Journal of Multimedia Information System VOL. 11, NO. 1, March 2024 (pp. 109-118): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2024.11.1.109

Improving Musical Expression by Capturing Psychological Changes with CNN Support

Xiaochan Li[1*], Yi Shi[1], Daohua Pan[2]

Abstract

In vocal performance, capturing psychological changes through electroencephalography (EEG) and integrating them with music can improve musical expression. We propose an EEG-based emotion recognition model using continuous convolutional neural networks (CNN). EEG signals in different frequency bands are extracted as features using differential entropy. An EEG enhancement method reassigns variable importance to channels by suppressing redundant information. The model is evaluated on the DEAP dataset for emotion recognition. Combining a rest-state baseline signal before each trial significantly improves accuracy. Following the simulation results, the suggested continuous CNN model has an average classification accuracy of 95.36% and 95.31 for arousal and valence on 22 channels, respectively. This is close to the average accuracy of 32 channels. Likewise, it can help capture the psychological changes of vocal performers and improve musical expression.

Key Words: Vocal Performance, EEG, Psychological Changes, DEAP, CNN, Differential Entropy.

I. INTRODUCTION

Many people focus on singing skills, ignoring the impact of psychological changes and the importance of psychology in vocal performance [1]. Psychology, as the theoretical basis of the performer's mental state, fully exerts the role of guidance to help performers overcome tension and relieve psychological pressure. Recently, people's research on psychology has become more and more profound, and more performers have realized the influence of psychology on vocal performance [2-3]. A perfect vocal performance requires not only the support of the performer's vocal skills but also solid psychological support to avoid mistakes. Psychology is used as a theoretical support to fully understand the characteristics of the psychological activities of vocal performances, scientifically and effectively clear the performers' minds, overcome the nervous state, and achieve good vocal performance effects. Therefore, psychology plays a vital role in vocal performance.

Psychology mainly acts on the psychological aspects of vocal performers and provides a theoretical basis for the psychological changes of performers in vocal performance [4]. We all know that a good mentality is essential for performers. Psychology is based on the characteristics of performers' activities, using scientific methods to help performers have a good mentality, reduce the blindness of learning, and improve learning efficiency [5]. It allows performers to perform at normal or exceptional levels to achieve excellent results. Good vocal performance requires rhythmic breathing, muscle coordination, and physical performance. As we all know, a bad mental state will disrupt the performer's breathing rhythm and cause muscle tension. Therefore, an excellent vocal performance mental state is an integral part of vocal performance, and the mental state of vocal performance directly affects the performer's performance.

Vocal performance is a complete manifestation of the human brain dominating the vocal, auditory, and respiratory systems. The quality of vocal performance is related to voice and singing techniques and is also affected by objective factors such as emotion [6-8]. Therefore, vocal performance not only tests the performer's singing techniques but also tests the psychological changes. The effect and emotion of vocal performance are inseparable. Currently, most studies agree with six basic emotions: happiness, sadness, surprise, fear, anger, and disgust [9]. Singing is an expression of emotion, so a person's emotions will be shown in singing. Usually, positive emotions will give people a comfortable feeling so that the audience will empathize with the music while enjoying the music. However, for negative

Manuscript received December 28, 2023; Revised February 21, 2024; Accepted February 25, 2024. (ID JMIS-23M-12-052)

Corresponding Author (*): Xiaochan Li, +86-15343277018, 15343277018@163.com
[1]Huaihua Normal College, Huaihua, China, 15343277018@163.com, shiyi18074597958@163.com
[2]Harbin Institute of Technology, Harbin, China, pandaohua@ftcl.hit.edu.cn

emotions, the same song brings different effects to the audience, with psychological changes [10-11]. Some performers are in negative emotions during the performance, causing vibrato and treble to go up, which seriously affects the quality of the performance. While some performers are unable to have a good performance due to poor rest and inertia in their voices. This happens from time to time, that personal emotions affect the level of vocal singing.

