

Research Article

Mind to Music: An EEG Signal-Driven Real-Time Emotional Music Generation System

Shuang Ran ¹, Wei Zhong ², Lin Ma ¹, Danting Duan ¹, Long Ye ², and Qin Zhang ²

¹Key Laboratory of Media Audio & Video, Communication University of China, Beijing 100024, China

²State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China

Correspondence should be addressed to Wei Zhong; wzhong@cuc.edu.cn

Received 11 January 2024; Revised 8 October 2024; Accepted 12 December 2024

Academic Editor: Alexander Hošovský

Copyright © 2024 Shuang Ran et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Music is an important way for emotion expression, and traditional manual composition requires a solid knowledge of music theory. It is needed to find a simple but accurate method to express personal emotions in music creation. In this paper, we propose and implement an EEG signal-driven real-time emotional music generation system for generating exclusive emotional music. To achieve real-time emotion recognition, the proposed system can obtain the model suitable for a newcomer quickly through short-time calibration. And then, both the recognized emotion state and music structure features are fed into the network as the conditional inputs to generate exclusive music which is consistent with the user's real emotional expression. In the real-time emotion recognition module, we propose an optimized style transfer mapping algorithm based on simplified parameter optimization and introduce the strategy of instance selection into the proposed method. The module can obtain and calibrate a suitable model for a new user in short-time, which achieves the purpose of real-time emotion recognition. The accuracies have been improved to 86.78% and 77.68%, and the computing time is just to 7 s and 10 s on the public SEED and self-collected datasets, respectively. In the music generation module, we propose an emotional music generation network based on structure features and embed it into our system, which breaks the limitation of the existing systems by calling third-party software and realizes the controllability of the consistency of generated music with the actual one in emotional expression. The experimental results show that the proposed system can generate fluent, complete, and exclusive music consistent with the user's real-time emotion recognition results.

Keywords: brain-computer interface; EEG signals; emotional music generation; real-time emotion recognition; transfer learning

1. Introduction

Music is an important way for people to express their emotions [1]. The traditional way of manual composition not only requires a solid knowledge of music theory but also is a time-consuming and labour-intensive process. Miranda [2] first introduced the concept of brain-computer music interface (BCMI), which aims to allow users to interact with music or control certain properties of music via EEG signals. As we know, brain-computer interface (BCI) can connect the brain to external devices directly enabling information exchange between them while bypassing peripheral nerves and muscles [3–5]. By capturing and converting brain

signals into electrical signals, BCI achieves the transmission and control of information and involves the research fields of medicine, neurology, signal processing, etc. [6–8]. As a branch of BCI, the BCMI has played an increasingly important role in the applications such as emotion regulation [9, 10], treating mental illnesses [11–13], and music generation [14–17]. Although many research works have been proposed to generate music from EEG signals, there are few feasible systems which can be applied in practice. Ehrlich, Guan, and Cheng [9] proposed a conceptual system for studying the interaction between human emotions and emotional stimuli, in which a rule for generating music from EEG signals was embedded in a real-time BCMI

architecture, allowing the EEG signals to control the generation of music for emotion regulation in a closed loop. Later, they also implemented the conceptual system by controlling the parameters to generate specific emotional music for the subjects [10]. Concretely, the predicted arousal and valence values of the subjects' emotion were mapped to the parameters needed for music generation, calling the software to generate a new emotional music to regulate the emotion of subjects, which forms an interactive closed loop. Kimmatkar and Babu [13] proposed an emotional feedback regulation method based on the BCMI therapy for the symptoms of depression. Wu, Li, and Yao [15] mapped the average power and amplitude of EEG signals to the intensity and pitch of music, respectively, allowing the EEG signals to be converted directly to music. Makeig et al. [16] used the paradigm of imaginary stimuli to obtain the signals of EEG together with scalp muscle and eye activity for emotion recognition, and then synthesized music with the Max/MSP software based on the recognition results. Bellier et al. [17] applied a stimulus reconstruction approach previously used in the speech domain and reconstructed a recognizable song from direct neural recordings.

From the above analysis, the existing EEG emotional music generation systems have the following limitations:

- For the real-time EEG emotion recognition, most of the methods use trained models to test the new subject. However, the statistics of the EEG signals is person-dependent, and if the samples for training and testing are collected from the different persons, the performance of model will decrease dramatically.
- For the music generation, all of the existing systems call the third-party software and lost the controllability of the consistency of generated music with the actual one in emotional expression. And thus, it is needed to implement a controllable music generation scheme according to the recognized emotion of the subject's EEG signals.

To address the above issues, this paper proposes and implements an EEG signal-driven real-time emotional music generation system, which can quickly and accurately extract the EEG emotional features of the subject and generate personal exclusive music. For the EEG real-time emotion recognition module, we introduce instance selection and propose a simplified style transfer mapping method, which achieves an accuracy of 86.78% in 7 s on the SEED dataset and an accuracy of 77.68% in 10 s on self-collected dataset, exceeding the baseline and comparison methods in accuracy and computing time. For the emotional music generation module, we use the recognized emotion of EEG signals and structure features of music as the conditional inputs to the generative network, while adding the perceptual optimization of the emotional classification network branch to generate music which are closer to the real emotional expression. The experimental results show that compared with the existing representative methods, the proposed emotional music generation module has improved emotional expression authenticity, music content richness, and quality indicators by 5.14%, 15.53%, and 17.79%,

respectively. From the results of the on-line experiments, the proposed system can generate smooth, complete, and exclusive music that is consistent with the real-time emotional recognition results of the subject's EEG signals.

Three main contributions in this work are listed:

- We propose a simplified style transfer mapping method based on instance selection (SSTM-IS) by obtaining the informative instances and optimizing the way for updating parameters in style transfer mapping to realize fast and accurate subject-independent EEG emotion recognition.
- Considering the structural characteristics of the music and consistency of generated music with the actual one in emotional expression, a music generative network is designed to control the emotion of generated music in accordance with the recognized emotion state.
- By embedding the modules of EEG real-time emotion recognition and emotional music generation, we propose an EEG signal-driven real-time emotional music generation system. The experimental results demonstrate that the system can be calibrated in 7 s with good performance for a newcomer, and the recognized emotion states are used for controlling the later music generation.

The structure of this paper is as follows: Section 2 briefly reviews and analyzes the recent research developments of the EEG emotion recognition and music generation. Section 3 presents the framework of the proposed system and describes the modules of real-time EEG emotion recognition and emotional music generation in detail. Section 4 introduces the concrete implementation process of the proposed system. The experiments are designed in Section 5 to demonstrate the effectiveness and practicality of the proposed system in two aspects of the EEG emotion recognition and emotional music generation. Finally, some conclusions and future works are drawn in Section 6.

2. Related Works

This section briefly reviews the relevant works in the fields of EEG emotion recognition and emotional music generation.

2.1. EEG Emotion Recognition. As known that EEG signals are nonstationary and there are significant differences among different individuals. The model trained by the saved data is usually not suitable to the new user. The EEG emotion detection application requires the model to recognize emotion states accurately and fast when a new subject comes. One type of method is to use all unlabelled data of the target domain in the training phase. By approaching the distribution distance between the source and target domains, the generalization of the model can be improved [18–21]. However, this method uses all of the unlabelled target data in the model training; thus, it is obviously inapplicable to recognize the emotion of a new subject in real

time. In contrast, another method mainly extracts the invariant features from multisource domains to obtain a model with good generalization ability [22, 23]. In this method, the target domain data does not need to appear in the training of the model. The performance of the recognition model trained without target data is generally not as good as that of enhancing the model generalization by using unlabelled target data for domain adaptation. The PPDA model, which used all target and source domain data for domain adaptation in model training, achieves higher performance than that of the same model which does not use target data [23]. It can be inferred from the analysis above that, combining the two approaches is also an idea to train a model with reliable recognition performance. By combining the marginal and conditional distributions, Chai et al. [24] proposed an adaptive subspace feature matching (ASFM) method and achieved a recognition accuracy up to 83.51% on the SEED dataset [25, 26]. Li et al. [27] proposed a multisource style transfer mapping (MS-STM) method by using a few labelled target data to calibrate the model and approach the distribution distance between target and multiple source domain data. Chen et al. [28] proposed a fast online instance transfer (FOIT) method for classifying emotion states in real-time, and they conducted the subject-independent experiments on the SEED dataset, which reached the accuracy of 82.05% with the computing time of 32 s. Besides, some methods focus on generating the emotion-related features from EEG signals to perform the EEG emotion recognition task [29–31]. Dogan et al. [31] proposed an automated PrimePatNet87 model by using q-factor wavelet transform to obtain the features with minimum redundancy and maximum relevance. To sum up, most of the current works are still difficult to find a balance between the computing time and performance for the applications. It is necessary to design a simple and effective transfer learning method to reduce negative transfer and obtain good performance while shortening computing time.

2.2. Emotional Music Generation. The research on emotional music, such as emotional music generation [32–36] and music emotion analysis [37], which can be applied to machine composition and other related applications, has received a lot of attentions. In the emotional music generation study, most of the works focus on the mapping relationship between music parameters and emotional labels to construct a rule of emotional music generation based on perceptual information. Ferreira and Whitehead [33] constructed a MIDI format emotional music dataset named as VGMIDI based on Russell's emotional model, which includes both positive and negative labels. By employing the emotions as the connection between the visual and auditory, Tan, Antony, and Kong [34] used pretrained image representations and explored music modelling based on RNN and transformer architectures to build models being capable of generating music for a given image input. Based on the emotional text generation method, Radford, Jozefowicz, and Sutskever [35] also proposed a multilayer long short-term memory (LSTM) emotional music generation network,

which employs the logistic regression to construct a music emotion analysis model and optimizes the parameters by the genetic algorithm used in [33]. Hung et al. [36] embedded emotion labels into the transformer model and built a four-category emotional music generative model based on Russell's emotional model. At present, the task of emotional music generation is still in the initial stage of exploration. For music with different emotion categories, their emotional expression is often highly related to the structural characteristics. It is worth mentioning that most current works either only consider the internal structure of music or require the generation of music within specific emotions, ignoring their relationships between them, and thus, it is important to take into account not only the structural characteristics of the music itself, but also the consistency of the generated music with the real music in terms of emotional expression.

