JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EMOTION-BASED MUSIC RECOMMENDATION SYSTEM

[1]Chandana B, [1]B Shraddha Shetty, [1]Kavya, [1]Lavanya A, [2]Dr Jithendra PR Nayak

[1] Student, [2] Professor,
[1] Department of Computer Science & Engineering,
[1] Srinivas Institute of Technology, Mangalore, India

Abstract : This work proposes a smart system for detecting emotions from audio input, capable of analyzing either live-recorded or uploaded speech to determine emotional tone and recommend music accordingly. The system uses deep learning models and sophisticated signal processing techniques to extract pertinent auditory data like pitch, energy, and spectral qualities in order to categorize emotions like happy, sorrow, rage, and neutrality. When an emotion is detected, a music recommendation engine correlates the user's emotional state with a carefully chosen selection of songs that fit their mood, improving user experience and emotional control. Applications for the suggested system include human-computer interaction, mental health monitoring, and individualized entertainment. It provides a fun link between music therapy and affective computing. The results of the experiment show that users are satisfied with emotion-matched music playback and that emotion categorization accuracy is good.

IndexTerms **- Emotion recognition, speech analysis, music recommendation, deep learning, affective computing.**

I . INTRODUCTION

The topic of affective computing has grown rapidly in recent years, especially in terms of comprehending and reacting to human emotions in a variety of ways. Speech is a particularly rich and organic channel for communicating emotional information among these. Vocal tone, pitch, rhythm, and intensity are all intrinsically linked to human emotions, which can be efficiently examined by computer techniques to yield insightful information. At the same time, music has long been acknowledged as a potent medium for expressing and affecting human emotions. This study bridges these two fields by presenting a novel system that uses spoken audio, either recorded live or uploaded as a file, to identify emotional states. The system then plays music that corresponds with the emotion it has detected. The growing need for emotionally intelligent technology that not only comprehend consumers but also improve their emotional health is what inspired this approach. Understanding and reacting to emotional cues can have a big impact on user pleasure and engagement in applications ranging from mental health care to personalized entertainment.

Emotion identification and emotion-driven music suggestion form the two main pillars of the suggested solution. To identify main emotional states like happiness, sadness, rage, and neutrality, the emotion detection module uses deep learning classifiers in conjunction with audio feature extraction techniques including pitch tracking, energy analysis, and spectral characterization . After identifying the user's emotional state, a music recommendation engine chooses songs from a carefully curated collection that best fit their present mood. Depending on the application environment, the engine's goal may be to either reinforce or regulate emotional states.

The system's design, implementation, and evaluation are covered in this paper, with particular attention paid to how well it can identify emotions from audio inputs and play music that is appropriate for the mood. The study's findings show encouraging accuracy in classifying emotions and encouraging user comments, establishing this system as a significant addition to the fusion of multimedia experiences and affective computing.

II . RELATED WORK

Sn No	Paper title	Authors & Year of publication	Methodologies advantage and disadvantages
1.	Content-based Recommender Systems	Pasquale Lops, Marco de Gemmis and Giovanni Semeraro. 2010	Advantages: Learning of profile is made easy. Quality improves over time. Considers . Disadvantages: Does not completely Overcome the problem of over-specialization and serendipity.
2.	Hybrid Recommender Systems	Robin Burke 2010	Advantages: The survey shows combine techniques for improved performance. It improves the user preferences for suggesting items to users.
3.	Association rule Mining for recommendation system on the book sale	Luo Zhenghua. 2012	Advantages: The website based on this has shown great performance. Disadvantages: It does not recommend quality content to the users. Does not consider new user cold start problem Not very efficient in terms of performance.
5	Collaborative filtering for recommender systems: User-based and Item-based CF	Gilbert Badaro, Hazem Hajj, Wassim El-Hajj and Lama Nachman. 2013	Advantages: solves the problem of finding the ratings of unrated items in a user-item ranking matrix. It improves the data sparsity problem. Disadvantage: It does not consider the demographic features which would give better results and solve the user cold-start problem
5.	Content Based Filtering, Collaborative Filtering, and Association Rule mining.	Anand Shanker Tewari, Abhay Kumar, and Asim Gopal Barman. 2014	Advantages: It considers Various parameters like content & quality of the book by doing collaborative filtering of rating of other buyers. It does not have performance problems. It builds the recommendation offline. Disadvantage: It still lacks the new user cold-start problem
6.	Non Personalized Recomm - ender Systems and User based Collaborative Recomm ender Systems	Anil Poriya, Neev Patel, Tanvi Bhagat, and Rekha Sharma. 2014.	Advantages: The system helps users find items they want to buy from a business. It overcomes the lack of personalization involved with non-personalized recommender systems. It is domain independent. Disadvantages: The recommendations are not very specific. It still lacks personalization. The computational time is low.