Emotion is a psychological and physiological reaction that occurs when people are stimulated by the outside or themselves. It affects people's daily activities, such as cognition, perception, and rational decision-making [12-13]. Recognizing and expressing emotions is the basis of communication and understanding. To realize human-computer interaction (HCI), computers must have emotional perception and recognition capabilities similar to humans [14]. Emotion recognition, one of the basic tasks to realize comprehensive computer intelligence, holds an important position in HCI and has broad application prospects in artificial intelligence, healthcare, distance education, military and other fields [15-19]. Among many emotion recognition methods, emotion recognition based on facial expression or voice is easy to distinguish but easy to disguise. In contrast, emotion recognition based on physiological signals is not easy to disguise and contains more information. Electroencephalography (EEG) is one of them since it has a high temporal resolution and can represent the electrophysiological activities of brain nerve cells on the scalp or surface of the cerebral cortex [20]. By capturing psychological changes through EEG, psychology and vocal performance will be integrated into each other, which is believed to improve the performer's musical expression. Many scholars regard EEG emotion recognition as the focus of research and combine machine learning and deep learning to achieve high accuracy [21-23]. Compared with machine learning, deep learning can extract features deeper. With the update and iteration of computer vision recognition technology, the recognition and classification effect of Convolutional Neural Networks (CNN) is also continuously improving [24-25]. CNN has a powerful ability to extract spatial features. Moreover, it has higher efficiency for distributing feature weights among different channels. As a result, the key contributions of this study are described in the following.

- We propose a continuous CNN-based emotion recognition model to improve the musical expression of performers in vocal performances.
- We re-assign different importance to all EEG physical channels through EEG channel enhancement.

The remainder of the study is organized as follows. Section 2 reviews the related works. In Section 3, we analyze the psychological changes of vocal performers based on EEG emotion recognition and then improve the musical expression. The simulation and results analysis are presented in Section 4. Section 5 gives the conclusion of the study.

II. RELATED WORKS

2.1. Vocal Performance and Psychology

The integration of vocal performance and psychology plays a role in psychological guidance for performers, allowing performers to use scientific methods to gain self-confidence and establish a positive psychological state. Carl Seashore, a prominent American psychologist, led the University of Iowa's psychological laboratory and achieved world-renowned results in the psychological test of musical talent; and wrote *The Psychology of Musical Talent* (1919) and *Psychology of Music* (1938) [26]. It is believed that the research on music and psychology has already started very early. As vocal music is a form of music, it has long been a common phenomenon to use psychological factors to guide vocal performance. The fact that music conveys feeling and controls affect is one reason why it is so universal. Music psychologists frequently investigate the connections between music and emotion. In [27], Swaminathan and Schellenberg discussed new results in three areas: emotional transmission and perception in music, emotional repercussions of music listening, and indicators of music preferences. Vocal psychology is a discipline that intersects vocal music and psychology, and belongs to the category of applied psychology. The general research direction is the psychology of vocal performance. Vocal psychology has been separated from vocal art and has developed into a distinct field of study that guides the evolution of vocal music. The goal of the study of the psychology of vocal performance, according to the author in [28], is to theoretically clarify the psychological phenomena, nature, and laws of the performers and to fully exploit the regulating effect of psychological factors on performance and even vocal art itself in order to advance the development of vocal art.

2.2. Emotion Recognition

Emotion recognition research is of great significance. Regarding HCI, emotion recognition can allow machines to understand better interactive behaviors, such as text input by users and keyboard tapping speed, and capture and analyze emotions to provide real-time feedback to users to achieve more effective communication between humans and computers [29-30]. Regarding safe driving, when faced with complex and changeable traffic conditions, he/she should stay awake and focus on the traffic conditions that may change at any time [31]. However, the psychological changes of the driver are different. Once abnormal driving is detected, a warning is issued in time to avoid dangerous driving behavior. Early emotion recognition methods relied on facial expressions, speech, and postures. For example, rhythm, timbre, and frequency spectrum features were extracted from audio and then classified using pattern recognition methods. Although the features extracted from this data is easy to identify, the behavioral signals are easy to disguise. In contrast, physiological signals are less likely to be disguised and contain more information [32]. Common physiological signals include EEG, electromyography, electrocardiograph, pupil distance, etc. Emotion recognition mainly includes two parts: feature extraction and emotion classification. Among many physiological signals, EEG is a signal that records changes in scalp potential, which can better reflect changes in people's mental states and emotions. Various studies have shown some correlation between EEG and emotion [33]. With the rapid development of EEG products and signal processing methods, EEG-based emotion recognition has become a relatively common method and has been more and more applied.

