Capstone Technical Report: Emotion-aware Music Recommendation System

Nghi Le
Earlham College
Richmond, Indiana, USA
nble19@earlham.edu

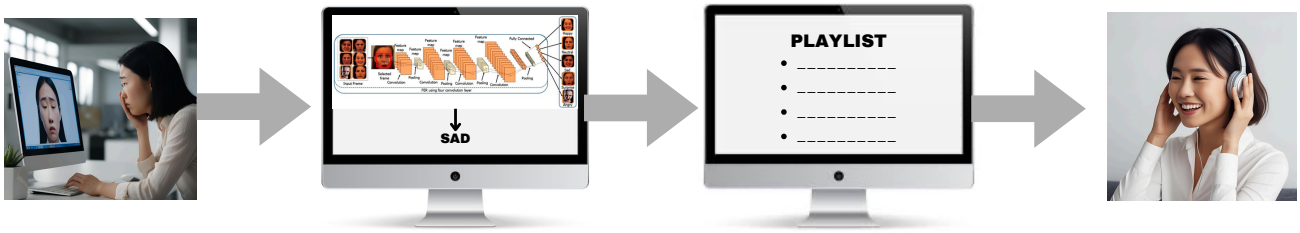

Figure 1: Graphical Abstract

Abstract

Existing emotion-based Music Recommendation Systems (MRSs) have remained underdeveloped due to the lack of focus on users' emotion approaches and the inherent psychological complexity of human emotions. To address this gap, this research introduces a novel recommendation framework that integrates users' emotional states into the music recommendation process. Specifically, the system first employs a facial expression analysis (DeepFace) for real-time detection of user emotions, then translates these emotions to scientifically validated musical attributes using Music Emotion Recognition (MER) tags guided by the Circumplex Model of Emotion. Finally, content-based filtering techniques are utilized to generate personalized playlists tailored to the user's current emotional state. Experimental results demonstrate that the proposed framework successfully delivers emotion-responsive music recommendations, effectively catering to both users' emotional needs and musical preferences.

Keywords

Facial expression recognition (FER), Emotion Detection, Deepface, Music emotion recognition (MER), The Circumplex model of emotion, Music recommendation systems (MRS), Content-based filtering

1 INTRODUCTION

Although research Music Recommendation Systems has been gaining substantial interest in both academia and industry, emotion-based MRSs are still underdeveloped in recent years. This comes from the fact that current MRSs typically center on core concepts of user-item interaction and content-based descriptors, neglecting factors that significantly affect listener musical tastes and needs such as personality and emotional states. The issue also stems from the psychological complexities in human emotion Schedl et al. (2017) [10]. As a result, current MRSs often yield unsatisfactory recommendations. To develop a more personalized and emotion-aware music recommendation system, I propose a new approach that integrates real-time emotion detection and scientifically grounded emotion-mood mapping within a traditional content-based music recommendation system. This proposed solution provides a foundational and more scientific approach to emotion-based MRSs that can address the complication in intrinsic aspects of listeners.

2 RELATED WORK

Overall, most existing related works in emotion-based music recommendation system, regardless of the approaches in emotion detection and/or music recommendation, have had major limitations on addressing how to effectively connect the inferred emotional states the listener is in and the inferred emotional concepts from the music itself. This is why this research is different from others since it introduces an innovative method of utilizing Music Emotion Recognition tags and the Circumplex Model of Emotion to bridge the gap. More details follow this section.

Below are the categories that represent the foundational and contemporary works that inform the design of the proposed emotional-aware music recommendation system.

2.1 Emotion Detection Techniques

Emotion detection plays a vital role in developing personalized recommendation systems. Traditional methods typically rely on explicit input, such as self-reported questionnaires or surveys utilizing categorical or dimensional emotion models (Russell, 1980; Zentner et al., 2008) [9] [13]. While these explicit methods provide clear emotional descriptors, they suffer from subjectivity and intrusiveness. More recently, implicit methods such as facial expression analysis (Ebrahimi et al., 2015) [2], speech emotion recognition (Erdal et al., 2016) [3], and text analysis (Dey et al., 2014) [1] have gained

popularity. Among these, facial expression recognition has emerged prominently due to its non-intrusive and real-time capabilities. Plus, FER is practical in terms of not requiring large amount of user input. Systems utilizing frameworks such as DeepFace (Serengil and Ozpinar, 2021) [11] demonstrate significant results for accurately inferring emotions from facial cues in real-time, which directly informs and justifies the selection of facial expression recognition for this research.