III.THEORIES OF EMOTIONS AND MUSIC

Understanding the relationship between emotion and music is essential to designing and developing efficient computational models for MER. Many aspects of music have been shown to be suggestive of specific feelings. Certain musical elements, like [10], can be used to express the basic emotions :

- Tempo: How quickly a piece of music is performed.
- Mode: A composition's scale or tonal structure, such as major or minor.
- Harmony: When two or more notes are perceived to produce a nice sound at the same time.
- Pitch: A single sound's location within the whole sound spectrum.
- Interval: The discrete transition from one pitch to another (major 3rd, perfect 5th, etc.
- Articulation: The sequence in which notes are played (staccato, legato, etc.).
- Rhythm: The arrangement of beats and their grouping.
- Melody: A series of notes that are heard as one cohesive whole.
- Dynamics: The music's loudness, including variations in volume and softness.

However, distinct emotions can be expressed in comparable ways using the same property. A quick tempo, for instance, can convey fear, rage, or enjoyment. Therefore, each aspect is neither essential nor definitive in and of itself.
In spite of this uncertainty, the more indications used, the more reliable the communication becomes.

Emotion	Musical Features
Happiness	Fast tempo, Small tempo variability, Major mode, Simple and consonant harmony, Medium-high sound level, Small sound level variability, High pitch, Ascending pitch, Perfect 4th and 5th intervals, Staccato articulation
Sadness	Slow tempo, Minor mode, Dissonance, Low sound level, Moderate sound level variability, Low pitch, Descending pitch, Small intervals (e.g., minor 2nd), Legato articulation
	Legato articulation.
Anger	Fast tempo, Small tempo variability, Minor mode, Dissonance, High sound level, Small loudness variability, High pitch, Ascending pitch, Major 7th and augmented 4th intervals, Staccato articulation
Fear	Fast tempo, Large tempo variability, Minor mode, Dissonance, Low sound level, Large sound level variability, High pitch, Ascending pitch, Staccato articulation
Tenderness	Slow tempo, Major mode, Medium-low sound level, Small sound level variability, Low pitch, Legato articulation
Surprise	Fast tempo, Large tempo variability, Major mode, High sound level, Large sound level variability, High pitch, Sudden dynamic changes, Staccato articulation, Unexpected pauses or rhythmic changes

Table 1. An overview of the characteristics of music that are associated with distinct moods.

IV. METHADOLOGY

Using a four-stage framework (Fig. 3), MMER research can be summed up as follows: modality and data selection (Stage 1), feature extraction (Stage 2), feature processing (Stage 3), and emotion prediction (Stage 4). Using a four-stage framework (Fig. 3), MMER research can be summed up as follows: modality and data selection (Stage 1), feature extraction (Stage 2), feature processing (Stage 3), and emotion prediction (Stage 4).

Methods The Music Emotion Recognition System was created with a modular and user-centered approach to ensure seamless emotion identification and music recommendation. In the following stages, the system integrates modules for real-time audio processing, machine learning, and user interaction:

1. Obtaining Audio Input Live Recording: Users record audio in real time using their smartphone's microphone. The technology records the audio and streams it for immediate processing. Uploading Files: Users can upload WAV and MP3 audio files. After the system confirms the file type, the audio signal is retrieved for further examination.
2. Preparation and feature extraction The supplied audio is preprocessed to remove noise and adjust the level. Mel-frequency acoustic properties such as energy, tempo, and pitch cepstral coefficients. In order to extract chroma and spectral contrast, audio processing libraries are utilized.
3. Emotion Intelligence Engine The retrieved features are then fed into a trained machine learning or deep learning model (e.g., CNN, LSTM, Random Forest). Using labeled datasets (such RAVDESS and CREMA-D), the model is trained to classify emotions into classes like happy, sad, furious, and calm.