The success of vocal performance depends not only on singing skills but also on expressing personal emotions. There is little research on enhancing vocal performance by capturing psychological changes. To this end, we analyze the psychological changes of vocal performers based on EEG emotion recognition and then improve the musical expression.

III. METHODOLOGY

3.1. Frequency Patterns Decomposition

Behavioral state and in-band correlation allow EEG signals to be divided into five frequency patterns, as shown in Table 1. In emotion recognition experiments, the original signal is generally decomposed by filtering. To explore the classification impact of the single and combined frequency bands, we decompose the EEG signals into each frequency band [34]. Theta, alpha, beta, and gamma are the four frequency bands that we use to divide the original EEG signals using filters. Then, we combine the four frequency bands into a multi-channel form.

3.2. Feature Extraction of EEG Signals

The EEG patterns' frequency bands can be distinguished using differential entropy (DE). It can precisely reflect changes in the stimulus's degree of response. Therefore, DE is used to illustrate the complexity of a fixed-length EEG. In fact, differential entropy is a generalization of Shannon entropy on continuous signals [35]. Let X_i be the EEG signal, defined as follows.

$$D(x) = -\int E(x) \log[E(x)]\, dx. \tag{1}$$

Here, $E(x)$ is the continuous EEG signals' probability distribution function, and DE is defined as follows.

$$D(x) = -\iint_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] dx = \frac{1}{2}\log 2\pi e^{\sigma^2} . \tag{2}$$

However, a section of the EEG signal approximately obeys the Gaussian distribution after 2–44 Hz band-pass filtering. The π and e in equation (2) are constants. As long as the value of σ^2 is obtained, the DE of a segment of EEG signal X_i can be calculated. To reduce the difference in the physiological signals of vocal performers, the baseline signal will be combined for emotion recognition. The difference between the EEG signal's DE and the baseline signal's DE for a specific frequency band can be described as follows.

$$final\_v_m^n = trial\_v_m^n - base\_v_m^n, \tag{3}$$

where $trial\_v_m^n$ represents the DE eigenvector of segment m on frequency band n, $base\_v_m^n$ represents the DE eigenvector of baseline signal segment k on frequency band n, $final\_v_m^n$ represents the final emotional state eigenvector of segment m on frequency band n. The above three eigenvectors belong to R^i, and i represents the EEG channel used.

Physiological signals vary significantly between different vocal performers, and the same vocal performer will also have significant differences at different moments and environments [36]. There will be a period of resting-state signals for vocal performers before each performance. These resting-state signals will specifically influence the EEG signals in the subsequent stimulus. Therefore, we use the DE of the EEG signal in the emotionally evoked state minus the DE of the EEG signal in the calm state (that is,

Table 1. EEG bands.

Band	Frequency (Hz)	Normally
Theta	4-8	• Drowsiness in adults and teens • Fatigue
Alpha	8-12	• Relaxed/reflecting • Closing the eyes
Beta	12-30	• Range span: active calm → intense → stressed → mild obsessive • Active thinking, focus, anxious
Gamma	30-100	• Excited

the baseline signal). The result is constructed as a 9×9 matrix according to the 10–20 system for EEG electrode placement positions in Fig. 1, and zero is used for the remaining positions filling [37]. The two-dimensional matrices of theta, alpha, beta, and gamma four frequency bands will be obtained according to the method in Fig. 2. To perform good feature extraction. The four two-dimensional matrices are stacked in the frequency dimension to obtain multi-channel input. The four-channel data is used as the input of continuous CNN for feature extraction and classification. The EEG signals feature extraction and transformation process is shown in Fig. 2.

