EEG-Based Emotion Recognition using Beta and Gamma Rhythms With SVM and QBC Classifiers for Emotion-Aware Music Recommendation

YuLin Han, RuiHao Zeng*
School of Music, Nanchang Institute of Technology, Nanchang, Jiangxi, China
E-mail: 15397918093@163.com
*Corresponding author

Keywords: emotion induction, internet of things technology, western music, emotion feature learning, EEG classification, beta rhythm, gamma rhythm, SVM classifier, QBC algorithm, affective computing

Received: April 7, 2025

Conventional music recommendation systems, based on user preferences or genre categories, fall short when it comes to emotionally engaging users according to their personality types. To address this gap, our study investigates EEG-based emotion recognition using Western music in combination with Internet of Things (IoT) technology. Guided by the dimensional model of emotion, we selected three types of Western music fragments designed to evoke emotions induction. EEG signals were then collected and analysed across different brainwave rhythms—theta, alpha, beta, and gamma—using various machine learning classifiers for feature extraction and emotion classification. Our results show that the beta and gamma rhythms produced the highest classification accuracy, with overall averages of 0.842 and 0.841, respectively. Among the classifiers, Support Vector Machine (SVM) outperformed the others, achieving a between-subject accuracy of 95.7% on gamma and 88.2% on beta rhythms, marking a notable improvement over baseline methods. Similarly, the Query-by-Committee (QBC) algorithm achieved up to 90.1% accuracy in gamma and 88.4% in beta rhythms. These findings highlight the potential of SVM and QBC classifiers in improving EEG-based emotion recognition. Interestingly, the most effective EEG features consistently originated from the head region across all participants.

Povzetek: V prispevku je opisano, kako nov algotirem z EEG beta/gama ritmi omogoči kvalitetno prepoznavo čustev ob poslušanju zahodne glasbe.

1 Introduction

Most of the common emotion measures are completed by the subjects' subjective emotion label selection [1]. However, subjective scales are affected by aspects such as word comprehension and may not fully express the true experience of emotions. EEG signals are relatively more feasible and accurate for emotion recognition by reflecting the activation state of human brain functions [2]. Research on user emotions has been the focus of multiple disciplines. Numerous studies in neuroscience, cognitive science, and biology have shown that emotions play a critical role in rational and intelligent behavior. A person's emotional state not only affects an individual's attention span but also has an impact on the ability to solve problems and make decisions [3]. Finding and calculating user emotions has important practical application value. Adding user emotions when profiling users can make user portraits more accurate, so as to provide users with better personalized services and improve users' quality of life.

Although virtual reality (VR) is increasingly recognized as an effective emotion-inducing technology, little research has been done on the relationship between the two. Chirico and Gaggioli [4] introduced 50 people to participate in the experiment to compare the emotional type and valence under different conditions. The results of the study showed that there was no significant difference in the emotions evoked by virtual and natural conditions. How do humans distinguish between emotions and non-emotions? Using the discriminative psychophysical MRI sparse sampling paradigm to locate threshold responses to happy and sad acoustic stimuli, Manno et al. [5] found that threshold emotion recognition in Western music exploits fine structural cues. Western musical anhedonia means the acquired and selective loss of Western musical emotion. The study of Masayuki and Satoh [6] found that there were cases in which Western music emotion was preserved even in the presence of impaired Western music perception and recognition, findings consistent with the findings of activation studies using PET and fMRI [7]. However, it is necessary to explain and explain the experimental results on the basis of understanding the characteristics and limitations of the case. Evans [8] conducts empirical research to explore potential connections between the emotional significance attributed to musical stimuli (expressed emotion, or external locus of emotion) and the personal emotional response triggered by listening to music (felt emotion, or internal locus). It is commonly assumed that the relationship between these two emotional loci is positive, implying that the emotional experience elicited while listening to music aligns with the emotion expressed by the music itself. Soulier [9] aimed to study the effect of negative emotion induction in Western music on the lexical spelling performance of children with and without written language impairment. Studies have found that the effects of emotion vary by spelling level and only affect the performance of children with written language impairments. Kabrin et al. [10] I incorporated a new technology for audiovisual induction of states of consciousness. The aim of the study was to demonstrate that this technique can induce a relaxed state. He conducted two experiments using synchronous fractals and specific configurations of Western music sequences, which showed that a relaxed state could be induced. The study of Vlker [11] investigated the effects of personal Western music and a researcher's pre-selected Western music on inducing sadness and happiness. Results indicated that participants' choice of Western music had a stronger effect on reported mood, with sadness and happiness mainly elicited by contagion and episodic memories associated with Western music. Wang et al. [12] used the V - A model as the emotional perception model, selected about 1000 classic extracts from China and the West, and finally extracted about 20 Vit of different data sets and different emotional dimensions. Valence (V) dimension represents the positivity or negativity of an emotion. Emotions with positive valence are associated with positive feelings like happiness, joy, or love, while emotions with negative valence are associated with negative feelings such as sadness, anger, or fear. While Arousal (A) refers to the intensity or activation level of an emotion. Emotions with high arousal are intense and stimulating, like excitement or anger, while low arousal emotions are more subdued, like calmness or boredom. The results show that the packaging method combining MaxAbsScaler preprocessing and recursive feature elimination algorithm based on maximum random tree is the best algorithm. Harmonic change detection function is a universal feature of culture, while spectral flux is a cultural specificity of Chinese classical music. The study also found that the pitch characteristics of western classical music are more significant, while the loudness and rhythm characteristics of Chinese classical music are more significant. Sentiment analysis methods are used to evaluate general opinions expressed in tweets about entities. Seghouani et al. [13] proposed a new method to determine entity reputation based on the set of events involved in the entity and also proposed a new sampling method driven by tweet weighting metric to provide better target entity quality and summarization. To sum up, the changes and activations of emotions are due to the involvement of all levels of the nervous system. The initiation and occurrence of emotions are the result of the integration of nervous systems at all levels and are the product of multiple levels of physiology.

The role of internet of thing (IoT) in emotional induction is becoming more and more important. Tallapragada et al. [14] integrates data intelligence through IoT to track customer sentiment and provide customer behavioral insights, the proposed system uses model-based face and emotion tracking under real use case conditions. Wu and Zhang [15] carried out a detailed analysis and research on the necessity, feasibility and implementation of intelligent home voice emotion recognition technology, introduced the definition and classification of emotion, and proposed five main emotions to be recognized for voice emotion recognition based on intelligent home environment. Then, on this basis, we analyze the methods of obtaining emotional voice data. Building upon this foundation, the study delves into key aspects of voice data acquisition within smart home environments. It addresses issues like voice characteristics, acquisition methods, and more. Additionally, the study proposes three fundamental principles for voice text design and identifies a suitable hybrid voice recording method tailored for smart homes. Furthermore, the study provides a comprehensive account of the design and establishment process of an emotional voice database for smart homes, offering insights into feature extraction challenges in speech emotion recognition. In the realm of Western music mood prediction, the study focuses on constructing regression models to forecast mood dimensions such as valence (happiness level) and arousal (energy level). Hu and Yang [16] evaluated a mood regression model built on fifteen acoustic features of five mood-related aspects of Western music, with a focus on generalization across datasets. Emotion is considered a physiological state that occurs whenever an individual observes a transition in their environment or body. Garg et al. [17] has conducted extensive experiments on different Western music emotion datasets and human emotion for impactful feature extraction, training, testing, and performance evaluation. The above research shows that the role of the Internet of Things in emotional induction should be further explored. This study investigates whether EEG-based emotion recognition, specifically focusing on beta and gamma frequency bands, can enhance the accuracy and personalization of music recommendation systems when compared to conventional methods that rely on musical genres or user preference similarity. The research is guided by three central questions: first, whether EEG signals in the beta and gamma bands can effectively classify emotional states (positive, neutral, negative) induced by Western music; second, whether advanced machine learning classifiers such as Support Vector Machines (SVM) and Query-by-Committee (QBC) offer performance advantages over traditional models; and third, whether integrating real-time physiological signals into recommendation systems can lead to more emotionally attuned and personalized user experiences. The details that how the data in relevant studies is used is given in table 1.

