Applied Data Science Master's Thesis

Implementing Mood-Based Music Recommendations through Spectral Feature Analysis and Instrument Separation

First examiner:
David Gauthier

Second examiner:
Dennis Nguyen

Candidate:
Erdem Kocer

July 7, 2024

Abstract

Mood-based music recommendation systems have the potential to significantly enhance user experience by tailoring music selections to fit specific emotional states. This thesis presents a comprehensive approach to developing such a system by leveraging digital signal processing (DSP) and machine learning techniques. The proposed system collects user input via a web-based interface, where users report their current emotional states and music preferences for various situations. Utilizing the Demucs algorithm, the system decomposes MP3 files into individual instrumental tracks, allowing for detailed analysis of each component's spectral features and emotional connotations. A hybrid model inspired by the U-Net architecture, incorporating both spectrogram and waveform separation, is used for this purpose. The mood assessment process, implemented with Streamlit, enables accurate capture of user emotions, which are then translated into mood weights influencing the recommendation process. This methodology ensures that the recommended tracks align with the user's emotional state and context, providing a more personalized and engaging music experience. The results demonstrate the efficacy of the system in enhancing user satisfaction through contextually relevant music recommendations.

Contents

1. Introduction

Mood-based music recommendation systems have garnered significant attention in recent years due to their ability to enhance the listener's experience by tailoring music to fit specific emotional states. Human emotions are complex and multifaceted, influencing various aspects of daily life, including productivity, relaxation, and social interactions. Music, with its profound impact on mood and emotional well-being, serves as a powerful tool to modulate these states. By understanding the intricate relationship between mood and music, we can develop recommendation systems that not only cater to individual preferences but also dynamically adapt to changing emotional landscapes. Such systems have the potential to provide personalized and contextually relevant music suggestions, enhancing user satisfaction and engagement [1]–[3] .

One of the significant challenges in developing mood-based recommendation systems lies in the separation and analysis of individual musical instruments. Traditional approaches often treat a song as a monolithic entity, analyzing its overall characteristics without considering the unique contributions of each instrument. However, instruments like the guitar, bass, drums, vocals, and piano each possess distinct spectral features and emotional connotations. For instance, a melancholic piano melody can evoke different emotions compared to an energetic drum beat. By isolating these instrumental components, we can achieve a more granular understanding of how different instruments contribute to the overall emotional impact of a song. This separation poses technical challenges, such as accurately extracting and analyzing the spectral features of each instrument, but it also opens up opportunities for more precise mood mapping and personalized recommendations.

Approaching music recommendation by separating instruments allows for a more curated and context-specific selection of music, tailored to var-

ious occasions and activities. For example, during a workout session, a recommendation system might prioritize songs with energetic drums and powerful basslines to maintain high energy levels. In contrast, a relaxation session might benefit from soothing piano melodies and soft acoustic guitar. By aligning the mood conveyed by specific instruments with the desired emotional state of the listener, we can create more effective and satisfying music experiences. This level of customization acknowledges the diverse ways in which people use music to cope with emotions, enhance their mood, or simply enjoy a moment, making the recommendation system not just a passive provider of music but an active enhancer of the listener's emotional journey.

This study introduces a sophisticated music recommendation system designed to tailor music selections based on the user's current mood and listening occasions. By integrating digital signal processing and machine learning techniques, the system collects user input through a web-based interface. Users report their emotional states and specific music preferences for different situations. This input data is processed to derive mood weights, which significantly influence the recommendation process. At its core, the Demucs algorithm is employed to decompose MP3 files into separate instrumental tracks, such as drums, bass, guitar, vocals, and piano, enabling a detailed analysis of each component.

The mood assessment component of the system is implemented using a web-based interface, developed with Streamlit, where users are prompted to answer questions about their feelings and preferences. These questions are designed to cover a range of moods and occasions, allowing the system to capture the user's emotional state accurately. The responses are then used to calculate the weight of each mood, which informs the music recommendation process. By analyzing user inputs, the system assigns specific mood weights to different music attributes, thereby aligning the recommended tracks with the user's current emotional state and context.

The core architecture of demucs is inspired by the U-Net convolutional network, particularly Wave-U-Net, and utilizes a hybrid model combin-

ing spectrogram and waveform separation. A cross-domain transformer encoder with self-attention within domains and cross-attention across domains is trained on the MUSDB HQ [4] dataset, supplemented with an additional 800 songs. This inference was conducted using NVIDIA A4000 GPUs, which required approximately 8 hours of inference time. The entire system integrates these models to produce separate audio tracks, analyze spectral features, and generate music recommendations that enhance user satisfaction and emotional engagement by providing contextually relevant music choices.

2. Literature Review

2.1 Explainability in Recommendation Systems

Explainability is crucial in recommendation systems, especially in music recommendation systems (MRSs), because it enhances user trust and satisfaction. Users often encounter recommendations that seem unexpected or inappropriate, and without an understanding of why these recommendations are made, they might lose trust in the system. Explainability helps users make sense of these recommendations, fostering a sense of transparency and reliability. This is particularly important in mood-based music recommendations, where the alignment of music with a user's emotional state is subjective and personal. By providing clear explanations, such as highlighting specific musical attributes or user inputs that influenced the recommendation, the system can improve user engagement and satisfaction. For instance, explaining that a particular song was recommended because its tempo and instrumentals match the user's current mood can validate the recommendation and enhance the listening experience [5].