2.2 Music Emotion Recognition Approaches

Music Emotion Recognition (MER) involves assigning emotional labels to musical pieces based on features like rhythm, melody, lyrics, and harmony. Earlier MER techniques often overlooked distinctions between intended, perceived, and induced emotions, resulting in generalized emotion tags with limited applicability (Kim et al., 2010; Yang et al., 2012) [7] [12]. Recent MER methodologies focus primarily on perceived or induced emotions, employing machine learning algorithms trained on annotated datasets to classify emotional content accurately. However, even these advanced methods face challenges because musical emotion perception remains highly subjective (Kuo et al., 2005) [8]. While current MER systems can reliably tag basic emotions or moods such as happiness, sadness, or excitement, the complex psychological nuances influencing individual user perception still pose significant limitations. Thus, integrating MER effectively into recommendation systems requires careful consideration of subjective emotional interpretations, an area where this project aims to provide experimentation by incorporating the Circumplex model of emotion.

2.3 Existing Emotion-Based Music Recommender Systems

A variety of existing emotion-based music recommendation systems leverage emotional states as contextual factors, typically either pre-filtering user preferences or post-filtering recommendation results based on detected emotions. For example, Kaminskas and Ricci (2012) [6] utilized users' emotional states to predict short-term musical preferences but did not deeply explore the direct mapping between user emotions and music characteristics. Similarly, systems by Ferwerda et al. (2015) considered personality traits as a factor influencing emotional music preferences, yet lacked a robust real-time emotion detection mechanism. Additionally, previous systems largely ignored the interrelationship between expressed, perceived, and induced emotions, overlooking the intricate emotional dynamics of musical interactions described by Schedl et al. (2017) [10]. This project's unique contribution lies in bridging real-time implicit emotion detection (via facial expressions) with scientifically grounded MER tags and content-based filtering, creating a more dynamic, nuanced, and user-centric recommendation experience.

3 DESIGN AND IMPLEMENTATION

There are two main phases in the proposed system: emotion to music mood module and content-based recommendation module.

3.1 Datasets

Two datasets will be used for the purpose of these modules. They are initially merged by inner join method to make up a song database that contains songs with both qualitative emotional-inferred musical attributes and quantitative acoustic features.

CAL500 Dataset

The dataset used for Music Emotion Recognition tags (MER) is the Computer Audition Lab 500 - song tag annotation. It contains 502 songs chosen from Western popular music. Each song has been annotated by at least 3 people with 48 musically-relevant concepts spanning three semantic categories:

- 18 emotions were rated on a scale from one to three (e.g., not happy",neutral", "happy"),
- 15 song concepts describing the acoustic qualities of the song, artist and recording (e.g., tempo, energy, sound quality),
- 15 usage terms (e.g., "I would listen to this song while driving, sleeping, etc.").

For the purpose of this project, the semantic inferred musical tags will be used to map detected emotions to, using the valence and arousal level in the Russel's Circumplex model of emotion, which will be presented in more depths in the design and implementation section (3.3).

Song	Music Characters
10cc-for_you_and_i	backing vocals
10cc-for_you_and_i	boring
10cc-for_you_and_i	calming
10cc-for_you_and_i	cleaning the house
2pac-trapped	aggressive
2pac-trapped	angry
2pac-trapped	arousing
2pac-trapped	backing vocals

Figure 2: CAL500 Dataset Snippet

Billboard 200 - Acoustic Features Dataset

In the content-based recommendation phase, I use the Billboard 200 dataset which contains acoustic features of 340,000 tracks from Billboard 200 albums from 1/5/1963 to 1/19/2019. Each row contains track ID on Spotify, track name, album name, artist name, values for Spotify EchoNest acoustic data (acousticness, danceability, energy, instrumentalness, liveness, loudness, speechiness, key, time signature, and valence), duration in milliseconds, album ID on Spotify, and release date of the album.