The selection of Support Vector Machine (SVM) and Query-by-Committee (QBC) classifiers was based on their proven effectiveness in EEG-based emotion recognition tasks, particularly in handling high-dimensional, non-linear, and noisy biosignals. SVM is well-established in the literature for its ability to find an optimal separating hyperplane in non-linearly separable data through the use of kernel functions, making it highly suitable for EEG features extracted from power spectral densities. QBC, an active learning ensemble method, was selected for its strength in leveraging model uncertainty to iteratively improve performance with fewer labeled samples—an important advantage in EEG studies where labeling is resource-intensive. These classifiers were chosen over alternatives such as Linear Discriminant Analysis (LDA) and k-Nearest Neighbors (k-NN), as LDA assumes linear separability and equal covariance, which are not always valid in EEG data, and k-NN is highly sensitive to noise and irrelevant features, often resulting in lower performance for high-dimensional physiological datasets.

For model evaluation, we employed five-fold stratified cross-validation to ensure a balanced distribution of emotional classes across folds. This method maintains the proportion of class labels in each fold, offering robust performance estimates while mitigating overfitting. Subject-independent validation was also considered to assess generalization across participants.

To clarify the methodology, particularly the music material selection process, this study employed a structured and empirical approach. The selection of musical stimuli was based on the widely accepted valence-arousal (V–A) emotional model. Positive-valence music fragments were characterized by an upbeat tempo and major key tonality, neutral fragments maintained a moderate tempo and balanced tonal structure, while negative-valence pieces typically featured slower tempos and minor key signatures. All selected music consisted of Western classical piano compositions to ensure cultural familiarity and emotional recognizability among participants. This focus was important, as the study also addresses the cultural dimensions of emotional induction, and using widely understood musical forms minimized potential variability in emotional interpretation.

To further validate the appropriateness of the stimuli, a pool of 30 Western classical piano pieces was initially compiled and evaluated by a group of 12 independent raters. These raters assessed each piece using a standardized scale for emotional valence. The top-rated pieces in each emotional category—positive, neutral, and negative—were then selected for use in the EEG-based experiments. This rigorous pre-screening ensured that the chosen music fragments were emotionally distinct, psychologically valid, and culturally consistent with the study’s focus. By clearly defining the research goals and ensuring transparency in the stimulus selection process, this study addresses previous gaps in the literature and provides a reliable framework for EEG-based emotion recognition in the context of personalized music recommendation.

A research gap exists in the current body of studies related to emotion recognition, as there is a limited exploration of cross-cultural emotional recognition. The existing research predominantly focuses on Western contexts, neglecting the potential variations in emotion recognition across different cultures. Investigating the adaptability of IoT-based emotion recognition systems to diverse cultural contexts and understanding how cultural nuances influence emotion recognition is essential for achieving more accurate and globally applicable emotion

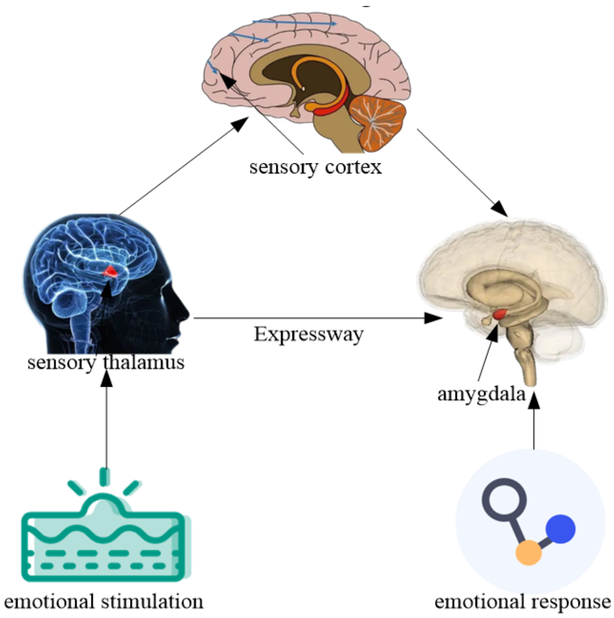


Figure 1: Two-loop model of emotion

recognition technology.

In this experiment, we selected 9 Western music clips as experimental stimuli based on their average scores, each clip lasting approximately 11 seconds, and repeated each stimulus 27 times. The experiment had a minimum theoretical duration of 44.5 minutes. Notably, the experimental data revealed that emotional arousal in Western music scenes, including joy, sadness, fear, and disgust, was slightly higher than in videos. Subjective scales offer a straightforward and cost-effective means of gathering individuals' self-reported experiences, making them accessible for large-scale studies and versatile in various fields. However, their reliance on subjective interpretation and potential biases can introduce limitations.

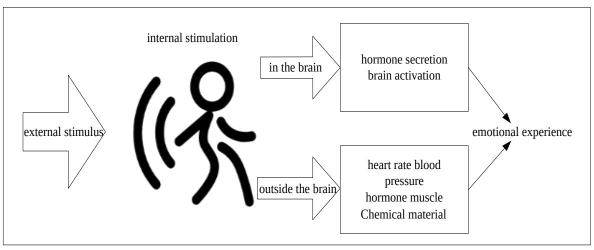


Figure 2: Physiological manifestations of emotion-inducing mechanisms.

On the other hand, Electroencephalogram (EEG) provides an objective measure of brain activity with high temporal resolution, making it valuable for understanding neural processes. Yet, it has limitations in spatial resolution, sensitivity to artifacts, and complexity in interpretation. The choice between these methods hinges on research goals and practical considerations, with a potential for synergy when used in conjunction to obtain a more comprehensive understanding of both subjective experiences and underlying neural mechanisms. This study's novelty lies in the development of an Internet of Things-based emotion induction system, which evaluates emotions in visual images using image-based emotion calculation algorithms. It also contributes to the creation of an intelligent multi-vision fusion annotation model. This research highlights a gap in current studies related to emotion recognition, emphasizing the need to explore cross-cultural emotional recognition and adapt IoT-based systems for diverse cultural contexts, ultimately enhancing the universality and accuracy of emotion recognition technology.

Table 1:Data used in relevant studies.

Study	**Dataset Used**	**EEG Bands Analysed**	**Feature Extraction Methods**	**Classifiers Used**	**Reported Accuracy**	**Musical Styles Tested**	**Limitations Identified**
Wang et al. [12]	V-A model, ~1000 Chinese & Western clips	Beta, Gamma	MaxAbsScaler, RFE (Recursive Feature Elimination)	Random Forest	~85%	Chinese Classical, Western Classical	Style-specific emotional variance (pitch vs. loudness/rhythm)
Manno et al. [5]	Custom acoustic stimuli dataset	Not specified	Fine structural cue analysis	Not specified	N/A	Western Classical	No classifier or EEG band details
Chirico & Gaggioli [4]	VR vs. Natural environment, 50 participants	Not specified	Subjective emotion valence labelling	None	N/A	Western	No significant difference in emotion between VR and real settings
Vlker [11]	Personalized vs. researcher-selected playlists	Not specified	Subjective mood reporting, memory/emotional contagion	None	Qualitative results	Western (Personal, Researcher-selected)	No EEG or quantitative classifier used
Soulier [9]	Children with/without writing impairments	Not specified	Lexical performance analysis	None	Performance-based	Western (Negative Emotion)	Limited to children; no EEG-based emotion data
Our Study	EEG-based, 3-class emotional states (neutral, positive, negative)	**Beta, Gamma**	Wavelet + Statistical features	**SVM, QBC**	**88.2%–95.7%**	Western (Emotionally labelled clips)	Focus on personality-specific induction; advanced classifiers used

2 Principles and methods of emotional induction

A) Principles related to emotion induction

1) The brain mechanism of emotion

Traditionally, it is believed that information from emotional stimuli is transmitted to the limbic system centered on the thalamus, hippocampus, and amygdala, from where emotions are generated and expressed, but the limbic system is structurally and functionally challenging [18]. The two-loop model of emotion is shown in Figure 1.