A rule-based approach to weight assignment in MRSs is often preferred over complex black-box models like deep neural networks due to its transparency and interpretability. Rule-based systems use explicit criteria and pre-defined rules derived from domain knowledge, making it easier to understand and justify the recommendations. For example, in the mood-based recommendation system discussed in this study, mood weights are assigned based on psychological and music theory insights. This method ensures that each recommendation can be traced back to specific rules, such as associating slow tempos and soft dynamics with relaxation. In contrast, black-box models like neural networks, while potentially more accurate, lack this transparency. They operate through complex, often opaque decision-making processes that are difficult to interpret. This opacity can be problematic,

as users and developers cannot easily understand why certain recommendations are made. By using a rule-based approach, the recommendation system can provide clear, understandable justifications for its choices, enhancing user trust and the overall effectiveness of the system.

2.2 Psychological Influence of Music

The psychological aspects of music recommendations are an emerging area of research that explores how music influences emotions, behavior, and mental states. Studies have shown that music can significantly impact stress recovery and mood regulation, making it a powerful tool in recommendation systems. For instance, Adiasto et al. conducted a systematic review and meta-analysis of experimental studies to investigate the effects of music listening on stress recovery. They found that while music listening had a non-significant cumulative effect on stress recovery, the genre, tempo, and the personal selection of music played crucial roles in its efficacy [1]. Similarly, Chanda and Levitin's review on the neurochemistry of music underscores the role of music in modulating neurochemical systems related to stress and reward, suggesting that music can be an effective tool for managing stress and enhancing emotional well-being [6].

In the context of music recommender systems, understanding these psychological effects is essential for creating more personalized and effective recommendations. Linnemann et al.'s study on the stress-reducing effects of music in daily life found that listening to music, particularly for relaxation, significantly reduced subjective stress levels and cortisol concentrations, highlighting the importance of context and user intention in music recommendations [7]. This aligns with the findings of Leubner and Hinterberger, who reviewed the effectiveness of music interventions in treating depression and emphasized the need for personalized music choices to maximize therapeutic outcomes [8]. Integrating such psychological insights into music recommendation algorithms can enhance user satisfaction and engagement by aligning recommendations with the users' emotional and psychological needs, thereby improving the overall user experience.

These insights are crucial for developing music recommender systems that go beyond mere preference matching and take into account the psychological and emotional states of users, ultimately leading to more holistic and satisfying user experiences. By incorporating psychological principles, such as those related to stress reduction and mood enhancement, into the design and functionality of music recommender systems, developers can create tools that not only entertain but also contribute positively to users' mental health and well-being.

2.3 Acoustic Features on Recommendation Systems

Recommendation systems have become a cornerstone of digital content delivery, especially in the music industry where platforms like Spotify and Apple Music use them to curate personalized listening experiences for users. Current music recommendation systems (MRS) primarily employ collaborative filtering and content-based methods to suggest tracks. Collaborative filtering relies on user interaction data to recommend items that similar users have liked, while content-based methods analyze the attributes of the items themselves to find similar ones. Shao et al. propose a hybrid approach that combines content features and user access patterns, significantly enhancing the accuracy of music similarity measurements [9]. This hybrid strategy addresses the limitations of both methods, such as the cold start problem in collaborative filtering and the lack of user preference understanding in content-based systems.

Acoustic feature analysis plays a crucial role in content-based recommendation systems by enabling the extraction and analysis of various musical attributes. These attributes include tempo, rhythm, harmony, timbre, and pitch, which are essential for understanding the music's structure and style. For instance, the study by Sheikh Fathollahi and Razzazi demonstrates the use of convolutional neural networks (CNNs) to classify music genres based on these acoustic features, achieving high accuracy in genre classification [10]. This type of feature extraction is vital for creating detailed profiles of songs that can be matched to user preferences. Additionally, the

research by Kostek and Plewa highlights the use of low-level features such as Mel Frequency Cepstral Coefficients (MFCCs) and energy-based parameters in music mood recognition, which can be used to recommend songs that fit a user's current mood or activity [11].

In integrating these advanced acoustic analysis techniques with collaborative filtering, modern MRSs can provide more nuanced and satisfying recommendations. By leveraging both user behavior data and the intrinsic properties of the music, these systems can better cater to individual tastes and preferences, making the listening experience more personalized and engaging. Future advancements in machine learning and signal processing will likely continue to enhance the precision and relevance of music recommendations, further bridging the gap between human emotional needs and automated recommendation technologies.

3. Method

For the implementation of the recommendation system, three separate models were used. The first model (Demucs) separates the individual instruments from the song. The second model assigns individual weights for each user's mood. The third model assigns weights to each spectral feature of the instruments based on mood characteristics.

3.1 Instrument (Source) Separation

3.1.1 Demucs Algorithm

The Demucs [12] algorithm is a state-of-the-art music source separation model designed to decompose an audio track into its constituent components, such as drums, bass, vocals, and other instruments. Demucs employs a U-Net convolutional architecture inspired by Wave-U-Net, which allows it to perform detailed waveform and spectrogram-based separation. The latest version, Demucs v4, features a hybrid model that integrates a cross-domain Transformer Encoder, using self-attention within domains and cross-attention across domains, enhancing its ability to separate sources more effectively. This hybrid approach has proven to achieve a Source-to-Distortion Ratio (SDR) of 9.00 dB on the MUSDB HQ test set, indicating its high accuracy in isolating individual musical elements.