3.3. EEG Channel Enhancement

The consistent importance of EEG channels, features that could be more relevant to emotion recognition, and multi-channel information overload are all issues in EEG emotion recognition [38]. The performance of emotion recognition in vocal performance can be improved by enhancing the effective emotional information extracted into features and suppressing the negative impact of redundant information. The EEG channel improvement module based on the attention mechanism is used to enhance the accuracy in EEG emotion recognition.

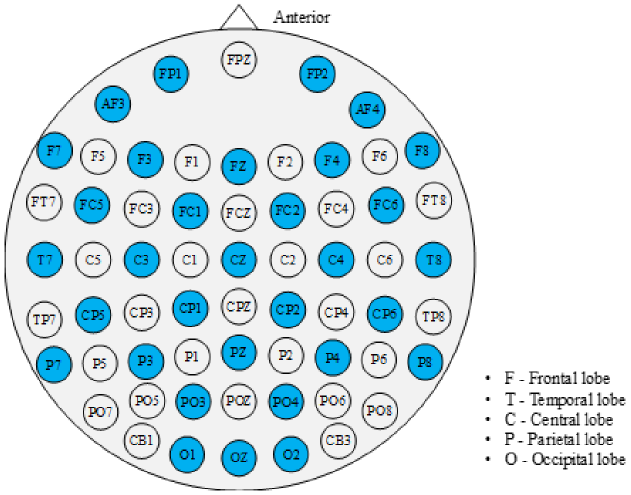

Fig. 1. 10–20 system for EEG placement.

The number of EEG channels, the temporal information, and the frequency information are all subject to channel enhancement. The EEG channel enhancement module can assign different importance to different EEG channels in emotion recognition. EEG channel enhancement mainly comprises a fully connected layer and activation function. Through EEG channel enhancement, a one-dimensional vector of the same size as the number of EEG channels can be obtained as the importance of each EEG channel. Then the importance of this one-dimensional vector is re-integrated with the three-dimensional time-frequency-like features tensor weighted and multiplied, and the enhanced features are obtained. We introduce scaling convolutional layers to extract one-dimensional time series signal features. Its input can be an EEG signal of any length. For each scaling convolutional layer, a scaling convolution kernel will perform cross-correlation calculations with the EEG signal. Each EEG channel will be assigned a scaling convolutional layer independently to extract time-frequency-like features of different EEG channel signals. After outputting the time-frequency-like features from the scaling convolutional layer, the EEG channel enhancement module is connected, where r represents the compression ratio of the dimensionality reduction process. This can reduce the amount of network computation on the one hand and increase the nonlinearity of the network on the other hand. Using two fully connected layers can enhance the nonlinear transformation capability