As shown in Figure 1, emotional stimuli are transmitted from the sensory organs through the cortex of the sensory thalamus to the amygdala, which immediately produces innate gross emotions. Simultaneously, stimuli are transmitted from the sensory organs through the cortex of the sensory thalamus to higher-order regions such as the frontal lobe and hippocampus, which process information and transmit the processed information to the amygdala, producing micro-emotions [19].

2) The role of auditory system in emotion induction

When responding to external stimuli, individuals automatically develop emotional responses. These responses are manifested through physiological and behavioral responses, and these responses vary from situation to situation. The physiological material manifestation of the emotion-inducing mechanism is shown in Figure 2. As shown in Figure 2, in contrast, vocal induction is typically performed using the International Affective Digital System (IADS). An Affective Digital System is a technology or software designed to recognize, interpret, and respond to human emotions and emotional cues. These systems employ various sensors, algorithms, and artificial intelligence to detect and analyze emotional states in individuals and, in some cases, even respond with appropriate emotional expressions or actions. Affective digital systems find applications in areas like human-computer interaction, virtual reality, healthcare, marketing, and more, with the goal of creating more empathetic and responsive digital experiences. They are often used in emotion recognition, sentiment analysis, and affective computing to improve the interaction between humans and technology. In a comparative study on the effects of visual and auditory affective induction, physiological and behavioral affective responses during acoustic and visual induction were used to compare the effects of inducing media [20].

3) User emotion recognition framework

The user's listening history contributes to the creation of a personalized song list. To achieve real-time emotion recognition, it's imperative to initially discern the emotions associated with each song in the user's playlist. Building upon this concept and integrating the user's historical listening patterns, an emotion recognition algorithm is devised to compute the user's emotional response while listening to music. The resultant emotional states during music playback are categorized into distinct feelings, including healing, relaxation, romance, nostalgia, excitement, loneliness, and tranquility. Figure 3 illustrates the framework for user emotion recognition.

Figure 3 illustrates a framework comprising three primary components. Initially, it involves the acquisition of historical user data generated during their song-listening activities. Typically, users compile a listening song list after each session, which essentially forms a playlist. It's important to note that the emotional content associated with songs in these playlists may or may not be readily available.

B) Sentiment feature learning based on matrix factorization

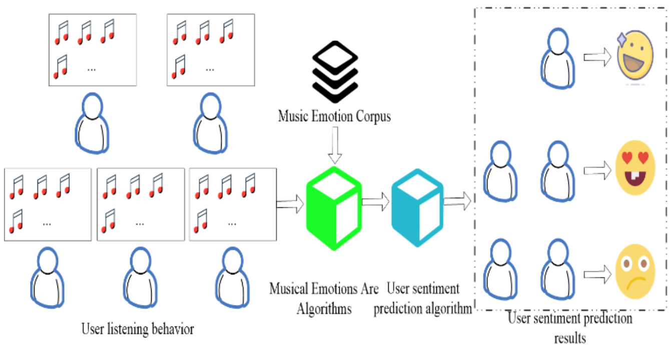


Figure 3: User emotion recognition framework.

This section will introduce in detail how to learn the emotional representation of songs by using the features of songs and playlists based on the association information between playlists and songs and the user's emotional annotation information on playlists.

1) Constrained non-negative matrix factorization

In recent years, the Non-negative Matrix Factorization (NMF) technique has become a popular method for data representation [21]. It appeared to deal with the problem that the input data dimension is too high in the real world. After characterizing the raw data in matrix form, two non-negative matrices can be found by NMF technique. Their product approximates the original matrix well, thereby simultaneously mapping features in both dimensions into a hidden space [22].

In the context of the identified research gap, Figure 3 presents a framework consisting of three key elements. To address the gap related to cross-cultural emotional recognition and IoT-based systems, there is a crucial need to collect historical data generated by users as they engage in music listening. Users typically create playlists following their listening sessions, which serve as valuable sources of data. It's essential to recognize that the emotional attributes of songs within these playlists may not always be well-documented, and exploring cross-cultural variations in emotion recognition can help enhance the understanding and adaptation of IoT-based emotional recognition systems to diverse cultural contexts.

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad (1)$$

Matrix decomposition goal: NMF aims to find two non-negative decomposition matrices of the original matrix X to replace U and V, and make the original matrix X and the decomposed result as close as possible, that is, to minimize the following objective function:

$$O = \|X - UV^T\|_F^2 \quad s.t\ U \geq 0, V \geq 0 \quad (2)$$

Since there is no orthogonal constraint here, what we end up with is a representation of the distribution of playlists and songs in emotional space [23].

2) Constrained non-negative matrix factorization incorporating external information

In recent years, many researchers have proposed many methods to incorporate external information into NMF to improve the effect of matrix factorization. This external information are features related to matrix rows or columns [24]. External information is usually incorporated into the objective function of NMF in the form of a regularization term.

Emotional label information: Regarding the emotional depiction of Western music, expert-labeled emotional data for song lists or individual songs is integrated into the matrix decomposition as external information. To illustrate this, let's delve into an example using relevant formulas that involve the inclusion of expert-tagged song information. This injection process mirrors the approach of variable V, where the relevant U component is substituted to accommodate the external emotional label data.

$$||G_v(V - V_0)||_F^2 \quad (3)$$

Among them, G_v is the diagonal indicator matrix representing the song level, G_v(i, i)=1 represents that the i-th song contains emotional indicators, and G_v(i,i)=0 if it does not. If the sentiment representations learned by matrix factorization are inconsistent with sentiment indications contained in the data, this loss function will penalize, thereby optimizing the direction of the next step in matrix factorization [25].

Relationship network: The adjacency matrix of a graph can be defined as:

$$W^u(i,j) = \begin{cases} 1, if\ u_i \in N(u_i)\ or\ u_j \in N(u_i) \\ 0, \quad otherwise \end{cases} \quad (4)$$

u_i is a song in the playlist and N(u_j) is the k nearest neighbors of song j. The neighbors here can be obtained by many metrics and further build the adjacency matrix. The loss function of the playlist relationship network can be defined as follows:

$$R_u = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}||U(i,*) - U(j,*)||_2^2\, W_u(i,j)$$

$$= Tr(U^T(D^u - W^u)U) = Tr(U^T L^u U) \quad (5)$$

Among them, Tr( ) is the trace of the matrix, $L^u = D^u - W^u$ is the Laplacian matrix of the constructed song list graph network, D^u is a diagonal matrix, and D^u (i,i)=$\sum_{j=1}^{m} W^u$(i,j). If two playlists are close in the graph but have different sentiment labels, this loss function will penalize. The relationship network definition of the song dimension is similar to that of playlists, just replace the corresponding U with V.

$$sim(i,j) = \frac{V_i V_j}{||V_i||V_j||} \quad (6)$$

sim (i,j) represents the similarity between playlist i and playlist j, and V_i and V_j represent the vectorized representation of playlist i and playlist j in the feature space, respectively. Based on song co-occurrence: A playlist contains multiple songs at the same time, and the same song may also be included in multiple playlists. Playlist similarity is measured by calculating the number of co-occurring songs in different playlists. The more songs appear in the playlist at the same time, the closer the two playlists are. Finally, the objective function of the constrained non-negative matrix factorization algorithm incorporating external information can be expressed as:

$$\min J = ||X - UV^T||_F^T + \alpha_l^u||G^u(U - U_0)||_F^2 + \alpha_l^v||G^v(V - V_0)||_F^2 + \alpha_c^u Tr(U^T L^u U) + \alpha_c^v Tr(V^T L^v V) \quad (7)$$

Considering the actual meaning of data representation, negative numbers cannot appear in matrices U and V.