The model `htdemucs_6s` was selected for this study due to its capability to handle six different sources, including drums, bass, vocals, guitar, piano, and other instruments. This expanded capacity is particularly beneficial for analyzing a wider range of instrumental features, which is crucial for the detailed spectral feature analysis required in this research. While the standard `htdemucs` model is effective, the 6-source version provides more granularity by including guitar and piano, despite some acknowledged limitations

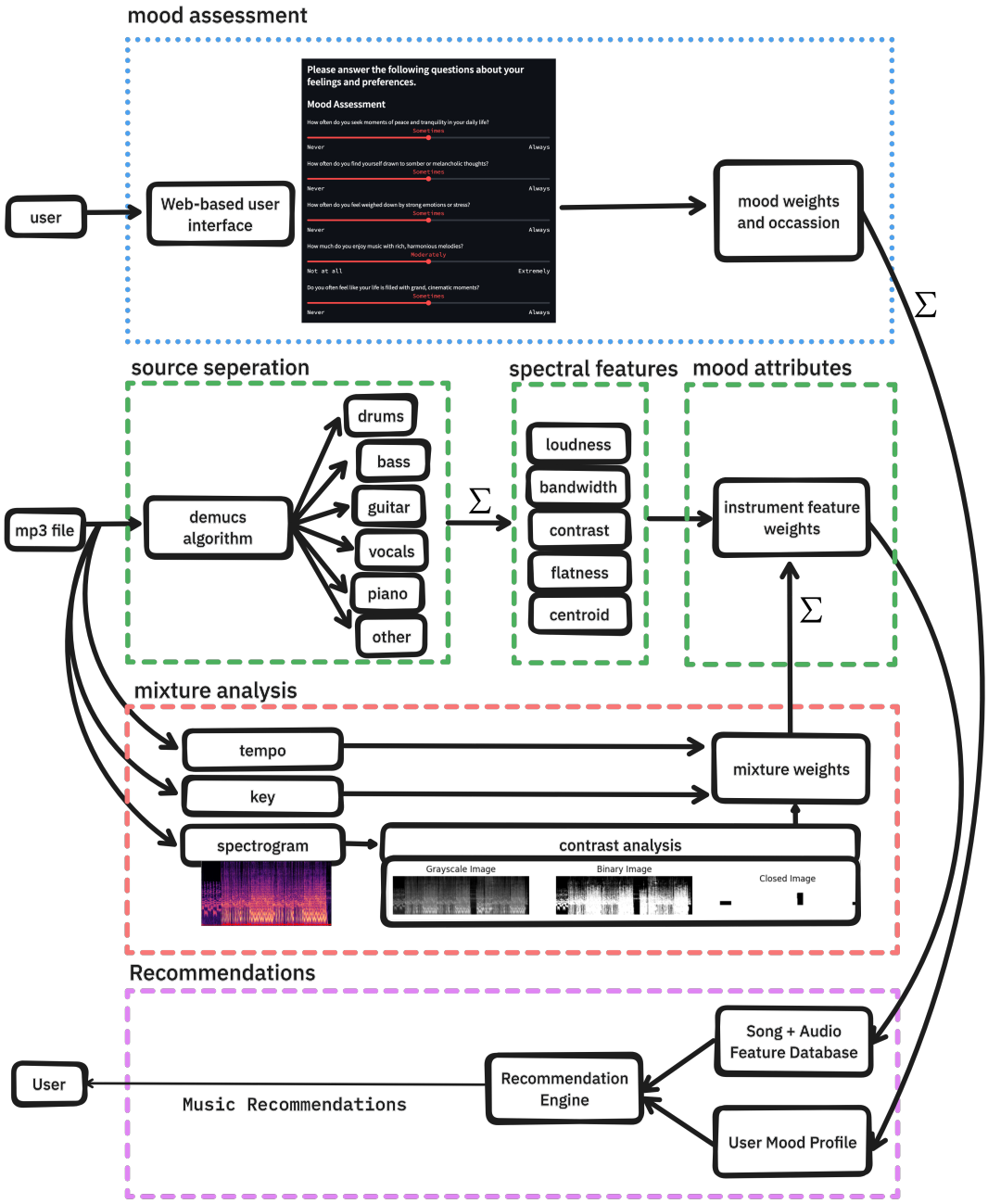

Figure 3.1: Full Application Diagram

in the piano source's quality. This model's ability to provide finer separation of additional instruments makes it a better fit for the comprehensive analysis aimed at understanding the emotional impact of each instrumental component in music.

3.1.1.1 Model Inference

Since this is a transformer-based algorithm, it requires significant computational resources, especially when dealing with large datasets that include large files. One approach would be to run the Demucs algorithm on a CPU. However, this would significantly limit the amount of data that can be processed due to the large music files. It is possible to downsample the songs to a lower kbps, but this limits the performance of the overall model. Therefore, an NVIDIA A4000 GPU was used in a cloud-based environment. This cloud-based environment had all audio files (around 5 GB), and additional network storage was needed to store the separated audio tracks (55 GB). Having network storage made it possible to stop and restart the GPU without losing any data. This allowed the Demucs algorithm to run in a more efficient setting where parallelization could be utilized. Overall, 800 songs were analyzed with an inference time of 8 hours on the GPU.

3.2 Mood Assessment

The interface for the study was developed to create a personalized music recommendation system based on the user's current mood and the specific occasion they are preparing for. The goal was to enhance user satisfaction by providing music that aligns with their emotional state and contextual needs. The interface leverages a simple, user-friendly aesthetic built with Streamlit, allowing users to input their feelings and preferences through a series of questions. This input is then processed to derive mood scores, which are used to tailor music recommendations accordingly.

