audio Using Key Dependent HMMS Trained on Synthesized Audio ‘ , *IEEE Transactions on Audio Speech and Language Processing*, 16(2). 291 - 301
- Oudre, L, Fevotte, C, and Grenier, Y. (2009) ‘ Chord recognition using measures of fit, chord templates and filtering methods’. 9-12
- Kyogu, L. (2006) ‘ Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile.
- Swift, A. (1997) ‘ A brief introduction to MIDI ‘, *Imperial College of Science Technology and Medicine*,
- Holmes, T. (2002) *Electronic and Experimental Music: Pioneers in Technology and Composition.* London: Psychology Press
- Rosner, A, and Kostek, B. (2018) ‘ Automatic music genre classification based on musical instrument track separation’ , *Journal of Intelligent Information Systems,* 50(2/3). 1-22
- Patil, S, Pradeepini, G, and Komati, T. (2023) ‘ Novel mathematical model for the classification of music and rhythmic genre using deep neural network’ , *Journal of Big Data*, 10(1)
- Manning, C. (2022) ‘ Human Language Understanding & Reasoning’ , *Daedalus*, 151(2). 127-138.
- Liu, J. (2023) ‘ An Automatic Classification Method for Multiple Music Genres by Integrating Emotions and Intelligent Algorithms’ , *Applied Artificial Intelligence*, 37(1)
- Lena, J, and Peterson, R. (2008) ‘ Classification as Culture: Types and Trajectories of Music Genres’ , *American Sociological Review*, 73(5) 697 - 718
- Sloboda, J. (1991) ‘ Music Structure and Emotional Responce: Some Empirical Findings’ , *Psychology of Music*, 19. 110-120
- Ogino, A, and Yamashita, Y. (2015) ‘ Emotion - Based Music Information Retrieval Using Lyrics’ , *Computer Information Systems and Industrial Management CISIM 2015,*

- Kim, Y, Schmidt, E, Migneco, R, Morton, B, Richardson, P, Scott, J, Speck, J, and Turnbull, D. (2010) ‘ MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW’ , *11th International Society for Music Information Retrieval Conference (ISIMIR 2010*),
- Mitri, G, Uitdenbogerd, A, and Ciesielski, V. (2004) ‘ Automatic Music Classification Problems’ 315-322
- Hevner, K. (1935) ‘ The Affective Character of the Major and Minor Modes in Music’ , *The American Journal of Psychology*, 47(1). 103-18
- Hevner, K. (1937) ‘ The Affective Value of Pitch and Tempo in Music’ , *The American Journal of Psychology*, 49(4). 621-30
- Hu, X. (2010) ‘ Music and Mood: Where Theory and Reality Meet’
- Zhong, B. (2021) *Social Media Communication: Trends and Theories* Hoboken: Wiley - Blackwell
- European Commission (2021) *Questions and Answers - New EU copyright rules* Luxembourg: Office for Official Publications of the European Communities
- Chau, C, Mo, R, and Horner, A. (2016) ‘ The Emotional Characteristics of Piano Sounds with Different Pitch and Dynamics ‘ , *Journal of the Audio Engineering Society*, 65(11). 918-932
- Barrett, F, Robbins, H, Smooke, D, Brown, J, and Griffits, R. (2017) ‘ Qualitative and Quantitative Features of Music Reported to Support Peak Mystical Experiences during Psychedelic Therapy Sessions’ , *Frontiers in Psychology*, 8, 12-38
- Python 2.7 - https://www.python.org/download/releases/2.7/
- Zotero - https://www.zotero.org/
- The Animals - House Of The Rising Sun - https://www.youtube.com/watch?v=N4bFqW_eu2I
- Superheaven - Youngest Daughter - https://www.youtube.com/watch?v=VMk6i7Q0k54
- Chord-Extractor Package - https://pypi.org/project/chord-extractor/
- Tkinter Package - https://docs.python.org/3/library/tkinter.html
- Spotify Web API - https://developer.spotify.com/documentation/web-api
- Arditi, D. (2017) ‘ Digital Subscriptions: The Unending Consumption of Music in the Digital Era’ , *Popular Music and Society*, 41(3). 302-318
- Hong, S. (2004) ‘ The Effect of Napster on Recorded Music Sales: Evidence from the Consumer Expenditure Survey’
- Woelfel, A. (2001) ‘ The Napster Phenomenon: Turning the Music Industry Upside Down’
- Downie, J. (2003) ‘Music Information Retrieval’ , *Annual Review of Information Science and Technology*, 37. 295-340
- GTZAN dataset - https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data
- Billy Joel - Piano Man - https://www.youtube.com/watch?v=gxEPV4kolz0