The EEG signal-driven real-time emotional music generation system needs to classify the emotion state of EEG signals from newcomer accurately and quickly, and then generates high-quality emotional music in accordance with the recognized results. From the above analysis, the current cross-subject EEG emotion recognition methods do not take into account the computing time of recognition model for the newcomer. Besides, it is also important to consider the consistency of generated music with the actual one in emotional expression. Therefore, it is necessary to propose an EEG signal-driven real-time emotional music generation system, which can generate personal exclusive music based on user's real-time emotional state in actual scenarios.

3. The Proposed System

This section provides a detailed introduction to the specific algorithm implementation of the two modules of EEG emotion recognition and music generation in the proposed system.

3.1. The System Structure. In this section, we propose and implement an EEG signal-driven real-time emotional music generation system, which can quickly and accurately extract the EEG emotional features of the subject and generate personal exclusive music. As shown in Figure 1, the proposed system consists of four main components: EEG signal acquisition, EEG signal recording and processing, real-time emotion recognition, and emotional music generation. The system first presents the video stimulus with a length of about 1–3 min to the subjects and records their EEG signals, and then the features are filtered and extracted from the collected EEG signals. After that a real-time EEG emotion recognition module is used for obtaining the emotion states, and finally an emotional music generation module is employed to generate music with the emotion recognition results. For the problem about real-time EEG emotion recognition, we propose a SSTM-IS method by obtaining the informative instances and optimizing the way for updating parameters in style transfer mapping to realize fast and accurate subject-independent EEG emotion recognition,

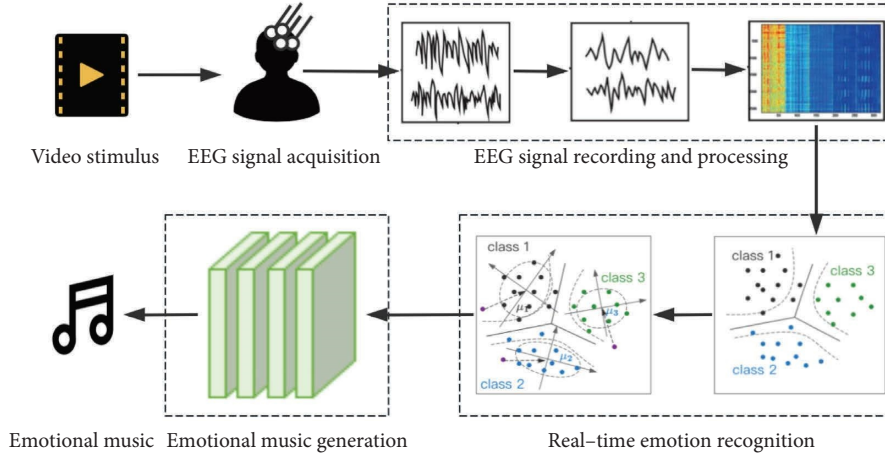


FIGURE 1: The structure of the EEG signal-driven real-time emotional music generation system.

satisfying the needs of the EEG signal-driven real-time emotional music generation system. In terms of music generation, we propose an emotional music generation network based on structure features and embed it into our system, which breaks the limitation of the existing systems by calling the third-party software and realizes the controllability of the consistency of generated music with the actual one in emotional expression.

3.2. Real-Time EEG Emotion Recognition Module. Since EEG signals vary greatly between individuals, an EEG emotion recognition model trained by the EEG data of an individual cannot be applied to the other one. Therefore, when a new individual comes, a large amount of labelled data needs to be collected to retrain a personalized model. In the proposed system, we want to use a small amount of labelled data from a newcomer to quickly calibrate a model suitable for the comer to achieve accurate and fast emotion recognition. Inspired by the idea of using learner to select unlabelled samples in active learning [38] and transfer learning of STM [27, 39], we propose the SSTM-IS method [18] for real-time EEG emotion recognition module. Its framework is shown in Figure 2, which consists of two main components: instance selection and data mapping.

The source domain is $S = \{s_i, i = 1, 2, \dots, n_s\}$, and its label is $L = \{y_s^i, i = 1, 2, \dots, n_s\}$, where n_s is the number of source data. The target domain is $T = \{t_i, i = 1, 2, \dots, n_t\}$ which consists of two parts of the labelled and unlabelled, and L_T is the label for the former, where n_t represent the number of target data. The marginal distributions of S and T are different. And then, we extract the differential entropy (DE) feature [26, 40] for source and target data. The DE feature of source data is noted as $S_L = \{s_L^i \in R^m | i = 1, 2, \dots, n_s'\}$, where n_s' is the number of DE feature and m is the size of feature. Same as the target domain, the DE feature of the labelled and unlabelled data is $T_L = \{t_L^i \in R^m | i = 1, 2, \dots, n_{t-L}'\}$ and $T_U = \{t_U^i \in R^m | i = 1, 2, \dots, n_{t-U}'\}$, where n_{t-L}' and n_{t-U}' represent the feature size of the labelled and unlabelled target data. $U' = \{U_i' \in R^m | i = 1, 2, \dots, n_{t-U}'\}$ represents the features mapped from T_U .

The instance selection is presented in the left part of Figure 2. The classifier C_0 is trained by T_L and T_L is extracted from the target labelled domain data with the label Y_L . In our method, C_0 is trained by the data which contains different emotion states from the labelled part of target domain. We regard the probability w_i predicted by classifier C_0 as the information in each sample as follows:

$$w_i = C_0(s_L^i), \quad i = 1, 2, \dots, n_s'. \quad (1)$$

Then, the instances will be selected by w_i . To achieve a balanced distribution of data for each emotion category, the source instances are sorted by the amount of information in each category and noted as S_{L-sort} . The top k informative instances are selected as S_{sel} .

$$\begin{aligned} S_{L-sort} &= g(s_L^i) = \{s_L^i | i = 1, 2, \dots, n_c\} \\ &= S_L^1 \cup S_L^2 \cup \dots \cup S_L^{n_c}, \end{aligned} \quad (2)$$

$$S_{sel} = S_L^1[1:k] \cup S_L^2[1:k] \cup \dots \cup S_L^{n_c}[1:k], \quad (3)$$

where $g(\cdot)$ is a function which conducts S_L^i by sorting the instance of source domain s_L^i with its contained information w_i and grouping them according to their emotion labels y_s^i in source domain. The resulting $\{S_L^i | i = 1, 2, \dots, n_c\}$ is the n_c sorted groups, and n_c is the number of emotion states. For the i -th emotion category, the group of S_L^i contains the top k selected instances with emotion label y_s^i . Therefore, the selected instance data S_{sel} will be used for the subsequent domain adaptation, further reducing data redundancy and improving computational efficiency.

And then, we approach the distribution distance between the target domain T and the source domain S by projecting the data to the common space with a style transfer method. The classifier C can recognize the emotion state for the target data in the common space. Assume that the DE samples of the source and target data are Gaussian distribution, so we project the target data T_L with the Gaussian model:

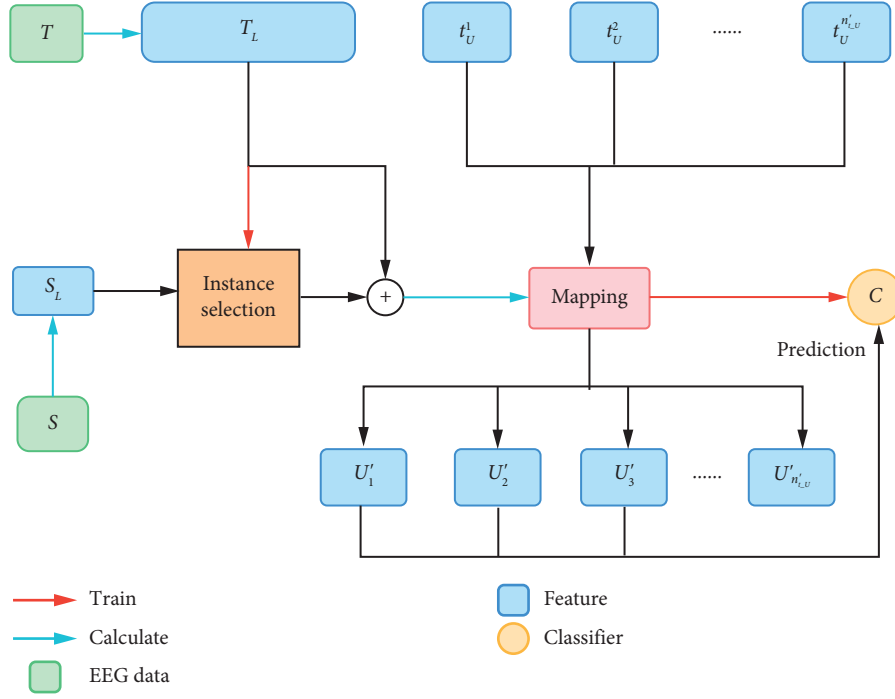


FIGURE 2: The pipeline of the proposed SSTM-IS method.

$$O = \{o^i \in R^m\}, o^i = u_c + \min \left\{ 1, \frac{\rho}{d(t_{L_c}^i, c)} \right\}, \quad i = 1, 2, \dots, n'_{L,c}, \quad (4)$$

$$d(t_{L_c}^i, c) = \sqrt{(t_{L_c}^i - \mu_c)^T \sum_c^{-1} (t_{L_c}^i - \mu_c)},$$

where o^i is the projected value of $t_{L_c}^i$, $d(t_{L_c}^i, c)$ is the distance in emotion label of c and $c = 1, 2, \dots, n_c$, $t_{L_c}^i$ is the target data with the same emotion label c in S_{sel} . Hyperparameter ρ is designed to adjust the deviation between o^i and μ_c .

Assume the affine transformation from o^i to $t_{L_c}^i$ is denoted as $Ao^i + b$, where the parameter matrix $A \in R^{m \times m}$ and $b \in R^m$ can be calculated by optimizing the equation (5):

$$\min_{A \in R^{m \times m}, b \in R^m} \sum_{i=1}^{n'_{L,c}} \|Ao^i + b - t_{L_c}^i\|_2^2 + \beta \|A - I_{m \times m}\|_F^2 + \gamma \|b\|_2^2, \quad (5)$$

where the hyperparameters β and γ control the transfer degree. The STM method [27] will reassign the values of β and γ in each turn, increasing computational complexity. We find that the values of β and γ can be fixed to appropriate values through experiments to save the computing time without degrading the model performance. To simplify the reassigning process of hyperparameters, we fix β and γ as constants and then calculate equation (5) to get the affine matrix.