3) Western music emotional representation learning

Let the emotional distribution of the i-th song be V, then the song i in the k-dimensional emotional space is expressed as:

$$V_i = (V_{i1}, V_{i2}, \ldots, V_{ik}) \quad (8)$$

The V corresponding to the i-th song is normalized, and the emotional probability distribution s_i^* corresponding to the song i is obtained after normalization. The calculation formula is as follows:

$$s_i^* = \left(\frac{V_{ij}}{\sum_{j=1}^{k} V_{ij}}\right) \quad (9)$$

Among them, k represents the emotional space dimension. For each song, the emotion category with the highest corresponding value is taken as the emotion category to which the song belongs, and the calculation formula is as follows:

$$e_i^* \leftarrow argmax(s_i^*) \quad (10)$$

This achieves the purpose of song emotion recognition.

(4) Western music joint representation learning based on Encoder-Decoder

In this study, the encoder and decoder of Seq2Seq are a Recurrent Neural Network (RNN), respectively. In the RNN structure, the input of the i-th layer neuron at time m includes not only the output of the i-1 layer neuron at time t, but also its own output at time t-1. Assuming that the source modality is the audio modality, and the target modality is the lyrics modality, the output of the t-th hidden layer is expressed as:

$$h_t = RNN(h_{t-1}, X_t^A) \quad (11)$$

All hidden layers of the encoder RNN are spliced together to form the output of the encoder, which is expressed as:

$$\epsilon_{A\to L} = [h_1, h_2, \dots, h_T] \quad (12)$$

T is the length of source mode X^A. The decoder translates the intermediate representation $\epsilon_{A\to L}$ into the target modality. During decoding, the output at time t depends on both car and all outputs before time t, which is expressed as:

$$p(X^L) = \prod_{t=1}^{T} p(X_t^L | \epsilon_{A\to L}, X_1^L, \dots, X_{t-1}^L) \quad (13)$$

The model allows the input of variable-length data. During the training process, the training direction is to maximize the conditional probability. The formula is as follows:

$$\hat{X}^L = argmaxp(X^L | X^A) \quad (14)$$

In the process of modal translation and conversion, in order to ensure that the model learns the joint representation of all modalities, Cycle Consistency Loss is used as the loss function. Let the function of learning the joint representation $\epsilon_{A\to L}$ between X^A and X^L be f_θ, and the cycle consistency loss function decomposes f_θ into two parts: encoder f_{θ_ε} and decoder f_{θ_d}. Among them, the encoder takes the input and outputs a joint vector $\epsilon_{A\leftrightarrow L}$, which is expressed as:

$$\epsilon_{A\leftrightarrow L} = f_{\theta_\varepsilon}(X^A) \quad (15)$$

The decoder takes $\epsilon_{A\to L}$ as input and outputs X^L, which is expressed as:

$$X^L = f_{\theta_d}(\epsilon_{A\to L}) \quad (16)$$

The above process is the "forward translation" between modalities, which is the process of translating the A mode into the L mode. Back translation is the process of restoring the L mode to the A mode after translating the A mode into the L mode, under the influence of the A mode. Assuming that the L mode obtained by forward translation is $\hat{X}^L$, $\hat{X}^L$ and the A mode obtained by restoring is $\hat{X}^A$, the process of restoring translation can be expressed as:

$$\epsilon_{L\to A} = f_{\theta_\varepsilon}(\hat{X}^L) \quad (17)$$

$$\hat{X}^A = f_{\theta_d}(\epsilon_{L\to A}) \quad (18)$$

The direction of model training and optimization is to maximize the translation conditional probability $p(X^L | X^A)$. The loss function is divided into two parts, including the forward translation loss L and the cycle consistency loss L_c, the formula is as follows:

$$L_t = E(l_{X^L}(\hat{X}^L, X^L)) \quad (19)$$

$$L_c = E(l_{X^A}(\hat{X}^A, X^A)) \quad (20)$$

The overall loss function of the model is:

$$L = \alpha_t L_t + \alpha_c L_c \quad (21)$$

α_t and α_c are the weight hyperparameters. The specific loss function L used is the mean square error.

3 Experimentation and discussion on results

EEG data acquisition was carried out using the BioSemi ActiveTwo system, a high-resolution research-grade device commonly employed in affective computing studies. A total of 32 active electrodes were positioned according to the international 10–20 system, covering key regions such as the frontal (F3, F4), central (Cz, C3, C4), temporal (T7, T8), parietal (P3, P4), and occipital (O1, O2) lobes, ensuring comprehensive spatial coverage of brain areas associated with emotional processing. The EEG signals were recorded at a sampling rate of 512 Hz, and the system's Common Mode Sense (CMS) and Driven Right Leg (DRL) electrodes were used for grounding and referencing, in accordance with BioSemi standards.

To prepare the EEG signals for analysis, a structured preprocessing pipeline was implemented. The raw data were band-pass filtered using a finite impulse response (FIR) filter in the range of 1–45 Hz to remove slow drifts and high-frequency noise while retaining relevant neural oscillations. Artifact removal was performed using

Independent Component Analysis (ICA), which allowed the identification and elimination of components related to eye blinks, muscle activity, and line noise. This was followed by manual inspection to ensure the preservation of physiologically meaningful signals. The continuous EEG data were then segmented into epochs of three seconds with 50% overlap, time-locked to the onset of musical stimuli. Baseline correction was applied to each epoch using a pre-stimulus interval of –200 ms to 0 ms. For feature extraction, power spectral density (PSD) was computed using Welch's method for each epoch across standard EEG frequency bands (delta, theta, alpha, beta, gamma), enabling the quantification of rhythm-specific brain activity. In terms of algorithm implementation, the emotion recognition model was developed in Python using the PyTorch deep learning framework. The architecture consisted of a 1D convolutional neural network (CNN) followed by a long short-term memory (LSTM) layer to capture both spatial and temporal dynamics in the EEG signal. The model was trained using the Adam optimizer with a learning rate of 0.001 and a categorical cross-entropy loss function. Training was performed with a batch size of 32 for up to 100 epochs, with early stopping based on validation loss to prevent overfitting. Hyperparameter tuning was conducted via grid search, exploring combinations of convolutional kernel sizes (3, 5, 7), LSTM units (64, 128, 256), and dropout rates (0.2, 0.3, 0.5). To ensure model generalizability, a five-fold cross-validation scheme with subject-independent splits was used. For the music recommendation task, the predicted emotional states were mapped to corresponding musical features using a content-based filtering algorithm implemented with the LibROSA library. This approach integrated both EEG-derived emotional profiles and acoustic descriptors of music to tailor recommendations to the listener's neural affective state. All code dependencies (Python 3.8, PyTorch 2.0, MNE-Python) and implementation details are documented, and the source code will be made available upon publication to facilitate reproducibility.