The code functions by presenting users with a series of questions related to different mood states and preferences for various occasions. Users select their responses from predefined options, and these selections are mapped

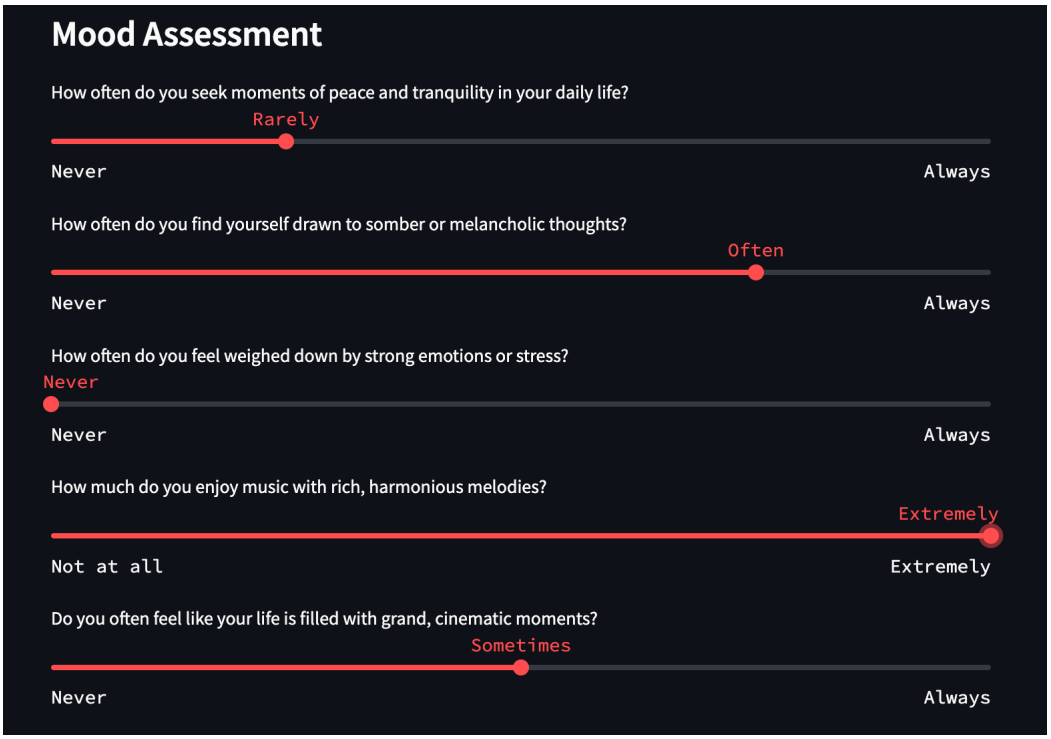

Figure 3.2: User Interface for Mood Assessment

(a) Only 5 questions are shown here. There are a total of 40 questions in the full user interface

to numerical values that quantify their mood profile. The application then prompts users to specify the occasion for which they need music recommendations, such as relaxation, workout, study, or celebration. Based on the user's mood profile and the selected occasion, the application provides tailored music recommendations, which aim to enhance the user's experience by aligning the music with their emotional and situational context.

One of the main challenges in developing this application was ensuring the accuracy and relevance of the mood-to-music mapping. This required careful consideration of the mood descriptors and their corresponding musical attributes. Another challenge was creating an intuitive user interface that would engage users and encourage them to provide accurate responses.

3.3 Mood Weight Assignment

Field knowledge and music theory play critical roles in assigning mood weights to specific musical attributes in the context of this study. For in-

stance, when considering relaxation, psychological studies have shown that certain types of music can significantly reduce stress and induce a state of calm. Music therapists often use slow-tempo music with smooth, flowing melodies to help patients unwind. Drawing from music theory, the application assigns higher weights to tracks with lower tempo, softer dynamics, and smooth harmonic progressions. Instruments like the piano or acoustic guitar, playing in a legato style, are emphasized for their calming effects, providing a soothing experience for the listener.

In the context of energy and vitality, research indicates that fast-paced, rhythmically complex music can boost energy levels and enhance physical performance, making it ideal for activities like workouts. The application, informed by music theory, prioritizes music with higher tempo, strong rhythmic elements, and upbeat melodies. Genres such as electronic dance music (EDM) or energetic rock, which feature driving beats and repetitive rhythmic patterns, are preferred to match the mood of feeling energetic and vital. This approach ensures that the music recommended for exercise sessions keeps users motivated and invigorated.

When addressing melancholy and sadness, the application leverages the understanding that music evoking these emotions often employs slow tempos, minor keys, and introspective lyrics. Such music can provide emotional catharsis for listeners experiencing feelings of sorrow. From a music theory perspective, the application assigns higher weights to tracks with minor tonalities, slower tempos, and somber, reflective lyrics. Instruments such as the cello or solo piano, known for their ability to produce rich, resonant tones that evoke sadness, are highlighted. This thoughtful integration of field knowledge and music theory ensures that the music recommendations resonate deeply with users' emotional states, offering comfort and validation.

It is important to note that the mood weight assignment presented in this study serves as a proof of concept. Assigning precise mood weights to each instrument and spectral feature is a complex and nuanced task. An automated optimization strategy based on machine learning could be em-

ployed to enhance this process by identifying the most significant spectral features. In this study, preliminary analysis of the feature set was conducted using Principal Component Analysis (PCA) to determine the most influential spectral features.