4. The technique offers instantaneous emotion feedback for live audio through real-time processing with lower latency. The mechanism that suggests songs Following the identification of an emotion, the recommendation algorithm selects music from a tagged music database.
5. Music is categorized based on its mood, genre, tempo, and degree of intensity. The system may suggest music based on how you're feeling right now. Contrasts are used to balance the mood. Enhancement of the emotion (e.g., enhancing gloomy moods). 5. Personalization and User Information Users can log in and create profiles in order to save: Favorite musicians or genres: Favorite moods based on emotional reaction history These profiles are used to personalize music recommendations in the future. The Visualization Module
6. An engaging user experience is provided via a real-time visualization tool that displays wave forms or spectrograms of the incoming audio. The visual feedback is updated instantly during live input.
7. Admin module management Administrators are responsible for overseeing the music database.There are tools available for system analytic, troubleshooting, and usage tracking.
8. Ongoing Education and Input Users can rate how relevant the suggested playlists are. The saved feedback is used to retrain the recommendation model to make better recommendations in the future.
9. Security and Error Management The system is made up of: Authentication of user accounts To get rid of faulty or unsupported files, use file validation. encrypting user data to safeguard privacy. Protection from malicious uploads
10. Cross-Platform Accessibility The system is built using responsive web technologies to allow: Workstation: A tablet Mobile device platforms It is certain that every device will have a consistent and user-friendly interface.

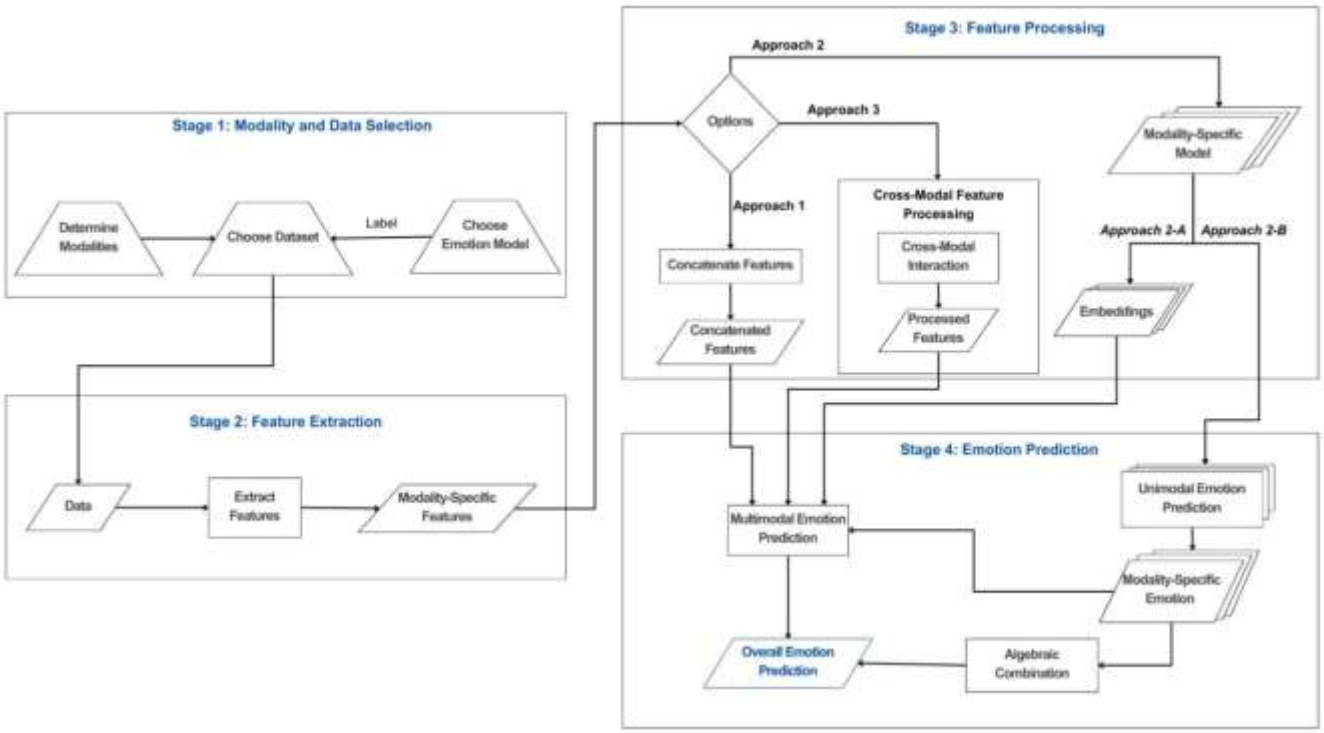

Fig 1. A framework that summarizes both past and present MMER approaches.