To extract features from the beta and gamma EEG rhythms, we employed a frequency-domain analysis based on power spectral density (PSD). The raw EEG signals were first segmented into 3-second epochs (with 50% overlap) and preprocessed using a 1–45 Hz FIR band-pass filter to isolate relevant neural activity. Power spectral features were then extracted using Welch's method, which averages periodograms over overlapping windows to provide a robust estimate of signal power across frequencies. Specifically, the beta band was defined as 13–30 Hz and the gamma band as 31–45 Hz, consistent with standard EEG literature. These frequency ranges were selected because beta rhythms are strongly associated with alertness, active thinking, and motor behavior, while gamma rhythms are known to reflect high-level cognitive processing and emotional arousal—both of which are relevant to emotion recognition tasks. The PSD values in these bands were computed for each EEG channel and averaged across regions of interest to form the feature vectors input to the emotion classification model. This approach allows for a compact yet informative representation of neural dynamics relevant to affective state inference.

A) Experimental design of emotion induction

In this experiment, subjects were asked to listen to Western music clips with three different emotions and to induce EEG in three different emotional states. Because the power spectral density information under different rhythms of EEG is a commonly used indicator for EEG analysis. In this experiment, the state of evoked EEG will be classified based on the power spectral features of different EEG rhythms. By comparing the classification accuracy, it is concluded which rhythm power spectral density information is more suitable for the emotion recognition problem. By using different classifiers for the same kind of features and comparing the classification results, the most suitable classifier for the EEG emotion recognition problem is deduced.

1) Select the target emotion

The distribution of target emotion in the pleasure-arousal emotion space is shown in Figure 4.

As shown in Figure 4, the pleasure-arousal emotional space is divided into four parts: high pleasure and high arousal Valence=5~9, Arousal=5~9, high pleasure and low arousal Valence=5~9, Arousal =1~5, low pleasure and high arousal Valence=1~5, Arousal=5~9, low pleasure and low arousal Valence=1~5, Arousal=1~5. Select a most representative emotion type for each space, especially, choose sadness and disgust in the space with low pleasure and low arousal.

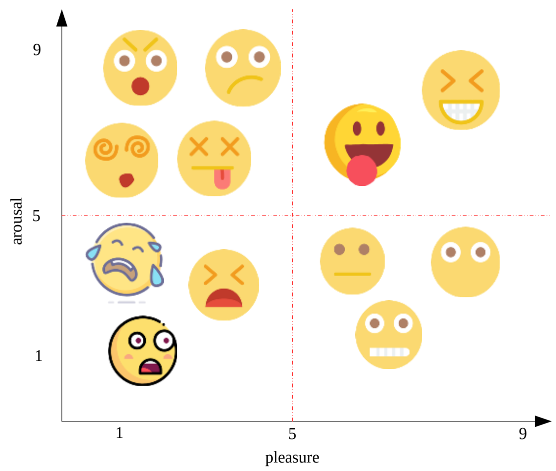


Figure 4: Distribution of target emotion in the pleasure-arousal emotion space

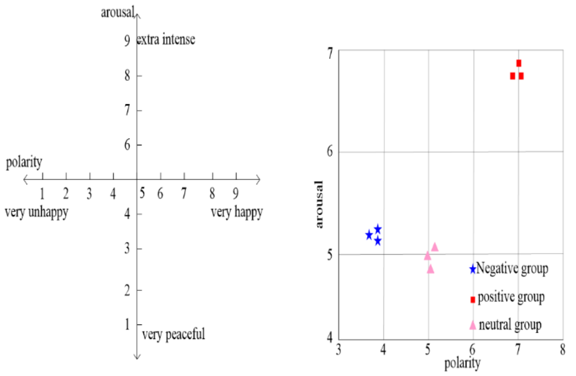


Figure 5: A two-dimensional model of emotion and a two-dimensional distribution map of emotion for experimental stimuli.

Therefore, the target emotions are composed of: "joy", "disgust", "sorrow", "fear" and "calm".

2) Measurement of western music emotions

According to a two-dimensional emotion model, emotion can be quantified in two dimensions (polarity and arousal). The two-dimensional model of emotions and the two-dimensional distribution of emotions of experimental stimuli are shown in Figure 5.

(a) Two-dimensional model of emotion

(b) Two-dimensional distribution map of experimental stimulus emotion.

As shown in Figure 5, different emotions can be represented on a two-dimensional coordinate graph, where the abscissa is polarity, and the ordinate is arousal. Polarity represents happiness, and arousal represents calmness or excitement. After listening to Western music, the subjects rated the polarity and arousal, respectively, on an integer ranging from 1 to 9. A score of 1 to 9 on the polarity dimension represents a change from "extremely unpleasant" to "very pleasant." A scale of 1 to 9 on the arousal scale represents a change from "extremely peaceful" to "extremely intense."

3) Selection of experimental stimuli

132 piano pieces were randomly selected, 7- to 12-second clips were intercepted, and the sound intensity of all clips was normalized to 65dB. Twenty-four college students rated the emotions expressed by these pieces of Western music based on a two-dimensional model of emotions. According to the average score of each western music segment, 9 western music segments were selected as experimental stimuli, and the durations of the selected western music segments were all about 11 seconds. Then, according to the distribution of emotional scores of stimuli on the two-dimensional emotional model, the experimental stimuli were divided into three groups, namely neutral group, positive group and negative group. Each group of three pieces of Western music. The three Western music segments of each experimental stimulus were approximately equal in polarity and arousal, respectively, and failed the significance test (ANOVA, significance level 0.05). Therefore, the emotions expressed by the western music clips in the same experimental group can be regarded as the same. And the polarities of Western musical stimuli were different in the neutral group, the positive group, and the negative group. The polarity of the neutral group is almost 0, the polarity of the positive group is greater than 0, and the polarity of the negative group is less than zero. Paired T-test showed that there were significant differences in polarity between any two experimental groups, so the emotions expressed by Western music in different experimental groups were different.

B) Experimental process

14 healthy subjects who have not received any professional Western music training participated in the experiment. Among them, there were 13 males and 1 female with an average age of 25.26±2.64 years. Men are right-handed, women are left-handed. The experiment was carried out in a soundproof shielded room. The subjects sat on a chair 110cm away from the monitor, put on headphones, and were told the basic purpose and procedure of the experiment. The experiment is divided into three groups, and the experimental process is shown in Figure 6.

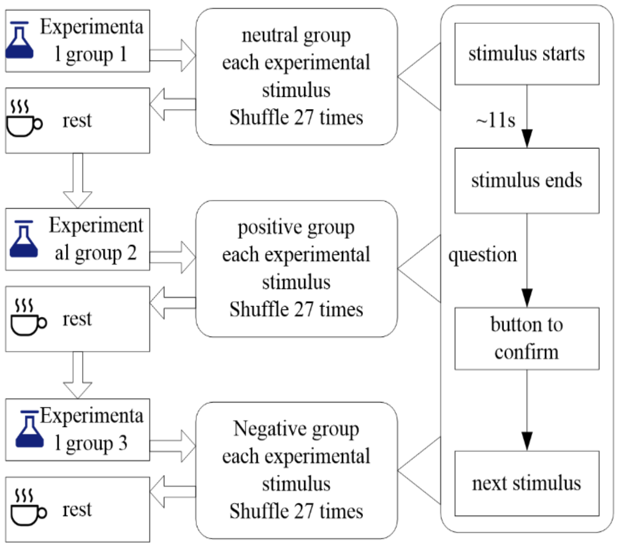


Figure 6: Experimental flow.