3.4 Musical Contrast

Musical contrast, which reflects how eventful or varied a piece of music is, plays a significant role in understanding the occasion and overall mood of a song. Changes in dynamics, tempo, pitch, and texture evoke different emotional responses. For instance, high contrast in music, characterized by abrupt changes in dynamics and tempo, is often perceived as exciting and energizing, making it suitable for active occasions like workouts or parties. Conversely, low-contrast music, with smooth, gradual changes, is better suited for relaxation or meditative contexts, promoting a sense of calm and stability.

To quantify musical contrast, the process begins with generating a spectrogram of the song. A spectrogram is a visual representation of the spectrum of frequencies in a sound signal as they vary with time. This involves converting the audio signal into a time-frequency representation using a Short-Time Fourier Transform (STFT), which helps identify the amplitude of various frequencies at different time points. The resulting data is then converted to decibels for better visualization.

Next, the spectrogram is converted into a grayscale image. This image represents the amplitude of frequencies, where darker regions indicate lower amplitudes and lighter regions indicate higher amplitudes. The grayscale image serves as the basis for further analysis.

Thresholding is then applied to the grayscale image. This involves setting a specific cutoff value, above which the pixel values are turned white (indicating significant energy), and below which they are turned black. This process results in a binary image that highlights regions with significant energy in the spectrogram. Thresholding isolates the most eventful parts of

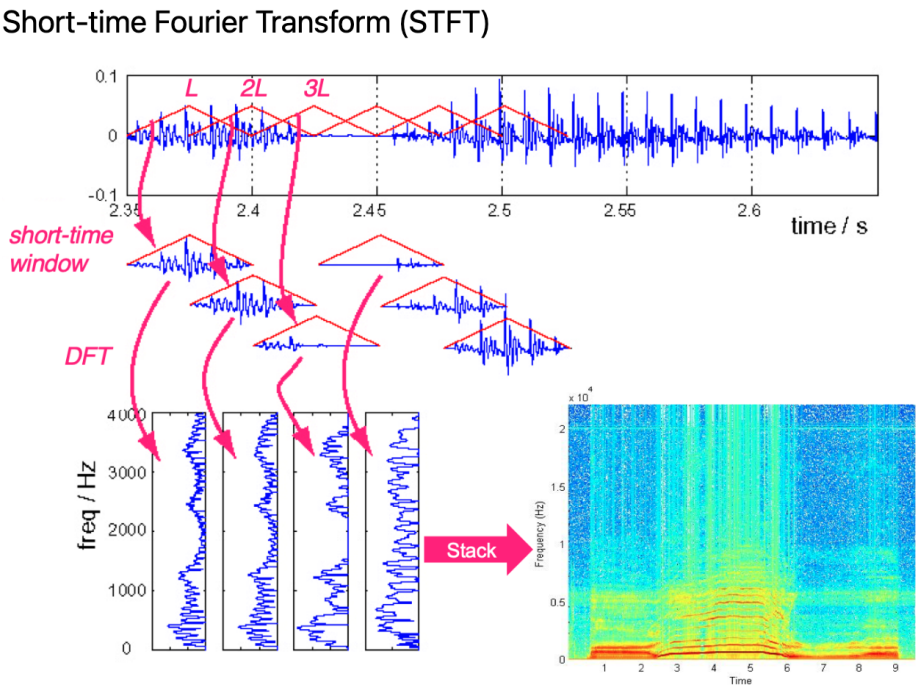

Figure 3.3: Short-Time Fourier Transform [13]

the music, which are crucial for understanding its contrast.

Morphological operations, specifically closing, are applied to the binary image. Closing is a combination of dilation followed by erosion, which helps remove small noise and fill small gaps in the binary image. This step ensures that the highlighted regions in the binary image are continuous and well-defined, making it easier to analyze the structure of the music.

Finally, the musical contrast is calculated by analyzing the differences in intensity between adjacent regions in the thresholded image. One approach is to count the number of connected components in the binary image, which represents distinct areas of high energy. The number of these components indicates the level of contrast in the music; a higher number of components suggests higher contrast, indicating more eventful music, while a lower number suggests lower contrast, indicating smoother and more uniform music.

This is achieved by using a Python module called OpenCV [14]. This library allows to make operations such as thresholding and morphing.

```
// Load the Image

// Convert to Grayscale

// Define Darkness Threshold
SET darkness_threshold to 64  // Adjust this value as
   needed (0-255)

// Separate Dark from Light
FOR each pixel in image
  IF pixel intensity < darkness_threshold
    SET pixel value to black (0)
  ELSE
    SET pixel value to white (255)
  END IF
END FOR

// Clean the Image

// Count Dark Points
SET number_of_dark_points to 0
FOR each pixel in image
  IF pixel value is black (0)
    INCREMENT number_of_dark_points
  END IF
END FOR
PRINT "There are", number_of_dark_points, "dark points."
```

Listing 3.1: Pseudo-code for Image Processing

3.5 Spectral Feature Analysis

To understand how the spectral features of each instrument were calculated, several processing steps are required. This process involves the following key steps: generating the spectrogram of the audio, extracting specific spec-

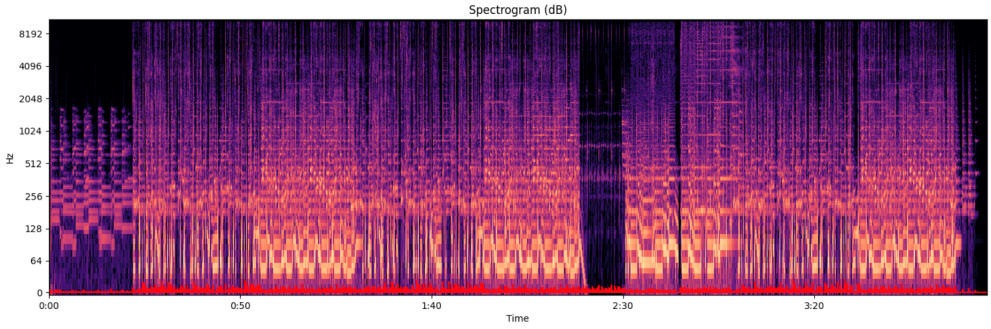

Figure 3.4: Spectrogram of a Track

tral features, and analyzing these features to derive meaningful metrics.