In the EEG emotion recognition module, we reduce the outliers and minimize the impact of noise on the model by selecting informative instances, which largely prevents overfitting. Besides, by adjusting the slack variable in main classifier SVM for emotion recognition, the complexity of the model can be controlled which also helps to prevent overfitting. According to the description given above, our method can improve the emotion recognition performance and shorten the time of the model calibration for newcomer, meeting the need of application in real-time EEG emotion recognition.

3.3. Emotional Music Generation Module. The existing systems call the third-party software and lost the controllability of the consistency of generated music with the actual one in emotional expression. For the emotional music generation module in the proposed system, we use the recognized emotion states and music structure features as the conditional inputs to the generative network, while adding the perceptual optimization of the emotional classification network branch to generate music which are closer to the real emotional expression.

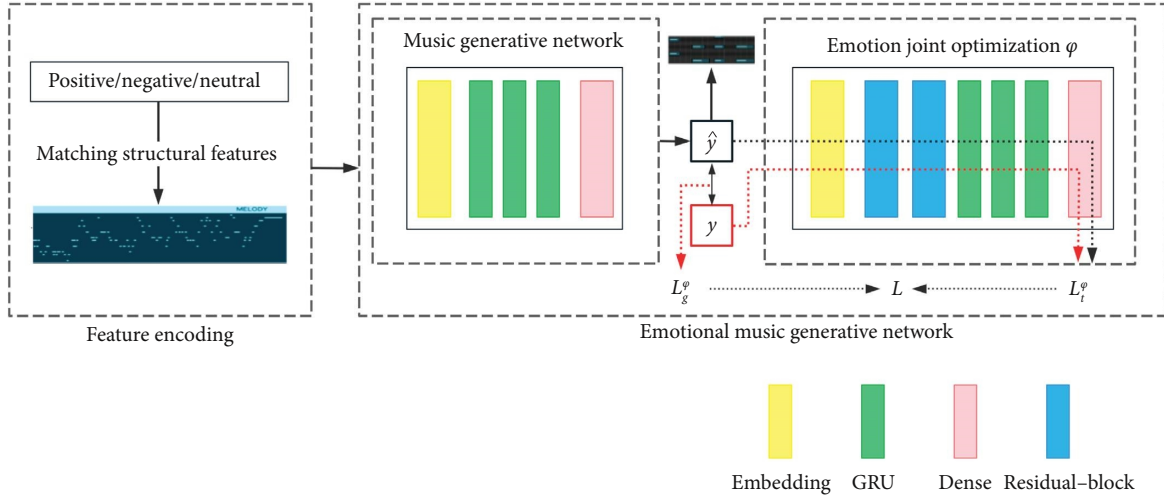


FIGURE 3: The framework of the emotional music generation method. L_g^ϕ represents the loss function of the music generation network and L_t^ϕ is the loss function of the music emotion classification network.

In Figure 3, the music generation network is presented with a feedforward neural network consisting of three parts [32]: input layers, network layers, and a full-connected output layer. Firstly, the input layers encode emotional labels and structural features of music into conditional vectors, which are used as input features for the network layers. In the conditional vector encoding, the emotional labels are mapped to binary vector representations of integer values with the one-hot word vector encoding. The index of integer value is 1, and the rest index positions are 0. Different from the encoding of emotional labels, since the structural features and feature vectors are both MIDI note sequences, they use the same encoding way; that is, the corresponding word dictionaries are constructed based on a series of words that appear in the note sequence, and the conditional vector and feature vector can be obtained through Word2vec word vector encoding. Then, the network layers are composed of three gated recurrent units (GRUs) in order to solve the problem of time dependence in the modelling of note sequences. Here, the number of GRUs is set to 512, and the size of the convolution kernel is set to 3. Compared with LSTM structure, GRU can greatly reduce training time and improve training efficiency while achieving the same performance. Meanwhile, GRU effectively solves the long-term dependency problem in the information transmission process by introducing a gating mechanism, thereby inferring the probability of subsequent states based on the previous state. Finally, the full-connected layer is used to predict the probability distribution of the output note sequence. That is, the softmax activation function σ is used to predict the position with the highest probability of the output logits l_j of the network layer and perform the gradient backpropagation and parameter update of the network, ultimately predicting the actual output note sequence. Among them, the number of units in the fully connected layer must be consistent with the size of the note word dictionary.

Inspired by word vector modelling and combined with the modelling of MIDI note sequences, the proposed module uses the negative maximum likelihood estimation function as the loss function of the music generative network as shown in the following equation:

$$L_g^\phi(p, q) = -\log_p[q],$$

$$p = \frac{1}{a} \sum_{j=1}^a \sigma(l_j[0: d_c-1]), \quad (6)$$

where a is the time-stamp of each event note, d_c is the number of MIDI event notes, and l_j represents the output logits of the generation model. The generator is designed for predicting a distribution p over the next note sequence and $q \in [0, N-1]$ represents the next MIDI event note, where N represents the length of the word dictionary.

In addition, to enhance the realism of the generated emotional music expression, we utilize a pretrained classification network as a loss network to capture different emotional features. The loss function defined as L_t^ϕ calculates the difference between the output \hat{y}_m and the target y_m and computes a scalar value $L_t^\phi(\hat{y}_m, y_m)$ as follows:

$$L_t^\phi = \|\varphi(\hat{y}_m) - \varphi(y_m)\|_2^2, \quad (7)$$

where $\varphi(\cdot)$ is as the feed-forward feature extractor which consists of dense layers and obtains features from output logit, as shown in Figure 3.

With the generation loss L_g^ϕ and classification loss L_t^ϕ , we also design a joint optimization strategy to obtain the consistency of generated music with the actual one in emotional expression, which captures the feature distributions of different emotions and jointly optimizes the distance between the loss function and feature distributions of the generative network, as shown in the following equation:

$$L = \operatorname{argmin}_{\theta} \sum L_{\theta}^p(p, q) + \lambda L_t^p(\hat{y}_m, y_m), \quad (8)$$

where θ represents the optimization parameter, and λ represents the hyperparameter.

In the practical application scenarios, the human emotions changes over time. The proposed system uses the designed real-time EEG emotion recognition module to describe users' emotions more accurately and timely, and then the recognized emotions and music structure features are fed as the conditional inputs to the emotional music generation module. As a result, the proposed system can generate smooth, complete, and exclusive music that is consistent with the emotion of the subject.

4. System Implementation

Figure 4 shows the implementation of the proposed real-time emotional music generation system. It can be seen from Figure 4 that, the implementation process mainly includes four parts: the audio-visual stimulation, EEG signal acquisition, EEG signal processing involving storage and analysis, and result visualization and display. The main work completed includes the design of hardware and software, as well as, the wrapping of emotion recognition and emotional music generation modules.

4.1. Hardware Design. The hardware design of the proposed system is as follows: the audio-visual stimulation provides the subjects with specific stimuli to induce their emotions; the EEG signal acquisition device collects the signals by EEG cap and stores them on the server through receiver; the emotion recognition module running on the server recognizes the emotion state based on the EEG signals; and then the emotional music generation module combines the recognized emotion state with the music structure to generate the exclusive music. Meanwhile, the server sends the EEG features, the results of emotion recognition, and the generated emotional music to monitor for displaying.

4.1.1. Audio-Visual Stimulation. The device used for playing the audio-visual stimuli is a Philips 23.8" monitor with a refresh rate of 75 Hz and equipped with speakers for the sound playing. We select the stimulus videos according to the native language of the subjects, and chose 24 short videos as stimuli and each video stimulus is about 1–3 min with one emotion category of positive, neutral, and negative. In the experiment, the videos are played pseudo-random and the emotions of the neighbouring videos are different.

4.1.2. EEG Signal Acquisition. As shown in Figures 5(a) and 5(b), the portable EEG headset Emotiv EPOC X is used for EEG signals collecting in the proposed system. The cap just contains 14-channel electrodes, allows easy rehydration and prevents interruptions. The sampling frequency of the used EEG headset is 256 Hz and the bandwidth is 0.20–43 Hz. In the EEG signal recording and processing, the system uses the

EmotivPRO software to record the subject's EEG signal, and the raw signals are shown in Figure 5(c).

4.1.3. EEG Signal Processing, Analysis, and Storage. The program for EEG signal processing and analysis runs on a server with an operating system of Window 10, a GPU of NVIDIA GTX 1070, and a CPU of Intel i5. The connection between the server and EEG data acquisition device is made through Bluetooth. For the processing and analysis, the EEG signals are filtered and conducted the operations of artifact removal and principal component analysis to get higher signal-to-noise ratio signals. And then, the server recognizes the emotion state from preprocessed EEG signals and generates the corresponding emotional music. In order to accurately and automatically get the time points corresponding to the EEG signals obtained when watching different emotion-evoking stimuli, we record key markers of the relevant pages triggered by the subjects and transmit them to the server for storage through the HTTP protocol.

4.1.4. Result Visualization and Display. For the visualization and display interface, we use another computer as monitor to connect the server by HTTP protocol for accessing the web page of result visualization and playing the generated emotional music.

4.2. Software Design. Based on the above hardware, we design the software for the proposed system. As shown in Figure 6, the software of the system is mainly divided into three parts: the server, the subject experimental interface, and result visualization interface.