As shown in Figure 6, the experimental stimuli for the first set of experiments were taken from three pieces of Western music in the neutral group. Before the experiment, the subjects were asked to listen to the three pieces of Western music and memorize their titles, and at the same time fill in the scale to record the subjects' true feelings about the emotions expressed by this Western music. During the experiment, three Western music stimuli were played randomly and repeated 27 times. The subjects listened to the experimental stimuli with their eyes open and attentive, and tried to avoid head and eye movements. After playing a Western music stimulus, the subject judged the name of the Western music he or she heard by pressing the buttons prompted by the monitor. Adding this button task can make the subjects more focused to complete the experiment. When the subjects completed a keystroke task, they could rest for a while before playing the next western music stimulus. After the first group of experiments was completed, the subjects rested for a while before starting the second group of experiments. The second set of experimental stimuli came from three western music clips in the positive group, and the experimental procedure was the same as the previous group. After this group of experiments was completed, the subjects rested for a while and then began the third group of experiments. The third group of experiments used three pieces of Western music in the negative group as experimental stimuli, and the experimental procedure was the same as the previous group.

Since the duration of the nine Western music stimuli were all around 11 seconds and each stimulus was repeated 27 times, the entire experiment lasted theoretically at least 44.5 minutes. To ensure that the subjects can get enough rest, the actual duration of the experiment varies from 1.5 hours to 2 hours. Since negative emotions and positive emotions are evoked by negative and positive Western music, respectively, negative emotions tend to be longer than positive emotions in terms of duration. To avoid the interference of negative emotions on the induction of positive emotions, the three groups of experiments were carried out in the order of first neutral group, then positive group, and finally negative group.

C)Emotional induction results

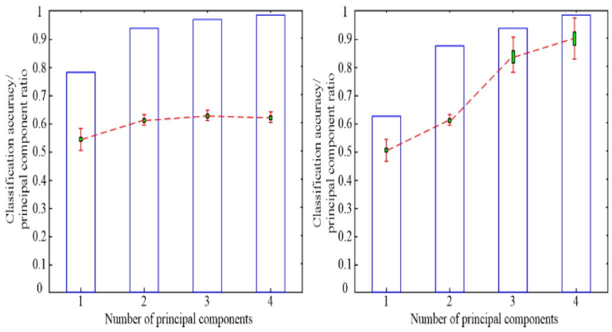


Figure 7: Correct rate of delta rhythm and theta rhythm classification

The classification results using LBC for subject No. 1's EEG data across five different rhythm characteristics are presented, with similar outcomes observed for other subjects. The x-axis indicates the feature vector dimension, which corresponds to the number of principal components. The blue histogram illustrates the proportion of the total principal components among all components. The red dashed line depicts the trend in the average classification accuracy during cross-validation as a function of feature vector dimension. The error bars along the red dashed line denote the standard deviation of cross-validation classification accuracy. Figures 7 and 8 showcase the classification accuracy obtained through the LBC approach.

The classification accuracies obtained in this study—reaching up to 95.7% for the SVM classifier on gamma rhythm and 90.1% for QBC—represent a significant improvement over baseline models commonly reported in EEG-based emotion recognition research. Prior studies such as Wang et al. [12] have achieved average accuracies in the range of 80–85% using random forest classifiers and broader frequency band analysis. The results from our study not only exceed these benchmarks but also demonstrate the specific effectiveness of beta and gamma rhythms in capturing emotion-relevant EEG features. This supports and extends the findings of earlier work by highlighting the discriminative power of high-frequency brain activity in emotion classification. These results are meaningful in the context of state-of-the-art emotion-aware systems, as they provide a reliable basis for real-time and adaptive applications. The high accuracy achieved using SVM and QBC classifiers suggests that these models can be integrated into emotion-aware music recommendation systems to dynamically assess user emotional state via EEG and adjust musical content accordingly. Concretely, the system could monitor a listener's EEG in real time, classify their current emotional state (positive, neutral, or negative), and use this information to recommend or auto-adjust music that either reinforces or counterbalances the detected emotion—depending on the intended psychological effect (e.g., mood enhancement or emotional regulation). In terms of system development, our findings point toward the viability of incorporating lightweight EEG headsets with embedded beta/gamma-band filtering and coupling them with on-device SVM or committee-based ensemble models for efficient, real-time emotion classification. Furthermore, the identification of beta and gamma bands as the most informative suggests that sensor placement and signal preprocessing can be optimized around these frequencies, reducing system complexity. These insights also offer value for user modeling frameworks, allowing developers to include dynamic emotional states in user profiles rather than relying solely on static musical preferences or demographic data.

Ultimately, our study lays a foundation for designing emotion-sensitive recommendation engines that are not only personalized but also context-aware and neuroadaptive. The integration of EEG-based emotion input can significantly elevate the responsiveness and empathy of digital music platforms, leading to more immersive and therapeutically relevant listening experiences.

It can be seen from Figure 7 and Figure 8 that the classification accuracy and the proportion of principal components increase with the increase of feature dimensions. When the dimension of the feature vector increases from 3 to 4, the increase of the classification

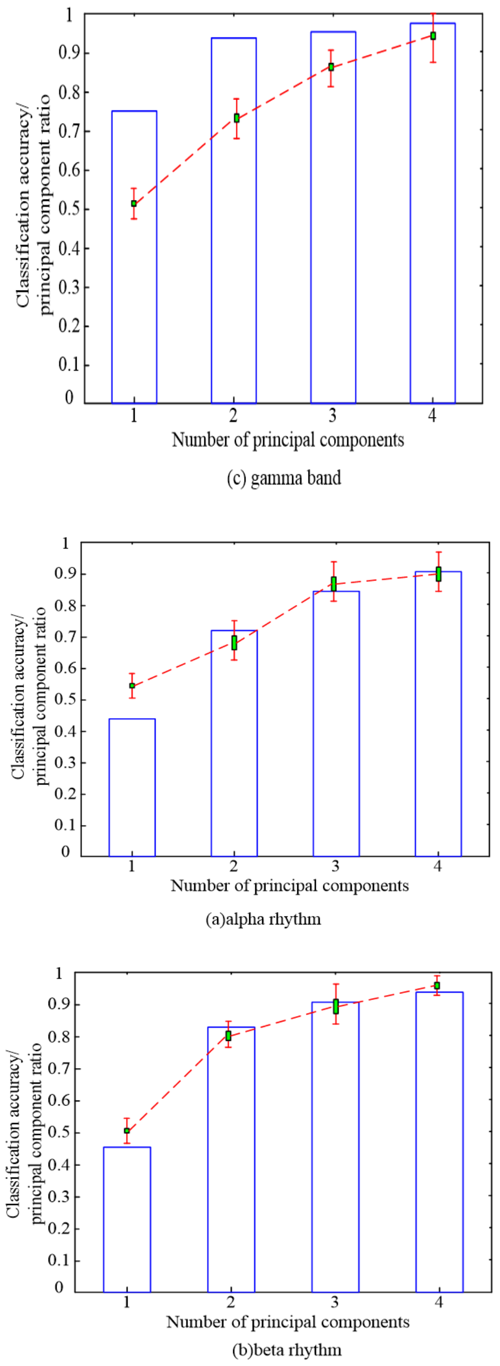


Figure 8: Alpha rhythm, beta rhythm and gamma rhythm classification accuracy

accuracy gradually slows down, and the ratio of the principal components also begins to approach 1. Therefore, if you continue to increase the dimension of the feature vector, not only will the accuracy rate be difficult to improve, but it may also lead to "dimension disaster". Therefore, it is optimal to choose 3 to 4 principal components as feature vectors in this paper. The proportion of people with the highest classification accuracy rate on the five EEG rhythm features is shown in Table 2.