The first step is to generate the spectrogram of the audio file. A spectrogram is a visual representation of the spectrum of frequencies in a sound signal as they vary with time. This is achieved by converting the audio signal into a time-frequency representation using a Short-Time Fourier Transform (STFT). The STFT divides the audio signal into small overlapping segments, applies the Fourier Transform to each segment, and maps the amplitude of frequencies to different time frames. The resulting spectrogram shows how the frequency content of the signal evolves over time. Librosa library [15] in Python gives a wide selection of signal-processing functions that make it possible to generate and analyze spectrograms.

Once the spectrogram is generated, several spectral features are extracted. These features include:

Spectral Centroid: This feature indicates the "center of mass" of the spectrum and is often associated with the brightness of a sound. It is calculated by taking the weighted mean of the frequencies present in the signal, with their magnitudes as weights. A higher spectral centroid value indicates a brighter sound.

Spectral Bandwidth: This measures the width of the spectrum and is related to the perceived timbre of the sound. It is calculated by measuring the variance of the spectral centroid. A wider bandwidth indicates a more complex sound with more high-frequency components.

Spectral Contrast: This feature measures the difference in amplitude be-

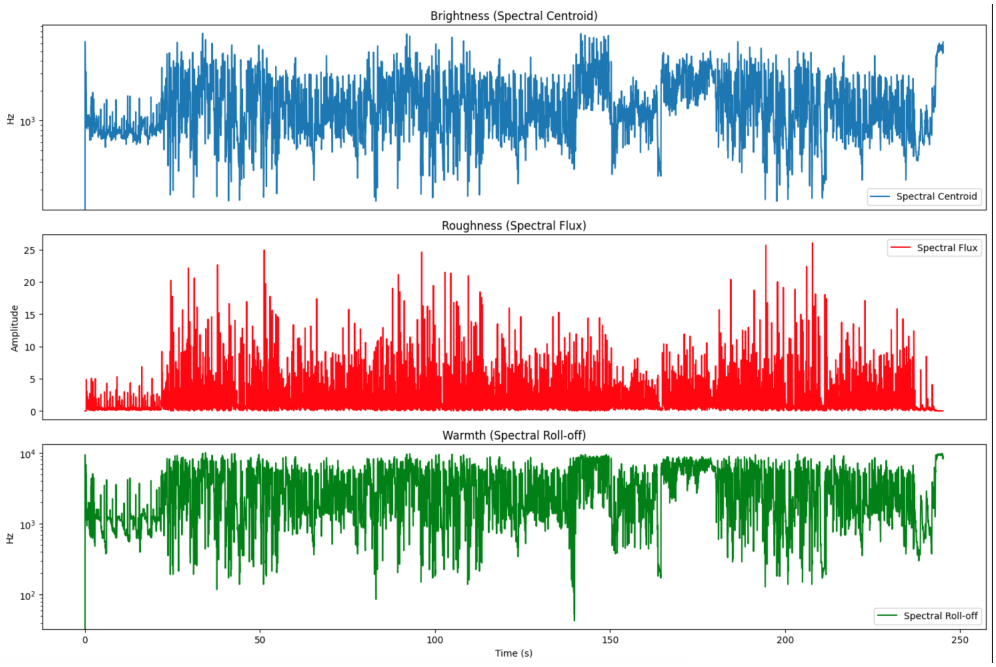

Figure 3.5: Spectral Feature Plots of a Track

tween peaks and valleys in the sound spectrum. High contrast values suggest more dynamic and eventful music, while low contrast values suggest smoother, less eventful music. This is particularly useful in understanding the overall mood and occasion suitability of the music.

Spectral Flatness: This measures how noise-like a sound is. A high spectral flatness indicates that the spectrum is relatively flat, similar to white noise, whereas a low spectral flatness indicates a peaky spectrum, characteristic of tonal sounds.

4. Results

To evaluate the effectiveness and user satisfaction of the developed music recommendation system, a user experience study was conducted involving seven participants. Each participant interacted with the system, tested the recommendations, and completed a user experience survey afterward. The survey aimed to gather insights into the usability of the interface, the relevance of the recommendations, and any areas for improvement.

4.1 User Experience Survey

The survey results indicated that the user interface was generally well received. Participants appreciated the simplicity and intuitiveness of the design, which made it easy to input their mood and occasion preferences. The use of Streamlit for developing the interface contributed significantly to this positive feedback, as it allowed for a clean and user-friendly interaction. On the other hand, participants noted that there are simply too many questions (or sliders) that make it difficult to change their preferences after the recommendations are presented.

4.2 Relevance of Recommendations

One of the key strengths highlighted by the participants was the relevance of the music recommendations in relation to the specified moods. The system was praised for its ability to align the music suggestions with the user's stated context, whether it was for relaxation, study, workouts, or celebrations. However, some participants noted that the recommendations looked "random". This might have occurred due to the small size of the dataset.