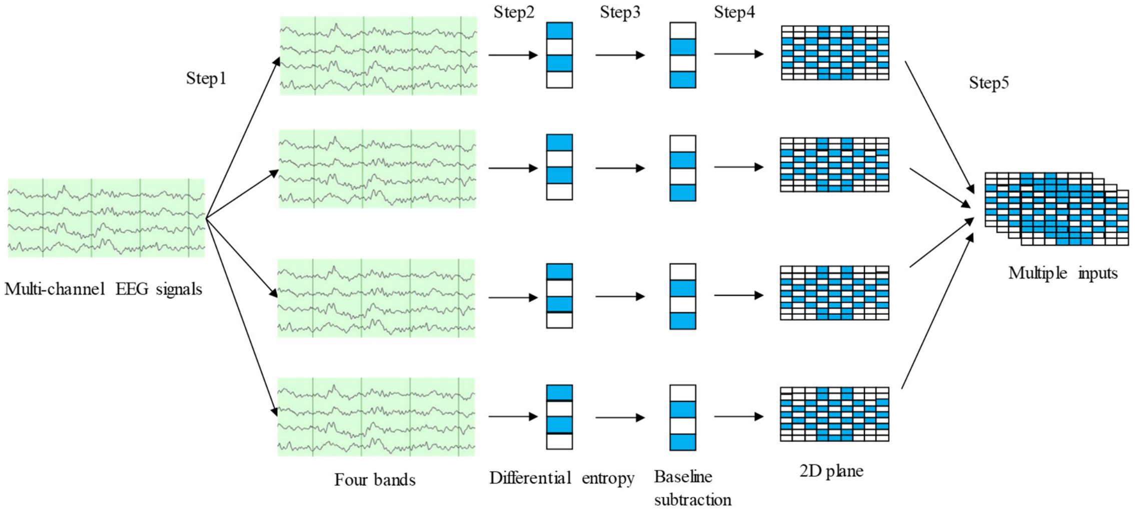

Fig. 2. Flowchart of EEG signals feature extraction and transformation.

of the network. Finally, it achieved the purpose of enhancing the importance of EEG channels related to EEG emotion recognition and inhibiting EEG channels unrelated to EEG emotion recognition.

The EEG channel enhancement module can assign importance to different EEG channels in emotion recognition. It mainly comprises a fully connected layer and activation function. Through EEG channel enhancement, a one-dimensional vector of the same size as the number of EEG channels can be obtained as the importance weight of each EEG channel. The module first extracts time-frequency features from each channel independently using scaling convolutional layers. After concatenating these features, a dimensionality reduction is performed to compress the features while increasing nonlinearity. Next, two fully connected layers further enhance the network's nonlinear transformation capability. Finally, a channel importance weight vector is outputted through a sigmoid activation, with larger values indicating higher relevance of that channel to emotion recognition. This vector is re-integrated with the input features tensor using element-wise multiplication. Thus, channels more pertinent to emotional states are enhanced while irrelevant channels are suppressed.

3.4. Continuous CNN

The convolutional layer is where most of the computation in CNN is focused, so using smaller convolution kernels and fewer convolutional layers can help accelerate the model's performance. The pooling layer follows the convolutional layer in the conventional CNN, filtering the features of the prior layer while minimizing the number of computations in the following layer. However, in emotion recognition, the features extracted by the convolution operation are limited, and the effective pooling layer is easy to lose when using the pooling layer. The pooling layer will increase the depth of the model and lead to overfitting. He et al. reduced the number of pooling layers in the image processing but did not affect the multi-channel input results. To decrease the depth of the model and the possibility of overfitting, continuous convolution is used in place of the pooling layer in the suggested CNN model.

Lastly, Fig. 3 depicts the seven-layer model developed by the continuous CNN. The convolutional model's input layer acquires the EEG data after processing it. Layers 2 to 5 are convolutional layers, and multiple convolutions are used to extract deeper psychological characteristics of vocal performers. Layer 6 is a fully connected layer with 1,024 neurons. The last layer is the output layer.

IV. SIMULATION

4.1. Setup

Since the number of EEG datasets is limited, the convolution is easy to lose the original data features. The same pattern is used for filling to ensure the same size of the input and output. The convolution kernel size is generally set to an odd number to ensure symmetrical filling. After many experiments, it is finally determined that the convolutional layer of the second to fourth layers uses a convolution kernel with a size of 3, the step size is set to 1, and the number of convolution kernels is doubled from 64 layer-by-layer. The fifth layer's convolution kernel is intended to be 1 in size. There are 64 convolution kernels since the feature dimensionality reduction is made before entering the fully connected layer. The continuous CNN model's activation function, sample size, and learning rate are scaled exponential linear unit (SELU), 128, and 10, respectively [39]. The cross-entropy loss function is combined with L2 regularization to prevent overfitting, and the Adam optimizer maximizes the objective function. The dropout is 0.5, and the L2 regularization weight coefficient is 0.5.