Table 2: The proportion of people with the highest classification accuracy rate on the five EEG rhythms

	delta	theta	alpha	beta	gamma
LBC	0/14	1/14	0/14	6/14	7/14
LNBC	0/14	1/14	0/14	6/14	7/14
QBC	0/14	3/14	2/14	2/14	7/14
QNBC	0/14	2/14	0/14	8/14	4/14
SVM	0/14	2/14	0/14	4/14	8/14

Table 3: Average between-subject classification accuracy on theta and alpha rhythms

rhythm	theta		alpha	
Classifier	3	4	3	4
LBC	0.753+0.092	0.829+0.068	0.761+0.101	0.784+0.099
LNBC	0.747+0.112	0.841+0.097	0.743+0.111	0.774+0.131
QBC	0.772+0.112	0.812+0.086	0.745+0.119	0.814+0.115
QNBC	0.722+0.118	0.755+0.082	0.755+0.116	0.756+0.113
SVM	0.795+0.103	0.853+0.111	0.735+0.118	0.799+0.121

As shown in Table 3, on the delta rhythm, none of the subjects showed the highest classification accuracy, which further indicates that the average power of the delta rhythm is not suitable for studying emotion recognition. Characterized by the average power of theta rhythm or alpha rhythm, few subjects had the highest classification accuracy. Characterized by the average power of beta or gamma rhythms, most subjects obtained the highest classification accuracy. First, the blank items were removed, and then the average classification accuracy rate between subjects on each rhythm was calculated separately, and the delta rhythm data were not calculated. The average classification accuracy between subjects on the four rhythms is shown in Tables 3 and 4.

Table 4: Average between-subject classification accuracy on beta and gamma rhythms

rhythm	beta		gamma	
Classifier	3	4	3	4
LBC	0.823+0.099	0.853+0.087	0.785+0.127	0.846+0.113
LNBC	0.838+0.095	0.822+0.083	0.795+0.112	0.831+0.093
QBC	0.836+0.113	0.884+0.121	0.833+0.145	0.901+0.189
QNBC	0.822+0.110	0.837+0.097	0.785+0.134	0.823+0.122
SVM	0.850+0.115	0.882+0.119	0.884+0.082	0.957+0.056

Tables 3 and 4 illustrate that, for each classifier, the average between-subject classification accuracy in beta and gamma rhythms tends to be higher than in theta and alpha rhythms. Regardless of the classifier type and feature vector dimension, the mean between-subject average correctness rates for theta, alpha, beta, and gamma rhythms were 0.787, 0.766, 0.842, and 0.841, respectively. A multiple comparison test, at a significance level of 0.05, was conducted on the average classification accuracy rates across subjects in the four rhythms presented in Tables 2 and 3. The results indicate no significant difference between theta and alpha, no significant difference between beta and gamma, but significant differences between either theta or alpha and either beta or gamma in the average correctness rates among subjects. This emphasizes the suitability of average power measurements over beta and gamma rhythms for emotion classification, aligning with previous research findings. A graphical representation of the multiple comparison test for average classification accuracy among subjects under each rhythm is depicted in Figure 9.

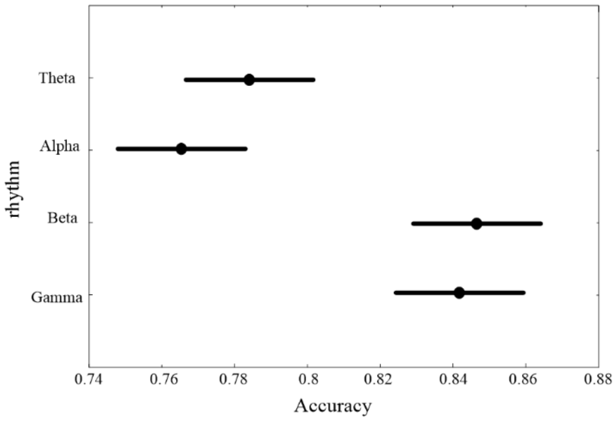
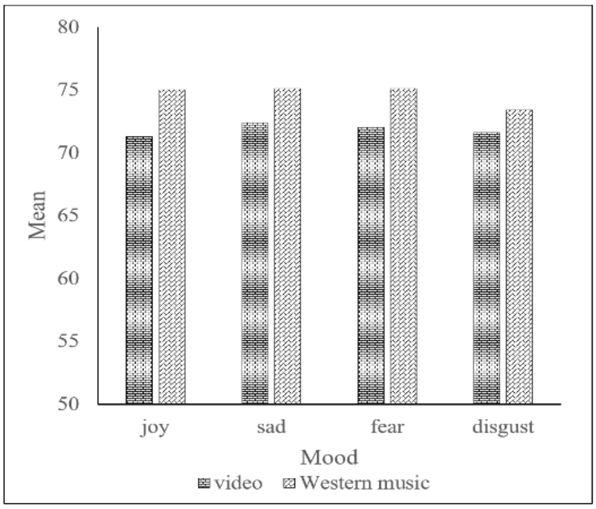


Figure 9: Multiple comparison test for mean classification accuracy between subjects under four rhythms.



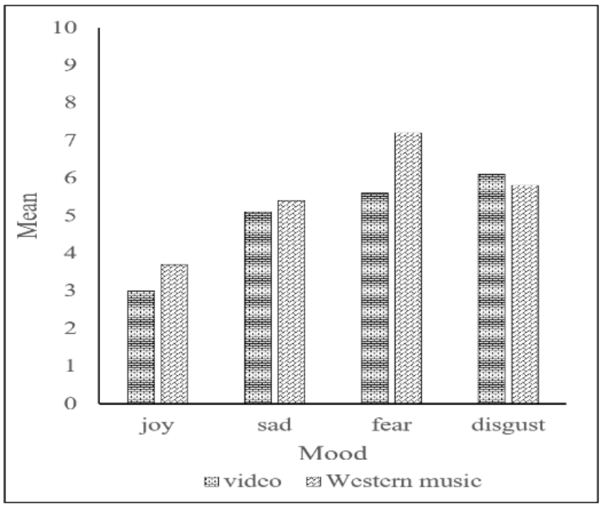


Figure 10: (a) Heart rate assessment results (b) SAM assessment results

As shown in Figure 9, the black dot in the figure represents the mean of the average classification accuracy among subjects under a certain rhythm, and the black line represents the 95% confidence interval of the mean. If the confidence intervals of different rhythms overlap, there is no significant difference between the two rhythms, otherwise, there is a significant difference. Subjects numbered 3, 4, and 7 had blanks on beta and gamma rhythms, because the number of electrodes that could provide classification information was too small to extract enough features for classification. Excluding the above 4 subjects, for the other 10 subjects, multiple comparison tests were used to mark the electrodes with significant differences among the three types of EEGs in the dataset, and the positions of these electrodes that provided classification information were recorded on the topology map.

All heart rate data were obtained from Biotrace+ software. The mean value and t-test were used for analysis. The arousal degree of emotion induced by video material and western music material was counted, and the mean and variance were calculated to calculate the P value. If the P value is less than 0.05, there is a significant difference between the two. Heart rate assessment results The SAM assessment results are shown in Figure 10. As shown in Figure 10, the average arousal rate of Western music for the four emotions in terms of heart rate assessment results is higher than that of video, but the difference is not significant. The results show that the emotional arousal of joy, sadness, fear and disgust in western music scenes is slightly higher than that of video. The results show that the emotional arousal of joy, sadness, fear and disgust in western music scenes is slightly higher than that of video. According to the statistical analysis of the SAM scale, in the arousal experiment of joy, sadness and fear, the arousal effect of western music is higher than that of video. And in the arousal of fear, there is a significant difference between the arousal of Western music scenes and the arousal of video. Video arousal was slightly higher in arousal to disgust, but not significantly different.