The server is mainly implemented based on the Flask 2.0.0 framework, which integrates the functions of controlling the subject experiment such as experimental process control, subject information storage, and EEG data acquisition, calling the core modules of real-time emotion recognition and emotional music generation, and providing the services relevant to result visualization interface. For the real-time emotion recognition module, the proposed SSTM-IS described in Section 3.2 is wrapped in a module and runs on the server. The EEG data from all of 14 electrodes is used for achieving better emotion recognition performance. While in the field of cognitive neuroscience, the research works [25–41] have shown that the brain areas are closely related to emotions. It is possible to determine which electrodes received more emotional information than others by evaluating the emotion recognition effect of collected signals from electrodes distributed in the different brain areas. In order to further improve the user experience, it is necessary to evaluate which electrodes received more emotional information than others to reduce the number of EEG cap electrodes in future work. For improving the signal-to-noise ratio of the collected EEG signals, we used a bandpass filter with the frequency band of 0.1–50 Hz to remove artifacts from EEG signals and preserve useful emotional information. This way ensures signal quality while performing fast signal preprocessing. In the practical

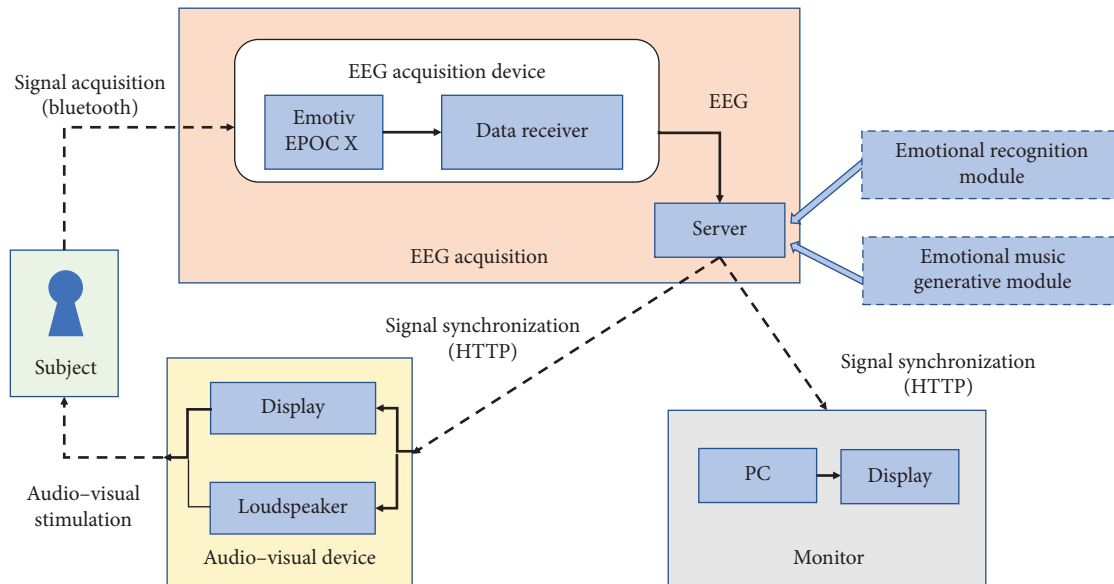


FIGURE 4: The proposed system architecture.

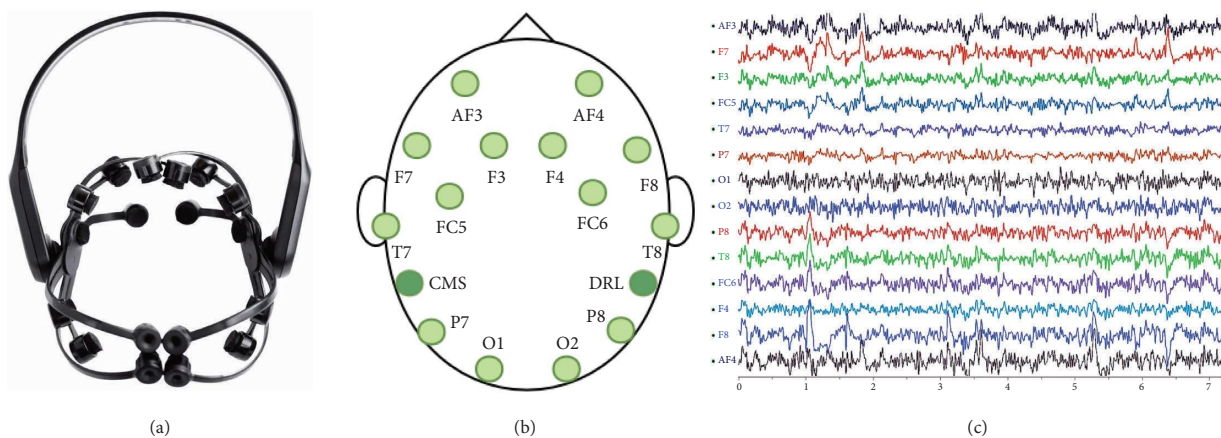


FIGURE 5: The EEG headset used in our system: (a) Emotiv EPOC X, (b) the electrode name and distribution, and (c) the recorded original EEG signals.

application, we put the collected EEG data into the EEG emotion recognition module to output the emotion recognition results in real-time. For the emotional music generation module, with the packaged emotional music generation method described in Section 3.3, the recognized emotion state combining with features of music structure are input to the wrapped module for generating the personal exclusive music.

The subject experimental interface focuses on the acquisition of personal information, playing the audio-visual stimuli, and subjective evaluation of the stimulus videos. The subject interface and experimental flow controlled by server is shown in Figure 7. The interface is coded with HTML5-based web pages, and the interactions are based on jQuery v2.1.1 framework and Bootstrap v3.3.7. As illustrated in Figure 7, the subject first wears an EEG cap, reads the experimental instructions, fills in personal information and then starts the experiment. During the experiment, the

subject first gazes at the white cross in the middle of the screen for resting with the duration of 15 s, and then watches a short video for emotion evoking. The subject is required to complete a subjective emotion questionnaire after watching a video. The above experimental steps are looped, and the subject's EEG signals are recorded. Besides, the subject can choose to stop the experiment at any time, and the EEG data collected under the first three videos are used for model calibration. The whole experimental process flows automatically and will not involve the experimenter.

As shown in Figure 8, the result visualization interface is used to capture the state of subject by displaying the facial expressions of the subject. It also monitors the display of the audio-visual stimuli on the experimental interface. At the same time, the emotion states recognized by emotion recognition module are displayed on the monitoring interface in real time. After a certain stage finish (we set the end of a stimulus video as the stage in the experiment), the

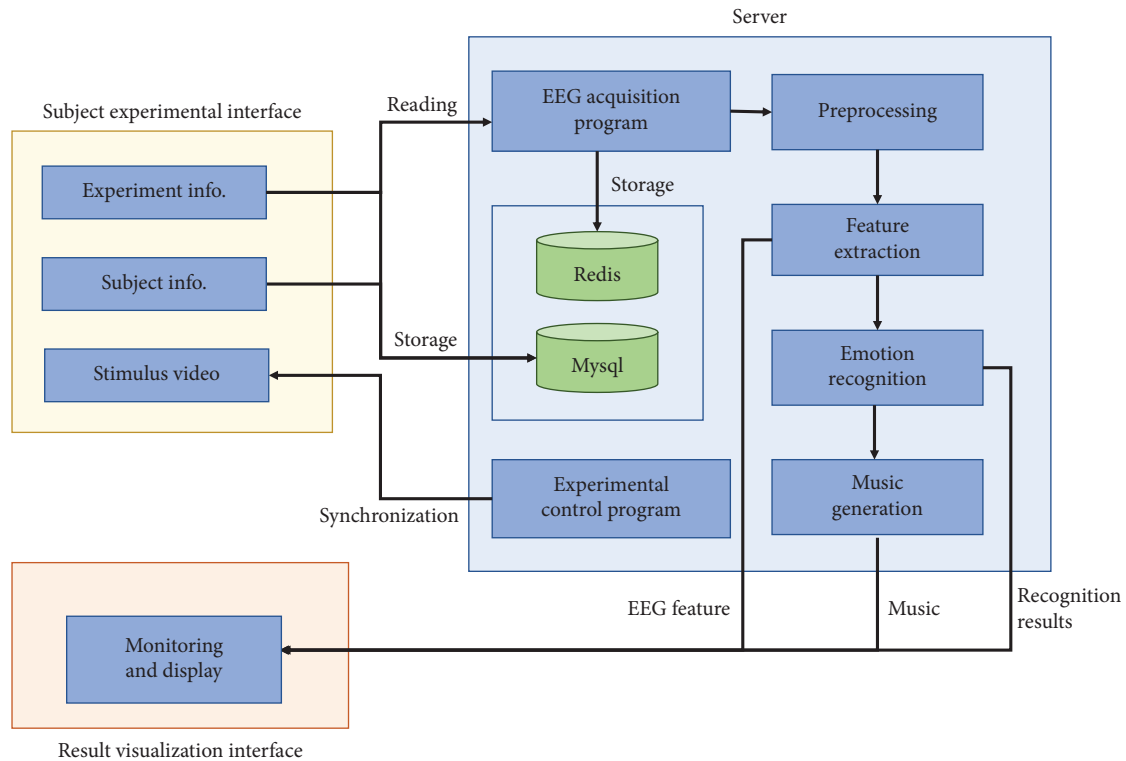


FIGURE 6: The software architecture for the proposed system.

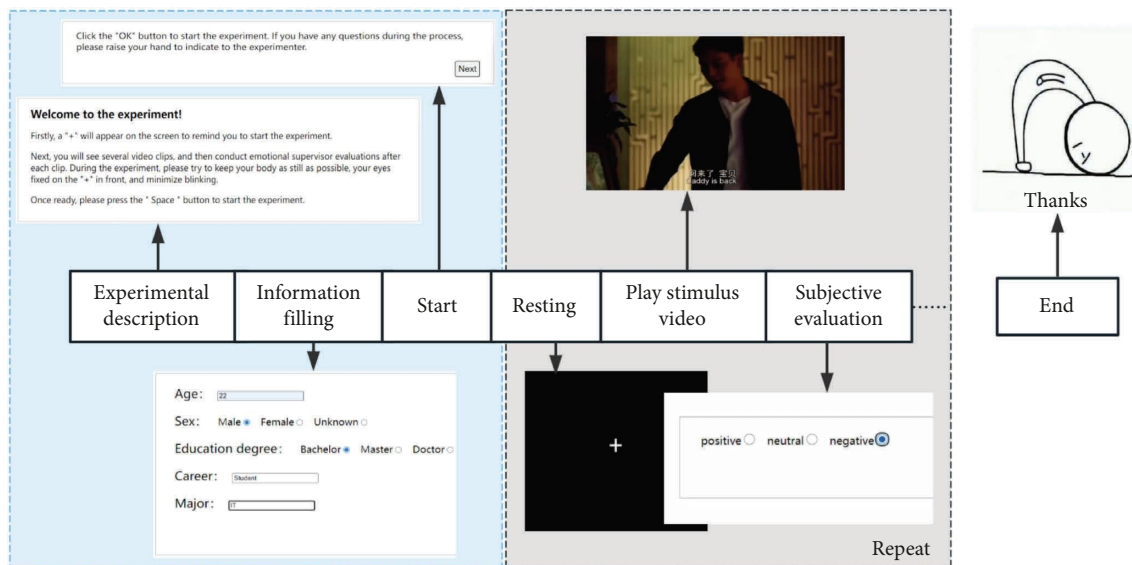


FIGURE 7: The controlling of experimental flow.

emotional music is generated corresponding to the EEG emotion states, and it can be played when the button of “Play” on the visualization interface is clicked.