IV. EXPERIMENTAL RESULT

The RAVDESS, which comprises 1440 audio samples tagged with eight different emotions—neutral, calm, happy, sad, angry, afraid, disgusted, and surprised—was used in our studies to assess the effectiveness of our emotion detection system. Using a mix of spectral contrast, chroma features, and Mel-Frequency Cepstral Coefficients (MFCCs), we preprocessed the audio data to extract the speech signals' spectral and temporal properties. We used a hybrid model that included Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) for classified tasks. Local time-frequency patterns were extracted using CNN layers, while the audio signal's sequential dependencies were modeled by LSTM layers. With a batch size of 32 and the Adam optimizer, the model was trained for 50 epochs using categorical cross-entropy loss.

80% of the dataset was used for training, and 20% was used for testing. The model's overall accuracy on the test set was 84%. For emotions like happiness (F1-score: 0.90), anger (F1-score: 0.87), and sadness (F1-score: 0.86), class-wise performance demonstrated good precision and recall; however, performance for calm and neutral emotions was marginally lower (F1-scores of 0.82 and 0.79, respectively). Due to their auditory similarity and difficulty in differentiation, calm and neutral were frequently misclassified, according to a confusion matrix. These findings show that the model can reliably identify emotions in speech, especially when those emotions are more expressive.

V. CONCLUSION

The goal of the developing discipline of music emotion recognition (MER), especially in its multimodal form (MMER), is to bridge the gap between human emotional comprehension and machine perception. Despite the encouraging outcomes of unimodal techniques, which are mainly focused on audio, contemporary MMER systems still per

form below human levels because of issues with emotional modeling, dataset variability, and established benchmarks. Recent developments have great promise for capturing complex emotional impressions, such as the combination of individualized meta-learning techniques (e.g., DSAML) with dual-scale attention models. The greatest outcomes to date, attained with a few number of well chosen features, highlight the significance of both high-level and audio-based multimodal feature design. Developing emotionally meaningful features, building individualized and user-centric systems, and growing multimodal datasets (such as audio, lyrics, MIDI, and physiological inputs) should be the top priorities for future research. With improved real-time applications in fields including sports, healthcare, recommendation systems, and human-computer interaction, MER is a field with enormous multidisciplinary promise and societal influence.

VI. FUTURE WORK

To advance the *Music Emotion Recognition System* in future projects, numerous important improvements can be implemented to enhance the system's precision and engage users more effectively. One notable approach is to incorporate *pre-trained deep learning models—such as CNNs, RNNs, or transformers like **Wav2Vec, **YAMNet, or **MusicBERT*—that have been developed on extensive music or audio emotion datasets. These models can offer powerful and context-sensitive feature extraction, enhancing the system’s capacity to accurately categorize the emotional tone of songs, even amidst noise or different musical styles.

Rather than creating models anew, we will investigate and assess the performance of these existing pre-trained models within our framework, possibly fine-tuning them using domain-specific data to better interpret emotions across various musical genres.

In addition, the user experience can be enhanced by developing *dynamic, emotion-driven playlists. Instead of static lists, the system can create *emotion-adaptive playlists* that change automatically in response to real-time emotion detection or user mood entries. For instance, if a user's identified mood transitions from sad to energetic, the playlist will gradually shift to include songs that align with this emotional transition. This would make the listening experience feel more personalized and engaging. Moreover, integrating *user feedback mechanisms*—where the system learns from tracks that are skipped or liked—can assist in improving future playlist generation, fostering a smarter and more responsive recommendation system. These improvements aim to transform the Music Emotion Recognition System from merely a classification tool into a comprehensive and continuously evolving platform for emotional and musical interaction.

VI. REFERENCES

[1] Yang, Y.-H., Lin, Y.-C., Su, Y.-F., and Chen, H. H. (2008). A Regression Approach to Music Emotion Recognition. IEEE Transactions on Audio, Speech, and Language Processing, 16(2):448–457. This study proposes a dimensional method for recognizing music emotions through regression models, focusing on representing emotions in a continuous space (e.g., valence-arousal), which is crucial for music emotion recognition systems.

[2] Hu, X., and Downie, J. S. (2010). When lyrics outperform audio for music mood classification: a feature analysis. In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010), 619–624. This research conducts a comparative analysis of lyrical and audio features for detecting mood, underscoring the relevance of multimodal data, which is highly significant in contemporary music emotion recognition approaches.