4.2. DEAP Dataset

In emotion recognition experiments, discrete and dimensional models are usually used to construct emotional spaces. The dimensional model believes that a two-dimensional arousal-valence coordinate system can represent emotional space. A database for emotion analysis using physiological signals (DEAP) was created by Koelstra et al. using the emotion dimension model [40]. This database is used frequently by many EEG-based emotion recognition

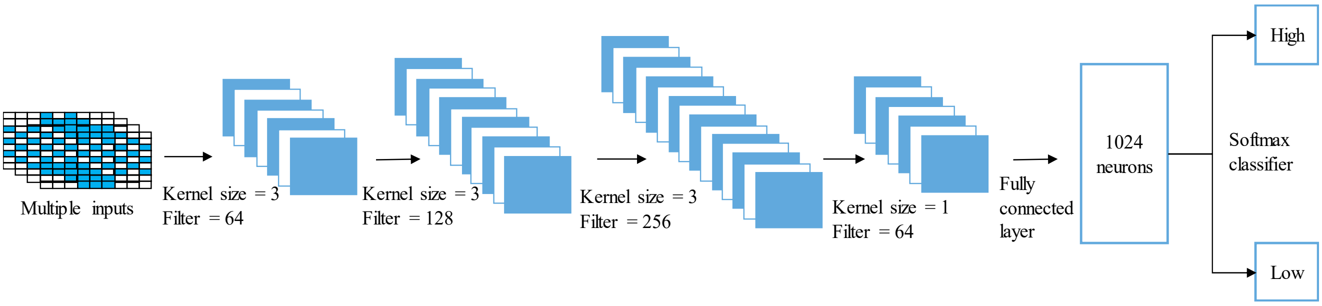

Fig. 3. Continuous CNN.

methods. Compared with movies, music videos can make people enter readiness faster and induce emotions faster, so this paper uses DEAP to verify the effectiveness of the emotion of vocal performers.

4.3. Data Preprocessing

To increase the data sample size, the EEG of vocal performers is first divided into 1s. Each participant can get 60 EEG sample data by watching a music video, and the size of the processed EEG dataset is 40 (number of channels) ×128 (segment length)×60 (number of segments). This study discusses EEG-based emotion recognition, which uses emotions to study the psychological changes of vocal performers, and then provides support for improving musical expression. We select the first 32-channel EEG signals in DEAP. The DE is used to extract the characteristics of the EEG signals in each of the four frequency bands—theta, alpha, beta, and gamma—that the Butterworth filter decomposes into.

There will be a period of resting-state signals for vocal performers before each performance. These resting-state signals will specifically influence the EEG signals in the subsequent stimulus. Therefore, we use the DE of the EEG signal in the emotionally evoked state minus the DE of the EEG signal in the calm state (that is, the baseline signal). The vocal performer's baseline signal is obtained by having them sit quietly with their eyes closed before watching the video stimulus. The EEG data from this rest period is divided into three 1s baseline signal segments. The DE features extracted from these baseline segments are then subtracted from the DE features extracted during the trials, accounting for each individual's inherent variability. This baseline subtraction allows for improved normalization across participants.

4.4. Baseline Subtraction Interference Experiment

To study the influence of baseline signal on EEG emotion recognition, this study conducts experiments on selecting the DE of baseline signal. The DE of a segment of the baseline signal, the mean of DE of the baseline signal per second, and the DE of the baseline signal in the last second are used as the participant's baseline signal DE features. The continuous CNN model is evaluated using ten-fold cross-validation. After being normalized and randomized, the processed sample data is divided into a training set and a test set in a 9:1 ratio. After ten-fold cross-validation, the average accuracy of each participant is used to determine that participant's accuracy.

As shown in Table 2, the classification accuracies of arousal and valence are 69.96% and 67.58% without combining the baseline signal. Using the deviation of all baseline signal DE features and EEG signal DE as the input of continuous CNN, the classification accuracies on arousal and valence reach 89.08% and 88.86%. The classification accuracy of emotion recognition on arousal and valence by combining the mean of baseline signal DE is 89.53% and 88.74%, and the classification accuracy on arousal and valence by combining the last second baseline signal DE is 95.59% and 95.45%. Fig. 4 shows the classification accuracy of 32 participants on arousal and valence. The results of the three experiments combined with the baseline signal are much higher than those without the baseline signal. For