5. Experiments and Results

This section first introduces the experimental environment of the proposed system, and then conducts off-line and on-line experiments on the EEG emotion recognition module.

The quality of the music generated by the emotion music generation module is also evaluated subjectively and objectively.

5.1. Experimental Environment. The diagram of the experimental environment is shown in Figure 9, which is divided into two parts: the subject room and the experimenter room. The subject room is used to acquire EEG signals. To reduce



FIGURE 8: The result visualization interface.

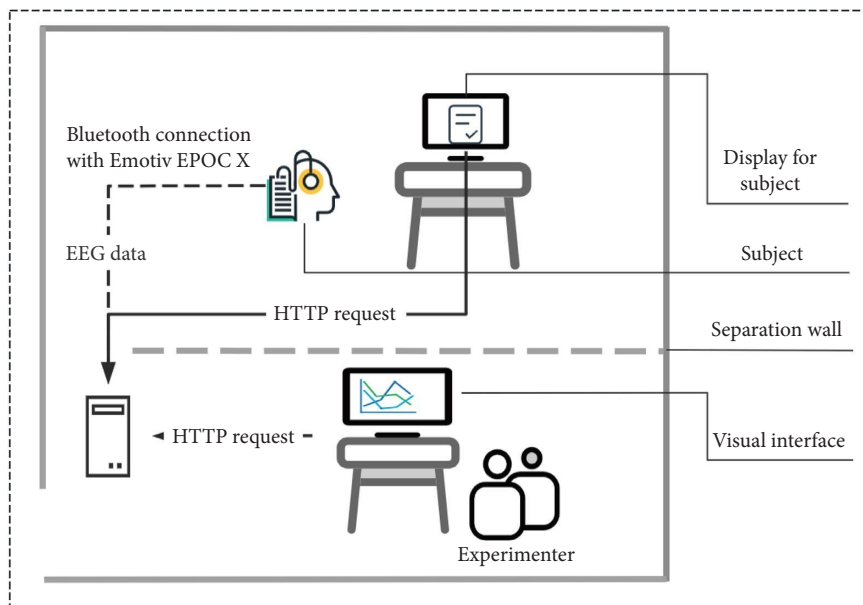


FIGURE 9: The diagram of the experimental environment.

the interference of wireless devices on EEG signals, the hardware devices such as cameras, mice, and keyboards in the subject room are wired to the host computer. The experimenter room is used to monitor the status of the subjects and ensure the progress of the experimental process. The host computer is located in the experimenter room, which is used to record EEG data and provide services for the experimental interface and system display interface of the subjects. The actual experimental environment is shown in Figure 10, and the subject wears the EEG cap and receives the emotional stimuli played on the monitor for conducting experiments in the subject room. The subjects are not asked

to show their facial expressions explicitly, but rather stay still throughout the experiments.

5.2. Experiments on EEG Emotion Recognition

5.2.1. Off-Line Experiments. To verify the performance of the EEG emotion recognition module, we select the public available SEED dataset [25, 26] and the self-collected EEG dataset with 14 electrodes to conduct the off-line experiments. The SEED dataset uses 15 Chinese film clips as stimuli which contains three emotion states of positive,

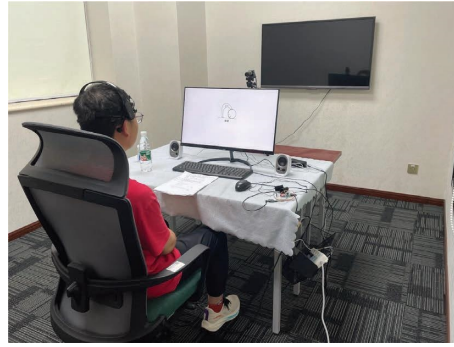


FIGURE 10: The actual experimental environment.

neutral, and negative. Each subject participated in three experiments by using the same stimulus video and playback sequence, with 1-week interval between the two experiments. An ESI NeuroScan System was used at a sampling rate of 1000 Hz from 62-channel electrode cap according to the international 10–20 system, and then they downsampled the signals to 200 Hz and used the 0.3–50 Hz band-pass filter for preprocess. The self-collected dataset recorded the EEG signals by using a 14-channel electrode cap named Emotiv EPOC X, which samples the data at a rate of 256 Hz. 10 subjects participated in the experiment. Considering the subjects' native language, we selected 27 Chinese short videos with emotion states of positive, neutral, and negative as stimuli. Each short video duration is about 1–3 min to avoid subject fatigue. Firstly, we investigate the relationship between the emotion recognition performance and selected instance number. Here for the SEED dataset, the number of instances is incremented in 1000 starting from 500 to all data in turn and is incremented in 500 starting from 500 to all for the self-collected dataset. The experimental results on SEED and self-collected datasets are shown in Figures 11 and 12, respectively. As shown in Figures 11 and 12, the x -axis represents the number of selected instances k as described in equation (3). The boxplot means the recognition accuracy corresponding to the left y -axis, and the blue line denotes the computing time of model calibration corresponding to the right y -axis. It can be seen from Figures 11 and 12 that, the recognition accuracy is not always increasing as the selected number of instances increases. When the selected number k exceeds a certain value, the accuracy tends to flatten out and fluctuates slowly. It indicates that we can select the informative instances to better transfer them with target data for building a more generalized model. And those excessive instances would not be selected to cause redundancy and negative transfer. We can also find from Figures 11 and 12 that on the SEED and self-collected datasets, setting the selected instance number k being 6500 and 5500, respectively in different emotion categories can train a generalized model reaching the performance of $86.78 \pm 6.65\%$ and $77.68 \pm 5.13\%$ with computing time of 7 s and 10 s.

Next, we compare the performance of our EEG emotion recognition module with that of baseline method MS-STM [27]. With leave-one-subject-out validation on the SEED and self-collected datasets, the experiment results are shown

in Figures 13 and 14. It can be seen from Figure 13 that on the SEED dataset, the accuracies of the proposed module are higher than those of MS-STM [27] for most subjects, except taking the fourth, sixth, and seventh subjects' data as the target domain. Even though for those 3 subjects, the recognition results by using MS-STM are better than those of our method, the difference is not significant. While for other 12 subjects, the performance of the proposed method has been significantly improved compared to the baseline MS-STM. Especially on the 5th and 15th subjects' data, the proposed module gains the accuracy improvements of 25.77% and 16.52%, respectively than MS-STM [27]. In addition, we can see from Figure 14 that on the self-collected dataset, our module achieves better performance than the baseline MS-STM on 9 subjects' data, except the ninth subject's data as the target domain. Especially when taking the 10th subject's data as target domain, the accuracy of our module is 11.61% higher than MS-STM.

In order to demonstrate the effectiveness of the proposed module, we also compare the performance of our method with several representative methods on the SEED and self-collected datasets. The results of accuracy and computing time for these methods are shown in Table 1. Here, MS-MDA [21], DResNet [22], PPDA [23], RGNN [44], BiHDM [45], and Saliency [46] are introduced as deep learning algorithms. What's more, we reproduct the results of MS-MDA [21] with the offering codes by its office. The accuracies and computing time of TCA [42], CORAL [43], MS-STM [21], and FOIT [28] on SEED are referenced from the results given in [28]. In addition, the results of TCA [42], CORAL [43], MS-MDA [21], and MS-STM [26] use the reproductions of open-source codes and test on the self-collected dataset. Besides, SSTM indicates that the emotion recognition module removes the part of instance selection and uses all the source domain data without selecting, and Instance-sel represents only to retain the instance selection part without transfer learning.

From Table 1, the results show that the proposed emotion recognition module with SSTM-IS can achieve superior performance compared to all competitive methods on both of the SEED and self-collected dataset datasets. On the SEED dataset, the proposed emotion recognition module reaches $86.78 \pm 6.65\%$, which is 4.28% higher than the baseline method MS-STM [27] with the standard deviation decreased by 52.36%. The performance of our module is

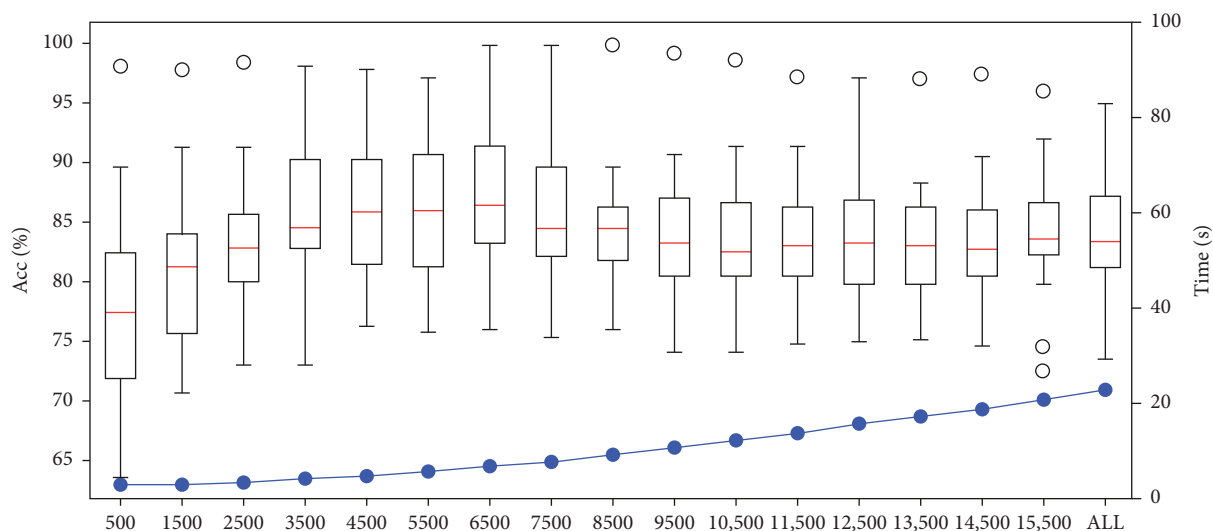


FIGURE 11: The performance of the EEG emotion recognition module on different number of instances on SEED dataset.