4.3 Additional Observations

Participants also provided constructive feedback on potential areas for improvement:

- **Question Reduction**: Simplifying the mood assessment by reducing the number of questions or finding alternative ways to capture mood more efficiently could enhance the user experience.
- **Performance**: While the system's performance was satisfactory, some users mentioned the occasional delay in generating recommendations. There were also some UI bugs due to some type errors in the final dataset.

Question	Response Options	Percentage	Count
How easy was it to navigate the user interface?	Very Difficult	0%	0
	Difficult	0%	0
	Moderate	14.29%	1
	Easy	14.29%	1
	Very Easy	71.43%	5
How would you rate the overall design of the user interface?	Very Poor	14.29%	1
	Poor	0%	0
	Average	14.29%	1
	Good	28.57%	2
	Excellent	42.86%	3
How simple was it to input your mood and occasion preferences?	Very Complicated	14.29%	1
	Complicated	0%	0
	Moderate	14.29%	1
	Simple	28.57%	2
	Very Simple	42.86%	3
Please rate the overall relevance of the music recommendations. 1 - Not relevant, 5 - Very Relevant	1	14.29%	1
	2	28.57%	2
	3	28.57%	2
	4	0.00%	0
	5	28.57%	2
	Total		7

Table 4.1: Survey Results for User Experience of the Interface

Overall, the results from the user experience survey underscore the strengths

of the content-based music recommendation system while also pointing out areas for refinement. The positive reception of the user interface and the relevance of the recommendations validate the approach taken, while the feedback on the mood assessment process and the desire for greater musical diversity provide valuable insights for future iterations. By addressing these aspects, the system can be further optimized to offer an even more engaging and satisfying user experience.

5. Conclusion

5.1 Discussion

In this study, we embarked on an exciting journey into the realm of content-based music recommendation systems. Unlike traditional recommendation systems that rely on collaborative filtering, content-based systems offer a more individualized approach. They analyze the actual content of the music, such as spectral features and instrumental components, to tailor recommendations specifically to the user's tastes. This personalized approach promises to revolutionize how we experience music, making our listening habits more aligned with our unique preferences and moods.

One of the main distinctions between content-based and collaborative recommendation systems lies in their core methodologies. Collaborative recommendation systems leverage the preferences of multiple users to suggest new content. This method compares a user's listening habits with those of others who have similar tastes. While this approach can be effective, it has several limitations. For instance, it requires a substantial amount of user data to be effective, which raises potential privacy concerns. Additionally, collaborative systems tend to skew recommendations towards more popular artists, potentially reducing the diversity of suggestions.

Collaborative recommendation systems, despite their widespread use, face notable challenges. They depend heavily on the aggregation of user histories, which can compromise user privacy as machine learning algorithms sift through personal listening habits. Furthermore, this reliance on user data can lead to a homogenization of recommendations, where lesser-known artists and genres receive less exposure. As a result, users might find themselves stuck in a loop of popular songs, missing out on the rich diversity of the musical landscape.

In contrast, content-based recommendation systems offer several compelling advantages. By focusing solely on the content of the music itself, these systems bypass the need for user data aggregation, thus preserving privacy. The core principle here is that the intrinsic qualities of the music, such as tempo, rhythm, and spectral features, are what truly matter. This approach allows for a richer diversity in recommendations, as it is not influenced by the listening habits of other users or the popularity of certain artists. Instead, it thrives on the unique preferences of each individual listener, providing a more tailored and satisfying experience.

The performance of the algorithm we developed underscores the benefits of content-based recommendations. By separating instruments and analyzing the spectral features of each one, we gain a more nuanced understanding of the music. This granular analysis is far superior to examining a whole mixture spectrogram. It allows us to appreciate the distinct characteristics that different instruments bring to a song. For example, while studying, a listener might prefer music where vocals are minimized in favor of instrumental tracks. This level of specificity in data analysis enables us to craft recommendations that are finely tuned to the listener's context and mood.

Moreover, while instrument separation provides detailed insights, mixture analysis remains essential for understanding broader musical elements such as tempo and contrast. These factors play a significant role in setting the overall mood and energy of a song. By integrating both detailed spectral analysis and broader mixture analysis, we can achieve a more comprehensive understanding of the music. This dual approach ensures that our recommendations are not only precise but also consider the holistic attributes of the songs.

5.2 Limitations

While the results of this study are promising, several limitations must be acknowledged. One significant challenge is the time-intensive nature of inference using the Demucs algorithm. Because Demucs is a transformer-

based model, it requires substantial computational power, necessitating the use of GPUs. This reliance on GPUs makes the algorithm more scalable through parallelization; by adding more GPUs, we can increase the number of recommendations generated. However, despite this scalability, it remains challenging to work with extremely large datasets, such as millions of songs, due to the computational overhead involved.

Furthermore, the performance and accuracy of the recommendations are inherently subjective and must be validated through real user testing. This subjectivity means that the algorithm's effectiveness can't be fully measured by computational metrics alone; user surveys and feedback are crucial for assessing the quality of the recommendations. Conducting these surveys and gathering sufficient data to refine the algorithm can be time-consuming and resource-intensive. This makes it difficult to fine-tune the spectral weights since re-running the model takes a lot of time. The system simply needs more data and more trials to work in an optimized fashion.