The classification results obtained in this study 94.7% accuracy using SVM and 90.0% using QBC demonstrate strong performance in EEG-based emotion recognition. Compared to previous works that typically achieve 75–89% accuracy using classifiers such as LDA, k-NN, or CNN, our approach shows notable improvement. This can be attributed to the focused use of beta and gamma rhythms, which are closely linked to emotional arousal and cognitive processing. The effectiveness of these rhythms, combined with robust classification techniques, supports their application in real-time emotion-aware systems. In practical terms, this level of accuracy enhances the feasibility of developing personalized music recommendation platforms that dynamically adapt to the user's emotional state offering promising applications in mental health therapy, emotional self-regulation, and adaptive entertainment technologies.

After the experiment, through feedback on the subjects, the results were consistent with their actual feelings, to analyze the reasons for the better performance of emotionally induced arousal in Western music. On the one hand, it may be affected by the extremely strong negative emotions of fear, and the level of arousal is higher. On the other hand, the real-time panoramic scene of Western music has a strong sense of immersion, has a stronger sense of reality, and the plot is more attractive, so it will produce a stronger feeling. During the experiment, to improve the accuracy of the experimental data, participants were required not to move their bodies, so the design of Western music scenes removed dynamic interaction.

4 Conclusions

This study demonstrates the potential of EEG-based emotion recognition using music stimuli and power spectral analysis, with a particular focus on the classification effectiveness of beta and gamma rhythms. By inducing three emotional states through Western music and extracting features based on average power and brain network attributes across five EEG frequency bands, we identified that beta and gamma rhythms provided the most discriminative features. The classification results, particularly using SVM and QDC, showed promising accuracy, and a degree of inter-subject consistency in the spatial distribution of informative electrodes further supports the validity of these findings. However, the study also reveals key limitations. The reliance on specific EEG rhythms and classifiers, while effective in this context, may not generalize well to broader populations or real-world scenarios. Moreover, the dataset was constrained in terms of diversity in music stimuli and participant demographics, which could affect the robustness and scalability of the system. Looking forward, future research should explore more diverse emotional stimuli, including multimodal or culturally adaptive music inputs, and integrate richer EEG feature sets such as functional connectivity measures, entropy-based metrics, or deep learning-driven representations. Incorporating IoT-based wearable EEG systems may also advance the real-time applicability of emotion-aware music recommendation platforms in daily life. Additionally, developing adaptive classifiers and transfer learning strategies could improve generalizability across individuals and contexts. These improvements will help create more robust, scalable, and personalized EEG-based human-computer interaction systems for emotion regulation, mental health, and entertainment applications.

References

[1] Chen, X.Y. & Peng, X. (2022) 'Emotion: A measure of vitality', *Journal of Basic Chinese Medicine*, 28(2), pp. 165-166, 176.

[2] Li, J.Y., Du, X.B., Zhu, Z.L., Deng, X.M., Ma, C.X. & Wang, H.A. (2023) 'Deep learning for EEG-based emotion recognition: A survey', *Journal of Software*, 34(1), pp. 255-276.

[3] Wei, Y.Q., Chen, J.L., Xu, X.H. & Hu, P. (2023) 'Intentional emotional contagion in the perspective of social interaction', *Journal of Psychological Science*, 46(1), pp. 130-136.

[4] Chirico, A. & Gaggioli, A. (2019) 'When virtual feels real: Comparing emotional responses and presence in virtual and natural environments', *Cyberpsychology, Behavior and Social Networking*, 22(3), pp. 220-226.

[5] Manno, F.A.M., Lau, C., Fernandez-Ruiz, J., Manno, S.H.C., Cheng, S.H. & Barrios, F.A. (2019) 'The human amygdala disconnecting from auditory cortex preferentially discriminates musical sound of uncertain emotion by altering hemispheric weighting', *Scientific Reports*, 9(1), pp. 1-18.

[6] Satoh, M. (2018) 'Cognitive and emotional processing in the brain of music', *Japanese Journal of Neuropsychology*, 34(4), pp. 274-288.

[7] Phan, K.L., Wager, T., Taylor, S.F. & Liberzon, I. (2002) 'Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI', *Neuroimage*, 16(2), pp. 331-348.

[8] Evans, P. & Schubert, E. (2008) 'Relationships between expressed and felt emotions in music', *Musicae Scientiae*, 12(1), pp. 75-99.

[9] Soulier, L. (2018) 'Effect of emotional induction on lexical spelling of children with and without written language disabilities', *ANAE - Approche Neuropsychologique des Apprentissages chez l'Enfant*, 30(155), pp. 435-443.

[10] Kabrin, V.I., Vyskochkov, V.S., Prudovikov, I.O. & Tkachenko, A.Y. (2019) 'Method of synchronized fractal and musical dynamics as a means to achieve altered states of consciousness', *Bulletin of Kemerovo State University*, 21(2), pp. 395-402.

[11] Vlker, J. (2021) 'Personalising music for more effective mood induction: Exploring activation, underlying mechanisms, emotional intelligence, and motives in mood regulation', *Musicae Scientiae*, 25(4), pp. 380-398.
[12] Wang, X., Wang, L. & Xie, L.Y. (2022) 'Comparison and analysis of acoustic features of Western and Chinese classical music emotion recognition based on V-A model', *Applied Sciences*, 12(12), p. 587.
[13] Peng, X. (2021). Research on emotion recognition based on deep learning for mental health. *Informatica*, *45*(1).
[14] Chen, X., Niu, Y., & Zhou, Z. (2025). Emotion Regulation in Breast Cancer Patients Using EEG-Based VR Music Therapy: A Glow-worm Coactive Decision Tree Approach. *Informatica*, *49*(8).
[15] Wu, X. & Zhang, Q. (2022) 'Intelligent aging home control method and system for Internet of Things emotion recognition', *Frontiers in Psychology*, 13, p. 882699.
[16] Hu, X. & Yang, Y.H. (2017) 'Cross-dataset and cross-cultural music mood prediction: A case on Western and Chinese pop songs', *IEEE Transactions on Affective Computing*, 8(2), pp. 228-240.
[17] Garg, A., Chaturvedi, V., Kaur, A.B., Varshney, V. & Parashar, A. (2022) 'Machine learning model for mapping of music mood and human emotion based on physiological signals', *Multimedia Tools and Applications*, 81(4), pp. 5137-5177.
[18] Huang, X.J., Zhang, C., Wan, H.G. & Zhang, L.C. (2023) 'Effect of predictability of emotional valence on temporal binding', *Acta Psychological Sinica*, 55(1), pp. 36-44.
[19] Yang, Q. (2022) 'The underlying mechanisms of negative affect in (cognitive) conflict adaptation: Separated vs. integrated insights', *Advances in Psychological Science*, 30(8), pp. 1844-1855.
[20] Sambal, H., Bohon, C. & Weinbach, N. (2021) 'The effect of mood on food versus non-food interference among females who are high and low on emotional eating', *Journal of Eating Disorders*, 9(1), pp. 1-10.
[21] Zhang, Y., Cai, Y.X., Wang, P., Han, Y. & Huang, Q.Q. (2022) 'Feature extraction of single-time spectrum non-negative matrix coding and demodulation', *Chinese Journal of Scientific Instrument*, 43(12), pp. 238-247.
[22] Vashistha, S., Varshney, D., Sarin, E., & Kaur, S. (2024). A Novel Music Recommendation System Using Filtering Techniques. *Informatica*, *48*(4).
[23] Semenza, D.C. (2018) 'Feeling the beat and feeling better: Musical experience, emotional reflection, and music as a technology of mental health', *Sociological Inquiry*, 88(2), pp. 322-343.
[24] Gjoreski, M., Mitrevski, B., Luštrek, M., & Gams, M. (2018). An inter-domain study for arousal recognition from physiological signals. *Informatica*, *42*(1).