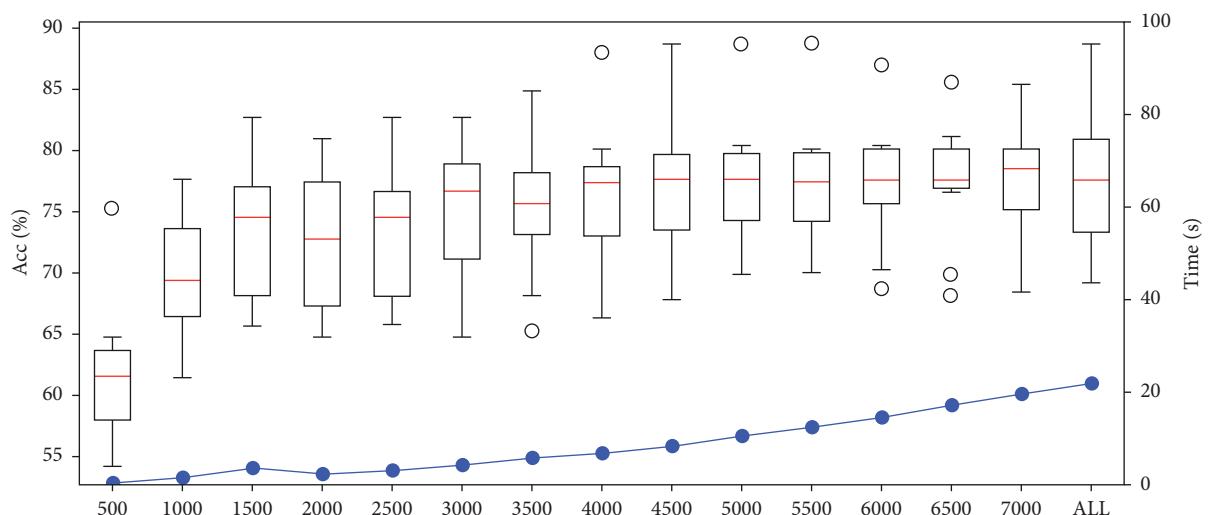


FIGURE 12: The performance of the EEG emotion recognition module on different number of instances on self-collected dataset.

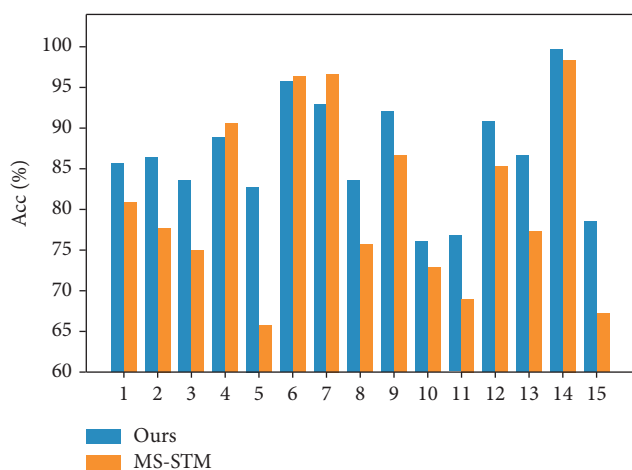


FIGURE 13: The comparison of the EEG emotion recognition module and MS-STM [27] with leave-one-subject-out on SEED dataset.

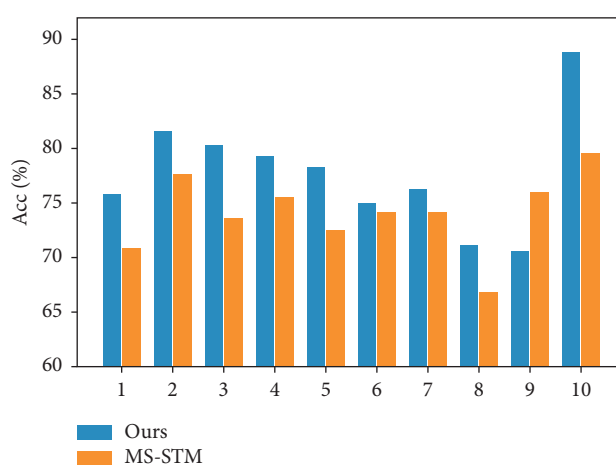


FIGURE 14: The comparison results of the EEG emotion recognition module and MS-STM [27] with leave-one-subject-out on self-collected dataset.

TABLE 1: Performance comparisons with competing methods on SEED and self-collected datasets.

Dataset	Method	Acc (%)	Runtime (s)
SEED	TCA [42]	64.24 ± 15.34	298
	CORAL [43]	63.59 ± 7.60	19
	ASFM [23]	83.51 ± 10.18	—
	MS-STM [27]	83.22 ± 13.96	776
	FOIT [28]	82.05 ± 12.36	32
	MS-MDA [21]	85.04 ± 7.85	1959
	DResNet [22]	85.30 ± 8.00	—
	PPDA_NC [23]	85.40 ± 7.10	—
	PPDA [23]	86.70 ± 7.10	—
	RGNN [44]	85.30 ± 6.70	—
	BiHDM [45]	85.40 ± 7.50	—
	Saliency [46]	84.11 ± 2.90	—
	SSTM	84.02 ± 5.38	22
Self-collected	Instance-set	79.66 ± 8.15	5
	SSTM-IS	86.78 ± 6.65	7
	TCA [42]	47.04 ± 4.91	203
	CORAL [43]	46.26 ± 2.56	31
	MS-STM [27]	67.83 ± 4.75	9
	MS-MDA [21]	68.98 ± 4.91	347
	SSTM	76.96 ± 4.08	9
	Instance-sel	70.31 ± 5.62	6
	SSTM-IS	77.68 ± 5.13	10

Note: The significance of the bold values is to highlight the experimental effect of the method presented in this paper.

77.68 ± 5.13%, which is 14.52% higher than baseline MS-STM [27] on the self-collected dataset. Furthermore, the module uses a few target data to calibrate the model for a newcomer in 7 s and 10 s on the SEED and self-collected datasets respectively. On the other hand, by comparing the proposed SSTM-IS module with SSTM and Instance-sel which just uses either transfer learning or instance selection, we can find that the proposed module which introduces instance selection into transfer learning can reach the best performance on the two datasets, also verifying the effectiveness of our module on EEG emotion recognition.

5.2.2. On-Line Experiments. In order to verify the effectiveness of the real-time EEG emotion recognition module in the proposed system, we also conduct an experiment of online EEG emotion recognition in actual environment for 4 subjects. Following the experimental process shown in Figure 7, we use the EEG data collected when watching the first three stimulus videos for a rapid model calibration, including selecting the most informative instances from the stored EEG data, and conducting transfer learning between the selected instances and calibration data. Once the model calibration is completed, the EEG data from the subject are divided into 20-s segments for real-time emotion recognition. Figure 15 shows the experimental results of 4 subjects on each segment of 20 s when watching 24 video stimuli, where the recognition results for each time segment are averaged cumulatively. The *x*-axis represents the segment number of time, and the orange, grey, and white blocks represent that the current time period is under the video stimuli with positive, negative, and neutral emotions, respectively.

It can be seen from Figure 15 that, the recognition accuracies of 4 subjects fluctuate between 65%–85% for most stimulus videos. And for the same stimulus video, the

performance of all the subjects is generally consistent. Concretely, when watching the first positive (orange colour) and second negative (grey colour) emotion stimulus videos, the subjects' emotions are well aroused, whereas for the third neutral (white colour) and fifth negative (grey colour) emotion stimulus videos, the subjects' emotion reactions are not obvious. This occurrence also indirectly illustrates that the content of stimulus video has a significant impact on the performance of emotion recognition. And thus, it is necessary to consider the emotion analysis of stimulus videos as a modality to assist EEG emotion recognition in future work.

5.3. Experiments on Emotional Music Generation. In order to verify the effectiveness of the emotional music generation module in the proposed system, we compare our method with the existing representative algorithms LSTM + GA [33] and CP transformer [36] from both objective and subjective aspects, as shown in Table 2. In terms of objective evaluation, the pitch range (PR), number of pitch classes (NPC), and polyphonic performance (POLY) are selected as the quality evaluation indicators for the generated music. In addition, due to the significant impact of subjective perception on the musical emotions, a subjective experiment is also designed to evaluate the exactitude of emotion expression in the generated music. In terms of subjective evaluation, the Humanness (authenticity of emotion expression), Richness (richness of music content), and Overall (overall quality of music) are selected as the evaluation indicators for generated music. The subjective experiments are collected from 10 audiences, where the test set includes 9 music clips with three emotion categories of positive, neutral, and negative, and each category contains 3 clips generated by LSTM + GA [33], CP transformer, [36] and the proposed method, respectively.

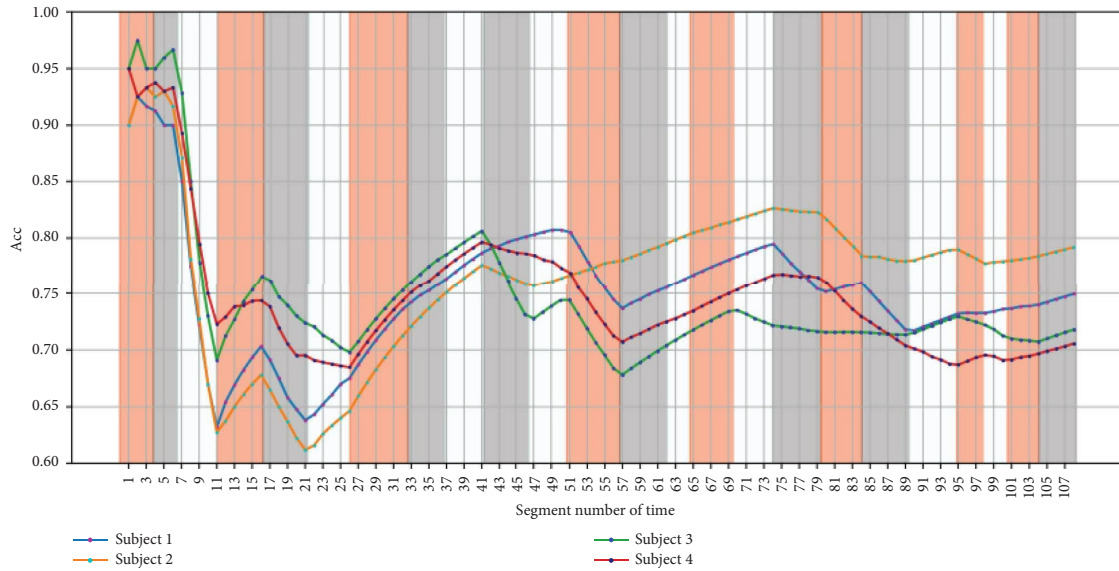


FIGURE 15: The on-line experimental results of 4 subjects on each segment of 20 s when watching 24 video stimuli.