id	song	album	artist	acousticne:	danceabilit	duration_n	energy	instrument
0Veyvc3n9	Voices In M	Hoodie SZN	A Boogie W	0.0555	0.754	142301	0.663	0
77JzXZonN	Beasty	Hoodie SZN	A Boogie W	0.292	0.86	152829	0.418	0
18yIlZD0Td	I Did It	Hoodie SZN	A Boogie W	0.153	0.718	215305	0.454	4.60E-05
1wJRveJZLs	Swervin (fe	Hoodie SZN	A Boogie W	0.0153	0.581	189487	0.662	0
0jAfdqv18g	Startender	Hoodie SZN	A Boogie W	0.0235	0.736	192779	0.622	0
3L19besdN	Demons ar	Hoodie SZN	A Boogie W	0.0114	0.81	214593	0.55	5.69E-06
1Gno5IBUH	Love Drugs	Hoodie SZN	A Boogie W	0.55	0.762	157863	0.553	0
2A4vCndSk	Skeezers	Hoodie SZN	A Boogie W	0.558	0.792	198054	0.484	0
1srm5YnZiu	Savage	Hoodie SZN	A Boogie W	0.188	0.78	169046	0.402	0
3UhRsQVrf	Come Clos	Hoodie SZN	A Boogie W	0.22	0.741	156512	0.556	0

Figure 3: Billboard 200 Dataset

3.2 Data Architecture Diagram

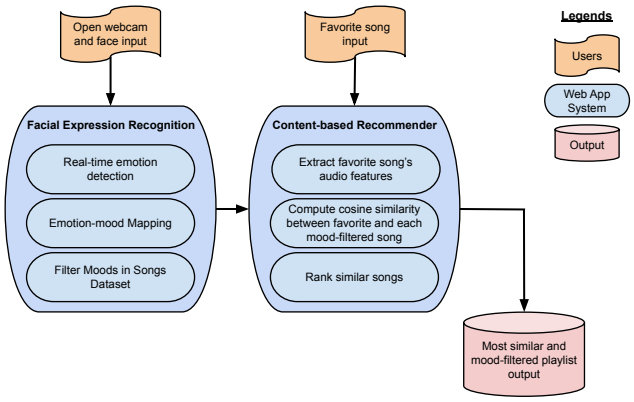

Figure 4: Software Architecture Diagram

As shown in the data architecture diagram, the system software will be deployed in a form of a web app, where two user inputs are collected, and fed into two modules of the system, emotion-mood recognition and content-based music recommendation, to output an emotion-aware recommendation playlist.

3.3 Emotion-Mood Recognition Module

Real-time Emotion Detection: First, I created a web app system using streamlit library which is essentially a web app developing tool. Then I built a real-time video processor using the facial expression analysis DeepFace and detector backend opencv to process, analyze and detect the emotion every 30 frames (about one second) and return the latest one. If the facial emotion is not detected, the system will automatically refresh the camera and pass the message to users to hold on for a clear face until an emotion is recognized. After successfully capturing and detecting the facial expression, the system will annotate the last seen emotion on the video screen using cv2 and parse the result to the next layer. To keep the project within scope, only four categories of emotions are being recognized, specifically including Happy, Sad, Angry and Neutral.

Emotion-mood Mapping: Next, the detected emotion is translated to five musical moods that will either modulate or enhance users' emotional states.

The emotion-mood mapping is based on the Russell's Circumplex Model of Emotion (2008) [9] that represents emotions as points in valence-arousal space, as in Figure 5; the Weinel et al. (2021) [4] audio features (energy, danceability translated to arousal level; according valence level translated to perceived happiness); and the Juslin et al. (2011) [5]: structural features (tempo, mode) linked to arousal/valence.

Definitions of four categories of emotions in terms of valence and arousal levels and general purpose in mapping are as follows:

- Happy (high valence, high arousal): amplify joy and energy,
- Sad (low valence, low arousal): uplift through gentle positivity,
- Angry (negative valence, high arousal): down-regulate via calm,