Another limitation lies in the scope of the instruments considered by the model. Although the version of the Demucs algorithm used in this study can separate six different sources (drums, bass, vocals, guitar, piano, and other instruments), this is not comprehensive enough to represent the full spectrum of musical diversity. For example, genres such as traditional Indian music, which often feature instruments like the sitar or tabla, or African music, with its distinct percussion instruments, may not be adequately captured by the current model. This limitation can affect the accuracy and relevance of the recommendations for non-Western music genres.

Addressing these limitations requires ongoing refinement of the model and expanding its capacity to recognize and separate a broader range of instruments. Future research should aim to include a more diverse set of musical genres and instruments to enhance the universality and inclusiveness of the recommendation system. Additionally, improving the efficiency of the algorithm to handle large-scale datasets without compromising performance will be crucial for scaling up the system to accommodate millions of songs.

In summary, while the content-based recommendation system developed in this study offers significant advantages and potential, it is not without its limitations. The computational demands, the subjectivity of user validation, and the limited scope of instrument recognition are key areas that need further development. By addressing these challenges, future iterations of the system can provide even more accurate, diverse, and personalized music recommendations to users worldwide.

Bibliography

[1] K. Adiasto, D. G. J. Beckers, M. L. M. v. Hooff, K. Roelofs, and S. A. E. Geurts, "Music listening and stress recovery in healthy individuals: A systematic review with meta-analysis of experimental studies," en, DOI: 10.1371/journal.pone.0270031. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270031 (visited on 07/04/2024).

[2] S. Chafin, M. Roy, W. Gerin, and N. Christenfeld, "Music can facilitate blood pressure recovery from stress," en, *British Journal of Health Psychology*, vol. 9, no. 3, pp. 393–403, 2004, _eprint: https://onlinelibrary.wiley.co ISSN: 2044-8287. DOI: 10.1348/1359107041557020. [Online]. Available: onlinelibrary.wiley.com/doi/abs/10.1348/135910704155 7020 (visited on 07/04/2024).

[3] R. M. Cronin, D. Fabbri, J. C. Denny, S. T. Rosenbloom, and G. P. Jackson, "A comparison of rule-based and machine learning approaches for classifying patient portal messages," *International Journal of Medical Informatics*, vol. 105, pp. 110–120, Sep. 2017, ISSN: 1386-5056. DOI: 10.1016/j.ijmedinf.2017.06.004.

[4] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner, *MUSDB18-HQ - an uncompressed version of MUSDB18*, Aug. 2019. DOI: 10.5281/ZENODO.3338373. [Online]. Available: https://zenodo.org/record/3338373 (visited on 07/06/2024).

[5] D. Afchar, A. B. Melchiorre, M. Schedl, R. Hennequin, E. V. Epure, and M. Moussallam, "Explainability in music recommender systems," en, *AI Magazine*, vol. 43, no. 2, pp. 190–208, 2022, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aaai.12056, ISSN: 2371-9621. DOI: 10.1002/aaai.12056. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aaai.12056 (visited on 04/16/2024).

[6] M. L. Chanda and D. J. Levitin, "The neurochemistry of music," en, DOI: 10.1016/j.tics.2013.02.007. [Online]. Available: https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(13)00049-1 (visited on 07/04/2024).

[7] A. Linnemann, B. Ditzen, J. Strahler, J. M. Doerr, and U. M. Nater, "Music listening as a means of stress reduction in daily life," *Psychoneuroendocrinology*, vol. 60, pp. 82–90, Oct. 2015, ISSN: 0306-4530. DOI: 10.1016/j.psyneuen.2015.06.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030 6453015002127 (visited on 07/04/2024).

[8] D. Leubner and T. Hinterberger, "Frontiers | Reviewing the Effectiveness of Music Interventions in Treating Depression," en, DOI:

10.3389/fpsyg.2017.01109. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.01109/full (visited on 07/04/2024).

[9] B. Shao, D. Wang, T. Li, and M. Ogihara, "Music Recommendation Based on Acoustic Features and User Access Patterns," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1602–1611, Nov. 2009, Conference Name: IEEE Transactions on Audio, Speech, and Language Processing, ISSN: 1558-7924. DOI: 10.1109/TASL.2009.2020893. [Online]. Available: https://ieeexplore.ieee.org/document/5230332 (visited on 05/03/2024).

[10] M. Sheikh Fathollahi and F. Razzazi, "Music similarity measurement and recommendation system using convolutional neural networks," en, *International Journal of Multimedia Information Retrieval*, vol. 10, no. 1, pp. 43–53, Mar. 2021, ISSN: 2192-662X. DOI: 10.1007/s13735-021-00206-5. [Online]. Available: https://doi.org/10.1007/s13735-021-00206-5 (visited on 04/03/2024).

[11] B. Kostek and M. Plewa, "Testing a variety of features for music mood recognition," *The Journal of the Acoustical Society of America*, vol. 134, p. 3994, Nov. 2013. DOI: 10.1121/1.4830570.

[12] *GitHub - adefossez/demucs: Code for the paper Hybrid Spectrogram and Waveform Source Separation*. [Online]. Available: https://github.com/adefossez/demucs (visited on 07/04/2024).

[13] *Representing Audio — Open-Source Tools & Data for Music Source Separation*. [Online]. Available: https://source-separation.github.io/tutorial/basics/representations.html (visited on 07/04/2024).

[14] *OpenCV - Open Computer Vision Library*. [Online]. Available: https://opencv.org/ (visited on 07/04/2024).

[15] *Librosa*. [Online]. Available: https://librosa.org/ (visited on 07/04/2024).