TABLE 2: Temperature and wildlife count in the three areas covered by the study.

Method	Objective evaluation			Subjective evaluation		Overall
	PR	NPC	POLY	Humanness	Richness	
EMOPIA [36]	51.00	8.48	5.90	—	—	—
LSTM + GA [33]	59.10	9.27	3.39	2.59	2.74	2.60
CP transformer [36]	49.60	8.54	4.40	3.31	3.22	3.26
Ours	50.10	8.27	4.76	3.48	3.72	3.84

Note: The significance of the bold values is to highlight the experimental effect of the method presented in this paper.

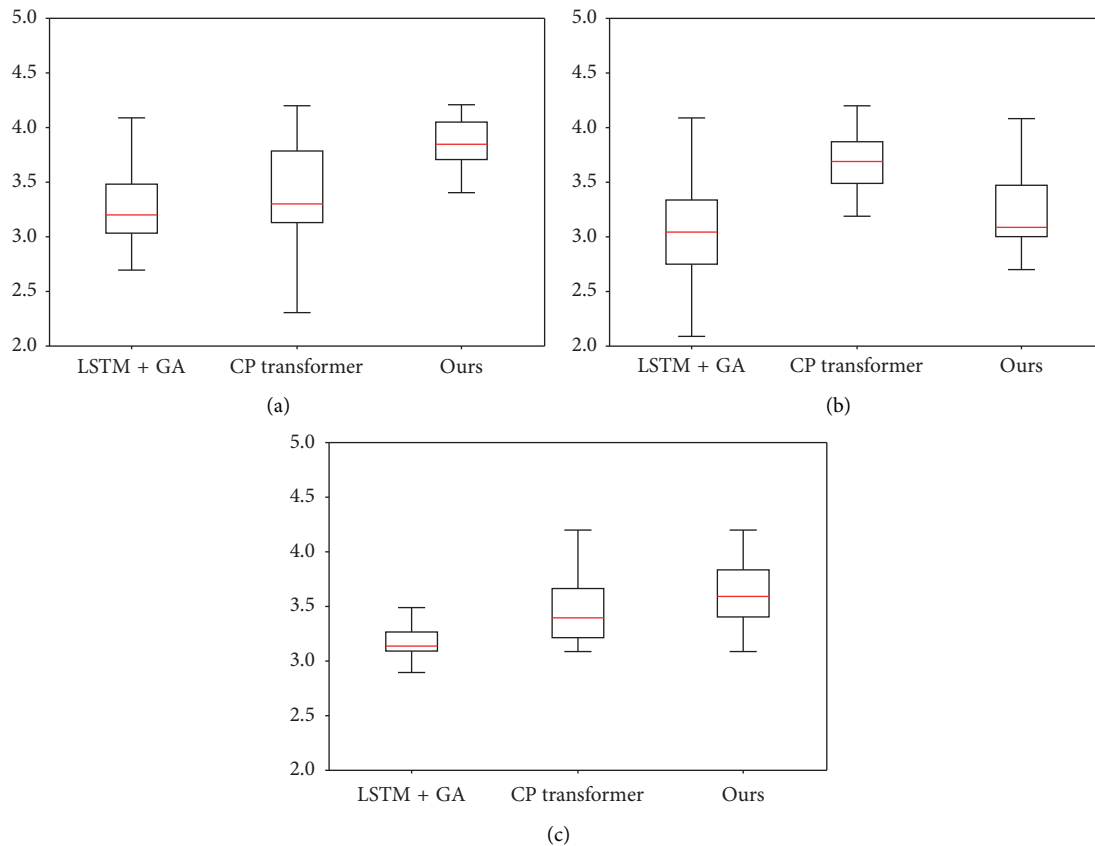


FIGURE 16: The comparison results of subjective evaluation on emotional consistency among the proposed method and existing typical methods [33, 36]: (a) positive, (b) neutral and (c) negative.

It can be seen from Table 2 that, compared to the existing emotional music generation algorithms, the performance of our method is closest to the dataset of EMOPIA [36] on two indicators of PR and POLY and second closest in NPC for the objective aspect. This indicates that our method not only ensures the quality of generated music, but also restores the real emotion of music to the greatest extent. From the subjective aspect, the emotional music generated by our algorithm has improved 34.36%, 35.77%, and 47.69% over LSTM + GA [33] in terms of Humanness, Richness, and Overall, respectively, and 5.14%, 15.53%, and 17.79% over CP transformer [36]. It is evidence that compared with LSTM + GA [33] and CP transformer [36], our method can achieve a significant improvement in all the subjective evaluation indicators, especially in Richness and Overall.

In order to further verify that the music generated by the proposed algorithm is resonant to the audiences emotionally, we further design a subjective experiment to evaluate the consistency between the emotions of audiences and music. Figure 16 shows the comparison results of emotional consistency scores among our method and existing representative algorithms [33, 36] in positive, neutral, and negative emotion categories. It can be seen from Figure 16 that, compared to the existing representative algorithms LSTM + GA [33] and CP transformer [36], the proposed method achieves the highest emotional consistency score in positive and negative emotions, while is inferior to CP transformer [36] in the category of neutral. It indicates that the emotional music generated based on structural features is capable of obtaining a higher degree of emotional resonance among audiences. Trained by the EMOPIA dataset, the emotional music generated by our system is monophonic piano solo and belongs to the POP genre. Some samples of the generated music by our system are provided from the link: https://8.146.209.97/music_sample/generated_music_samples.zip.

6. Conclusions

In this paper, we present and implement an EEG signal-driven real-time emotional music generation system. The proposed system is able to realize real-time emotion recognition by quickly obtaining a model suitable for the new user through a short calibration time. The recognized emotion states are input together with structural features of music to a generative network as conditions to generate exclusive music, which is closer to the real emotional expression. In order to achieve real-time emotion recognition on a new subject, an SSTM-IS algorithm is proposed to improve the recognition accuracy to $86.78 \pm 6.65\%$ on the SEED dataset and shorten the computing time to 7 s. In the emotional music generation module, an emotional music generation network is built based on the emotion recognition results, breaking the limitations of existing systems to call the third-party software and controlling the emotional expression of generated music being consistent with the real ones. In addition, the proposed system uses a wireless portable EEG device of Emotiv EPOC X with physiological saline humidity sensors to collect EEG signals, which is

convenient to carry and operate, greatly reducing the difficulty of EEG signal collection.

Our work still has some limitations, which should be considered in future works. Firstly, the proposed EEG emotion recognition module needs to use a few of data from newcomer for model calibration, which reduces the subject experience. In the future, the module can be explored without calibration. Secondly, the research works in cognitive neuroscience have shown that the brain areas are closely related to emotions, it is necessary to evaluate which electrodes received more emotional information than others to reduce the number of EEG cap electrodes to further improve the user experience. Thirdly for the proposed system, it is also necessary to adapt the result visualization interface on mobile, further improving the users' experience of mobile devices. In addition, the system can integrate the collection of other physiological signals for subsequent use to jointly improve system performance.

Data Availability Statement

The publicly available dataset of SEED can be found here: <https://bcmi.sjtu.edu.cn/home/seed/index.html> and we have obtained the permission to use the dataset in our research. The self-collected EEG data used to support the findings of this study is available from the corresponding author upon request.

Consent

The written informed consent has been obtained from the individuals for publication of the identifiable images included in Figures 8 and 10.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This work is supported by the National Natural Science Foundation of China under Grant No. 62271455 and the Fundamental Research Funds for the Central Universities under Grant No. CUC18LG024.

References

- [1] K. J. Kemper and S. C. Danhauer, "Music as Therapy," *Southern Medical Journal* 98, no. 3 (2005): 282–288, <https://doi.org/10.1097/01.smj.0000154773.11986.39>.
- [2] E. R. Miranda, "Brain-Computer Music Interface for Composition and Performance," *International Journal on Disability and Human Development* 5, no. 2 (2006): 119–126, <https://doi.org/10.1515/ijdh.2006.5.2.119>.
- [3] J. J. Vidal, "Toward Direct Brain-Computer Communication," *Annual Review of Biophysics & Bioengineering* 2, no. 1 (1973): 157–180, <https://doi.org/10.1146/annurev.bb.02.060173.001105>.
- [4] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-Based Brain-Computer Interface for Cursor Control," *Electroencephalography and Clinical Neurophysiology*