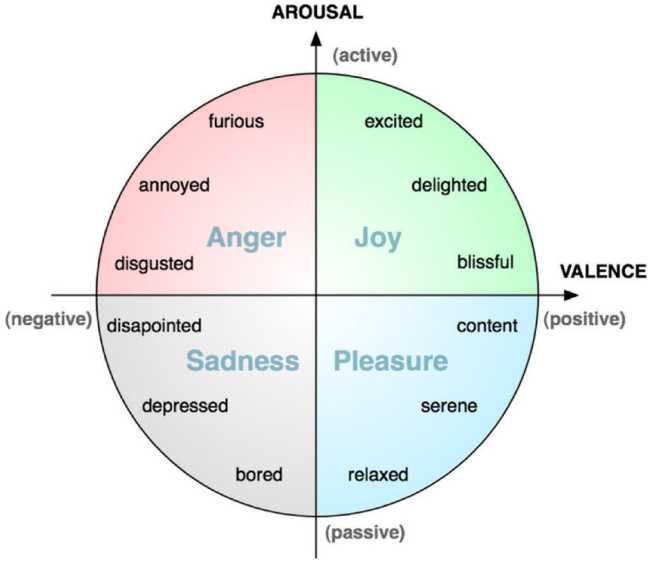

Figure 5: Circumplex model of emotion

- Neutral (mid valence, mid arousal): boost mood and engagement.

The detailed experimenting mapping results are shown in the table below.

Emotion	Musical Moods	Reasoning
Happy	energetic upbeat danceable cheerful party	fast tempo, bright timbre positive, major mode high arousal high valence social uplift
Sad	tender uplifting acoustic hopeful warm	low arousal, positive shift toward positive valence calming instrumentation emotional regulation comforting timbre
Angry	calming soothing mellow ambient relaxing	low arousal, slow tempo smooth melody, harmony gental instrumentation background textures to ease tension stress-reduction effect
Neutral	motivational bright optimistic uplifting energetic	moderate-high arousal, positive valence clear major-mode timbre foster positive affect encourage positive shift increase arousal to engage

Figure 6: Emotion-Mood Mapping

After the emotion is translated to desirable musical moods, the song database will be filtered with those only mood attributes ready for the coming music recommendation phase.

3.4 Content-based Recommendation Module

The second component of the system deals with content-based music recommendation. In this module, a second user input will be collected through the front end of the web app. Users will enter

their current favorite song that serves as the representation of their musical tastes.

Then, the system will parse the favorite song into the database and look for its audio features. All acoustic features in the database were pre-scaled beforehand using StandardScaler of scikitlearn package which is a Python module for Machine Learning purposes. The audio features that are chosen for the sake of this project are acousticness, danceability, energy, instrumentalness, liveness, speechiness, and valence.

After locating the favorite song's audio features, the system computes cosine similarity between the favorite song and each of the mood-filtered songs. It will then rank songs by similarity to the favorite track on the scale from 0 to 1 and return the top 10 similar and mood-filtered songs, which is the exact output playlist giving to the users.

4 Results

Within the first 30 testings, the system was able to detect users' emotion successfully with 98% accuracy in four categories Happy, Sad, Angry and Neutral and generate the playlist accordingly. The outputs show the emotion detected, and a list of songs with title and artist, tagged along with the moods mapped and similarity rates (Figure 7).

The average similarity rate to users' favorite track was 93% with the detected emotion being mapped to appropriate intended moods. This empirical test also finds that users start to recognize recommended songs either regulate or enhance their current emotional states.

This result means that the emotion-aware music recommender has the ability to detect users' emotions and adjust it suggestions accordingly. The outcomes represent both emotional needs and musical preferences of listeners.

However, to keep the project within scope, no thorough user-experience evaluation was conducted to test the users' satisfaction and the consistency in emotion detection performance.

Despite that fact, the system can be a great wellness or emotional-support tool for music listeners, especially those who suffer from neuro-divergence and mental health disorders. The framework can also be integrated into current streaming music apps' MRS which can automatically and constantly find and feed in users' current favorite songs and get more personalized and emotion-responsive recommendations.

5 Conclusion

The Emotion-aware Music Recommendation System is a novel approach in emotion-based MRSs, addressing the gap in bringing complexities in human emotion to song recommender, by introducing more scientific grounded solutions (MER and valence-arousal model).