[3] Christopher Akiki and Manuel Burghardt. 2021. MuSe: The musical sentiment dataset. Journal of Open Humanities Data, 7, 10. This work introduces a high-quality sentiment dataset specifically designed for music analysis, which is essential for developing essential for developing and assessing modern emotion recognition models. It is contemporary and directly

[4] Anna Aljanaki, Yi-Hsuan Yang, and Mohammad Soleymani. 2017. Developing a benchmark for emotional analysis of music. PLOS ONE, 12(3): e0173392. This article presents a solid benchmark for emotional analysis, tackling the deficiency of standard datasets and metrics, which is critical for comparing models and ensuring reproducibility in music emotion recognition research.

[4] Shashidhar G. Koolagudi, Ramu Reddy, Jainath Yadav, K. Sreenivasa Rao. "IITKGP-SEHSC: Hindi speech corpus for emotion analysis." IEEE (2011). This serves as a foundational reference for any research concerning Hindi speech emotion recognition.It offers a well-known labeled dataset vital for training and benchmarking speech emotion recognition models, especially pertaining to Indian languages

[6] Mehmet Cenk Sezgin, Bilge Gunsel, and Gunes Karabulut Kurt. "Perceptual audio features for emotion detection." EURASIP Journal on Audio, Speech, and Music Processing, 2012. This paper emphasizes audio feature-based methods for emotion detection, aligning with technical aspects of audio processing relevant to both speech and music emotion recognition systems.

[7] Fang, J., Grunberg, D., Luit, S., & Wang, Y. (2017). Development of a music recommendation system for motivating exercise. 2017 International Conference on Orange Technologies (ICOT), IEEE. This research discusses a practical application of a recommendation system that connects music selection to user context (exercise motivation), providing a strong real-world illustration of contextual or goal-driven recommendation.

[8] Nakamura, K., Fujisawa, T., & Kyoudou, T. (2017). Music recommendation system using lyric network. In 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE). This study showcases an innovative approach that utilizes lyrics as a network, introducing a content-based recommendation layer suitable for applications that involve emotional or sentiment analysis within music.

[9] A. Nediyanchath, P. Paramasivam, P. Yenigalla, presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Utilizing multi-head attention for recognizing emotions in speech with supplementary learning of gender recognition (IEEE, Barcelona, 2020), pp. 7179–7183.

[10] C.H. Park, D.W. Lee, K.B. Sim, explored emotion recognition in speech based on RNN. Nurse Lead. 4, 2210–2213 (2002). https://doi.org/10.1109/ICMLC.

[11] J. Niu, Y. Qian, K. Yu, in the 9th International Symposium on Chinese Spoken Language Processing, discussed acoustic emotion recognition through deep neural networks (IEEE, Singapore, 2014), pp. 128–132.

[12] Q. Mao, M. Dong, Z. Huang, Y. Zhan, focused on learning prominent features for speech emotion recognition via convolutional neural networks. IEEE Trans. Multimedia 16(8), 2203–2213 (2014).

[13] J. Lee, I. Tashev, presented in the Proceedings of Interspeech 2015 on high-level feature representation for speech emotion recognition using recurrent neural networks (ISCA, Dresden, Germany, 2015).

[14] M.A. Jalal, E. Loweimi, R.K. Moore, T. Hain, in the Proceedings of Interspeech 2019, discussed learning temporal clusters through capsule routing for speech emotion recognition (ISCA, Graz, 2019), pp. 1701–1705.

[15] R. Shankar, H.W. Hsieh, N. Charon, A. Venkataraman, in the Proceedings of Interspeech 2019, investigated automated emotion morphing in speech utilizing diffeomorphic curve registration and highway networks (ISCA, Graz, 2019), pp. 4499–4503.

[16] S. Siriwardhana, T. Kaluarachchi, M. Billinghurst, S. Nanayakkara, researched multimodal emotion recognition using transformer-based self-supervised feature fusion. IEEE Access 8, 176274–176285 (2020).

[17]S. Costantini, G. De Gasperis, P. Migliarini, in the 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), explored the engineering of multi-agent systems for emphatic human-robot interaction (IEEE, Sardinia, Italy, 2019), pp. 36–42.

[18]H. Okon-Singer, T. Hendler, L. Pessoa, A.J. Shackman, discussed the neurobiology of emotion-cognition interactions, addressing essential questions and strategies for future inquiry. Front. Hum. Neurosci. 9, 58 (2015).

[19]Q. Ma, D. Guo, examined the brain mechanisms underlying emotion. Adv. Psychol. Sci. 11(03), 328 (2003).

[20] S. Lee, S. Yildirim, A. Kazemzadeh, S. Narayanan, presented an articulatory analysis of emotional speech production at the Ninth European Conference on Speech Communication and Technology (ISCA, Lisbon, Portugal, 2005).