- 78, no. 3 (1991): 252–259, [https://doi.org/10.1016/0013-4694\(91\)90040-b](https://doi.org/10.1016/0013-4694(91)90040-b).
- [5] B. Hamadicharef, “Brain-Computer Interface Literature-A Bibliometric Study,” in *10th International Conference on Information Science, Signal Processing and Their Applications* (Kuala Lumpur, Malaysia, May 2010), 626–629.
- [6] W. R. Mu and B. L. Lu, “Examining Four Experimental Paradigms for EEG-Based Sleep Quality Evaluation with Domain Adaptation,” in *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (Montreal, QC, Canada, July 2020), 5913–5916.
- [7] W. Wu, W. Sun, Q. M. J. Wu, et al., “Multimodal Vigilance Estimation Using Deep Learning,” *IEEE Transactions on Cybernetics* 52, no. 5 (2022): 3097–3110, <https://doi.org/10.1109/tcyb.2020.3022647>.
- [8] R. Accordino, R. Comer, and W. B. Heller, “Searching for Music’s Potential: A Critical Examination of Research on Music Therapy With Individuals With Autism,” *Research in Autism Spectrum Disorders* 1, no. 1 (2007): 101–115, <https://doi.org/10.1016/j.rasd.2006.08.002>.
- [9] S. Ehrlich, C. Guan, and G. Cheng, “A Closed-Loop Brain-Computer Music Interface for Continuous Affective Interaction,” in *International Conference on Orange Technologies* (2017), 176–179.
- [10] S. K. Ehrlich, K. R. Agres, C. Guan, G. Cheng, and J. Najbauer, “A Closed-Loop, Music-Based Brain-Computer Interface for Emotion Mediation,” *PLoS One* 14, no. 3 (2019): e0213516, <https://doi.org/10.1371/journal.pone.0213516>.
- [11] T. Grimm and G. Kreutz, “Music Interventions and Music Therapy in Disorders of Consciousness—A Systematic Review of Qualitative Research,” *The Arts in Psychotherapy* 74 (2021): 101782, <https://doi.org/10.1016/j.aip.2021.101782>.
- [12] E. L. Sema and C. Baoul, “Effects of Progressive Muscle Relaxation Training with Music Therapy on Sleep and Anger of Patients at Community Mental Health Center,” *Complementary Therapies in Clinical Practice*, 43, 2021.
- [13] N. V. Kimmattkar and B. V. Babu, “Novel Approach for Emotion Detection and Stabilizing Mental State by Using Machine Learning Techniques,” *Computers* 10, no. 3 (2021): 37, <https://doi.org/10.3390/computers10030037>.
- [14] A. Pinegger, H. Hiebel, S. C. Wriessnegger, G. R. Müller-Putz, and S. Pierangelo, “Composing Only by Thought: Novel Application of the P300 Brain-Computer Interface,” *PLoS One* 12, no. 9 (2017): e0181584, <https://doi.org/10.1371/journal.pone.0181584>.
- [15] D. Wu, C. Y. Li, and D. Z. Yao, “Scale-Free Music of the Brain,” *PLoS One* 4, no. 6 (2009): e5915, <https://doi.org/10.1371/journal.pone.0005915>.
- [16] S. Makeig, G. Leslie, T. Mullen, D. Sarma, and C. Kothe, “First Demonstration of a Musical Emotion BCI,” in *4th International Conference on Affective Computing and Intelligent Interaction* (Memphis, TN, October 2011), 487–496.
- [17] L. Bellier, A. Llorens, D. Marciano, et al., “Music Can Be Reconstructed from Human Auditory Cortex Activity Using Nonlinear Decoding Models,” *PLoS Biology* 21, no. 8 (2023): e3002176, <https://doi.org/10.1371/journal.pbio.3002176>.
- [18] S. Ran, W. Zhong, D. T. Duan, L. Ye, and Q. Zhang, “SSTM-IS: Simplified STM Method Based on Instance Selection for Real-Time EEG Emotion Recognition,” *Frontiers in Human Neuroscience* 17 (2023): 1132254, <https://doi.org/10.3389/fnhum.2023.1132254>.
- [19] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, “Domain Adaptation for EEG Emotion Recognition Based on Latent Representation Similarity,” *IEEE Transactions on Cognitive and Developmental Systems* 12, no. 2 (2020): 344–353, <https://doi.org/10.1109/tcds.2019.2949306>.
- [20] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, “A Bi-Hemisphere Domain Adversarial Neural Network Model for EEG Emotion Recognition,” *IEEE Transactions on Affective Computing* 12, no. 2 (2021): 494–504, <https://doi.org/10.1109/taffc.2018.2885474>.
- [21] H. Chen, M. Jin, Z. Li, C. Fan, J. Li, and H. He, “MS-MDA: Multisource Marginal Distribution Adaptation for Cross-Subject and Cross-Session EEG Emotion Recognition,” *Frontiers in Neuroscience* 15 (2021): 778488, <https://doi.org/10.3389/fnins.2021.778488>.
- [22] B. Q. Ma, H. Li, W. L. Zheng, and B. L. Lu, “Reducing the Subject Variability of EEG Signals With Adversarial Domain Generalization,” in *26th International Conference on Neural Information Processing* (Sydney, NSW, Australia: Springer, December 2019), 30–42.
- [23] L. M. Zhao, X. Yan, and B. L. Lu, “Plug-and-Play Domain Adaptation for Cross-Subject EEG-Based Emotion Recognition,” *Proceedings of the AAAI Conference on Artificial Intelligence* 35, no. 1 (2021): 863–870, <https://doi.org/10.1609/aaai.v35i1.16169>.
- [24] X. Chai, Q. Wang, Y. Zhao, et al., “A Fast, Efficient Domain Adaptation Technique for Cross-Domain Electroencephalography (EEG)-Based Emotion Recognition,” *Sensors* 17, no. 5 (2017): 1014, <https://doi.org/10.3390/s17051014>.
- [25] W. L. Zheng and B. L. Lu, “Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition With Deep Neural Networks,” *IEEE Transactions on Autonomous Mental Development* 7, no. 3 (2015): 162–175, <https://doi.org/10.1109/tamd.2015.2431497>.
- [26] R. N. Duan, J. Y. Zhu, and B. L. Lu, “Differential Entropy Feature for EEG-Based Emotion Classification,” in *The 6th International IEEE EMBS Conference on Neural Engineering* (San Diego, CA, November 2013), 81–84.
- [27] J. P. Li, S. Qiu, Y. Y. Shen, C. Liu, and H. He, “Multisource Transfer Learning for Cross-Subject EEG Emotion Recognition,” *IEEE Transactions on Cybernetics* 50, no. 7 (2020): 3281–3293, <https://doi.org/10.1109/tcyb.2019.2904052>.
- [28] H. Chen, H. He, T. Cai, and J. Li, “Enhancing EEG-Based Emotion Recognition With Fast Online Instance Transfer,” in *Integrating Artificial Intelligence and IoT for Advanced Health Informatics: AI in the Healthcare Sector* (Springer, 2022), 141–160.
- [29] T. Tuncer, S. Dogan, and A. Subasi, “A New Fractal Pattern Feature Generation Function Based Emotion Recognition Method Using EEG,” *Chaos, Solitons & Fractals* 144, no. 144 (2021): 110671–110685, <https://doi.org/10.1016/j.chaos.2021.110671>.
- [30] A. Subasi, T. Tuncer, S. Dogan, D. Tanko, and U. Sakoglu, “EEG-Based Emotion Recognition Using Tunable Q Wavelet Transform and Rotation Forest Ensemble Classifier,” *Biomedical Signal Processing and Control* 68 (2021): 102648–102656, <https://doi.org/10.1016/j.bspc.2021.102648>.
- [31] A. Dogan, M. Akay, P. D. Barua, et al., “PrimePatNet87: Prime Pattern and Tunable Q-Factor Wavelet Transform Techniques for Automated Accurate EEG Emotion Recognition,” *Computers in Biology and Medicine* 138 (2021): 104867–104878, <https://doi.org/10.1016/j.combiomed.2021.104867>.
- [32] L. Ma, W. Zhong, X. Ma, L. Ye, and Q. Zhang, “Learning to Generate Emotional Music Correlated With Music Structure Features,” *Cognitive Computation and Systems* 4, no. 2 (2022): 100–107, <https://doi.org/10.1049/ccs2.12037>.
- [33] L. N. Ferreira and J. Whitehead, “Learning to Generate Music With Sentiment,” in *22nd International Society for Music*

- Information Retrieval Conference* (Milan, Italy, November 2020).
- [34] X. Tan, M. Antony, and H. Kong, "Automated Music Generation for Visual Art through Emotion," in *International Conference on Computer and Communications* (2020), 247–250.
 - [35] A. Radford, R. Jozefowicz, and I. Sutskever, "Learning to Generate Reviews and Discovering Sentiment," (2017), <https://arxiv.org/abs/1704.01444>.
 - [36] H. T. Hung, J. Ching, S. Doh, N. Kim, J. Nam, and Y. H. Yang, "Emopia: a Multi-Modal Pop Piano Dataset for Emotion Recognition and Emotion-Based Music Generation," in *23rd International Society for Music Information Retrieval Conference* (Delft, Netherlands, November 2021).
 - [37] E. Koh and S. Dubnov, "Comparison and Analysis of Deep Audio Embeddings for Music Emotion Recognition," (2021), <https://arxiv.org/abs/2104.06517>.
 - [38] I. Hossain, A. Khosravi, I. Hettiarachchi, and S. Nahavandi, "Multiclass Informative Instance Transfer Learning Framework for Motor Imagery-Based Brain-Computer Interface," *Computational Intelligence and Neuroscience* 2018 (2018): 1–12, <https://doi.org/10.1155/2018/6323414>.
 - [39] X. Y. Zhang and C. L. Liu, "Writer Adaptation With Style Transfer Mapping," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, no. 7 (2013): 1773–1787, <https://doi.org/10.1109/TPAMI.2012.239>.
 - [40] L. C. Shi, Y. Y. Jiao, and B. L. Lu, "Differential Entropy Feature for EEG-Based Vigilance Estimation," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Osaka, Japan, July 2013), 6627–6630.
 - [41] P. Fusar-Poli, A. Placentino, F. Carletti, et al., "Functional Atlas of Emotional Faces Processing: a Voxel-Based Meta-Analysis of 105 Functional Magnetic Resonance Imaging Studies," *Journal of psychiatry & neuroscience: JPN* 34, no. 6 (2009): 418–432.
 - [42] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain Adaptation via Transfer Component Analysis," *IEEE Transactions on Neural Networks* 22, no. 2 (2011): 199–210, <https://doi.org/10.1109/tnn.2010.2091281>.
 - [43] B. Sun and K. Saenko, "Deep Coral: Correlation Alignment for Deep Domain Adaptation," in *European Conference on Computer Vision Workshops: Computer Vision* (Amsterdam, The Netherlands: Springer, October 2016), 443–450.
 - [44] P. Zhong, D. Wang, and C. Miao, "EEG-Based Emotion Recognition Using Regularized Graph Neural Networks," *IEEE Transactions on Affective Computing* 13, no. 3 (2022): 1290–1301, <https://doi.org/10.1109/taffc.2020.2994159>.
 - [45] Y. Li, L. Wang, W. M. Zheng, et al., "A Novel Bi-Hemispheric Discrepancy Model for EEG Emotion Recognition," *IEEE Transactions on Cognitive and Developmental Systems* 13, no. 2 (2021): 354–367, <https://doi.org/10.1109/tcds.2020.2999337>.
 - [46] V. Delvigne, A. Facchini, H. Wannous, T. Dutoit, L. Ris, and J. P. Vandeborre, "A Saliency Based Feature Fusion Model for EEG Emotion Estimation," in *44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (Glasgow, Scotland, UK, July 2022), 3170–3174.