On the outlook, future works that can be done to enhance the system include:

- user experience and satisfaction testing,
- expanding to more emotions being detected (fear, disgust, surprise etc.) while increasing the complexity of detected emotions,

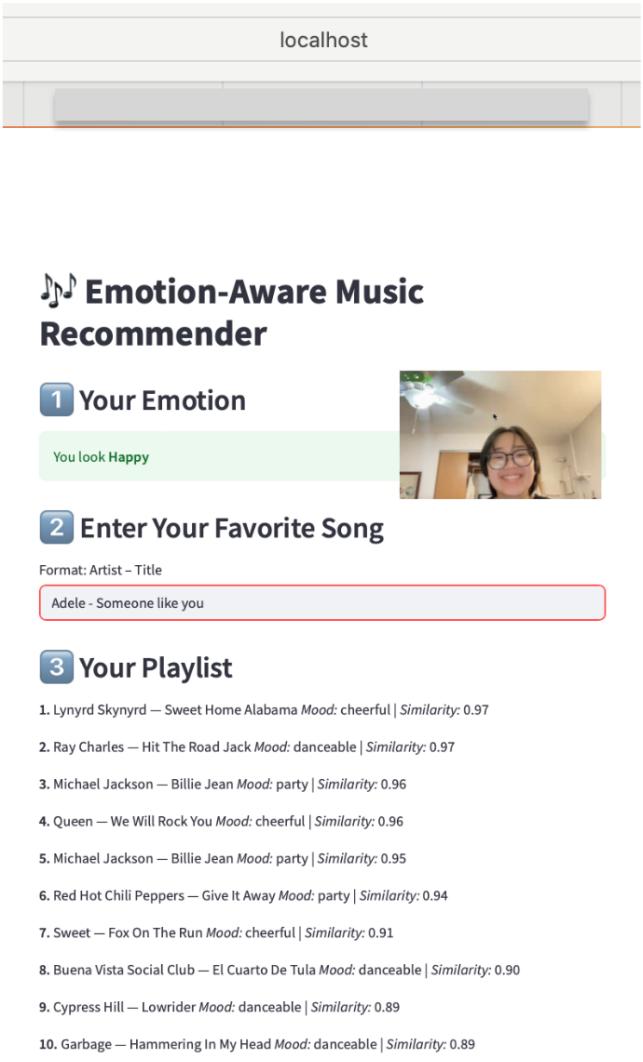

Figure 7: Sample run-through of the web app

- incorporate additional features such as related artists and lyrics to enhance users' preferences,
- real-time continuous emotion recognition and new playlists generation accordingly.

6 Acknowledgements

I would like to thank Dr. Charlies Peck and Dr. Yunting Yin from the Computer Science Department at Earlham College for their support and instruction in this Senior Capstone project. In addition, the researcher is also grateful for her immense work in completion of this research.

References

[1] Subhajit Dey Sarkar, Saptarsi Goswami, Aman Agarwal, and Javed Aktar. 2014. A novel feature selection technique for text classification using Naive Bayes. *International scholarly research notices* 2014, 1 (2014), 717092.

[2] Samira Ebrahimi Kahou, Vincent Michalski, Kishore Konda, Roland Memisevic, and Christopher Pal. 2015. Recurrent neural networks for emotion recognition in video. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*. 467–474.

[3] Mehmet Erdal, Markus Kächele, and Friedhelm Schwenker. 2016. Emotion recognition in speech with deep learning architectures. In *Artificial Neural Networks in Pattern Recognition: 7th IAPR TC3 Workshop, ANNPR 2016, Ulm, Germany, September 28–30, 2016, Proceedings 7*. Springer, 298–311.

[4] Darryl Griffiths, Stuart Cunningham, Jonathan Weinel, and Richard Picking. 2021. A multi-genre model for music emotion recognition using linear regressors. *Journal of New Music Research* 50, 4 (2021), 355–372.

[5] Patrik N Juslin and John Sloboda. 2011. *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

[6] Marius Kaminskas and Francesco Ricci. 2012. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review* 6, 2-3 (2012), 89–119.

[7] Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. 2010. Music emotion recognition: A state of the art review. In *Proc. ismir*, Vol. 86. 937–952.

[8] Fang-Fei Kuo, Meng-Fen Chiang, Man-Kwan Shan, and Suh-Yin Lee. 2005. Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th annual ACM international conference on Multimedia*. 507–510.

[9] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[10] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7 (2018), 95–116.

[11] Sefik Ilkin Serengil and Alper Ozpinar. 2021. Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 1–4.

[12] Yi-Hsuan Yang and Homer H Chen. 2012. Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3, 3 (2012), 1–30.

[13] Marcel Zentner, Didier Grandjean, and Klaus R Scherer. 2008. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 4 (2008), 494.