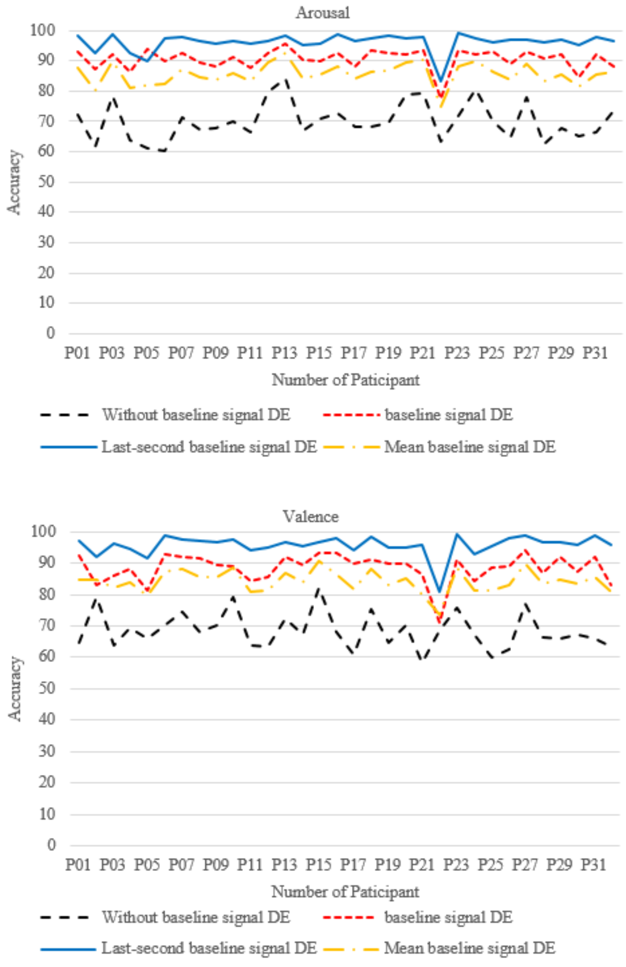

Fig. 4. Classification accuracy of 32 participants.

Table 2. Comparison of classification accuracy %.

	Without baseline signal DE	All baseline signal DE	Mean of baseline signal DE	Last second baseline signal DE
Arousal	69.96	89.08	89.53	95.59
Valence	67.58	88.86	88.74	95.45

the classification experiment, similar results were achieved when all baseline signal DEs were combined and when the baseline signal DE mean was used. The results obtained using the last-second baseline signal DE are six percent higher than the other two cases. Almost all participants' classification accuracy is higher than 90%. Relatively, the emotion recognition rate without using the baseline signal is low. From this analysis, it can be seen that subtracting baseline interference can improve the accuracy of emotion recognition and help capture the psychological changes of vocal performers and improve musical expression.

The correct number of EEG channels can help minimize computational complexity in EEG emotion recognition experiments. The preprocessed DEAP is used in this study to determine how the number of electrode channels affects recognition accuracy and then select the optimal number of EEG channels. The prefrontal, anterior hemisphere, and temporal lobe are three brain regions that control human emotions. Therefore, we selected three different electrode distributions, corresponding to the 10‒20 System for EEG: (i) 4 channels (FP1, FP2, T7, T8); (ii) 8 channels (FP1, FP2, T7, T8, CP5, CP6, P7, P8); (iii) 22 channels (FP1, FP2, AF3, AF4, F7, F3, FZ, F4, F8, FC5, FC1, FC2, FC6, T7, C3, CZ, C4, T8, CP5, CP6, P7, P8). Fig. 5 shows the classification accuracy of 32 participants on three different channels. The results demonstrate that the average classification accuracy of the 22 channels is higher than that of the 4 channels and 8 channels. The average classification accuracy of 22-channels is close to the average classification accuracy of 32 channels, so it can be inferred that FP1, FP2, AF3, AF4, F7, F3, FZ, F4, F8, FC5, FC1, FC2, FC6, T7, C3, CZ, C4, T8, CP5, CP6, P7, P8 electrode channels contain most of the emotion recognition information needed by vocal performers.

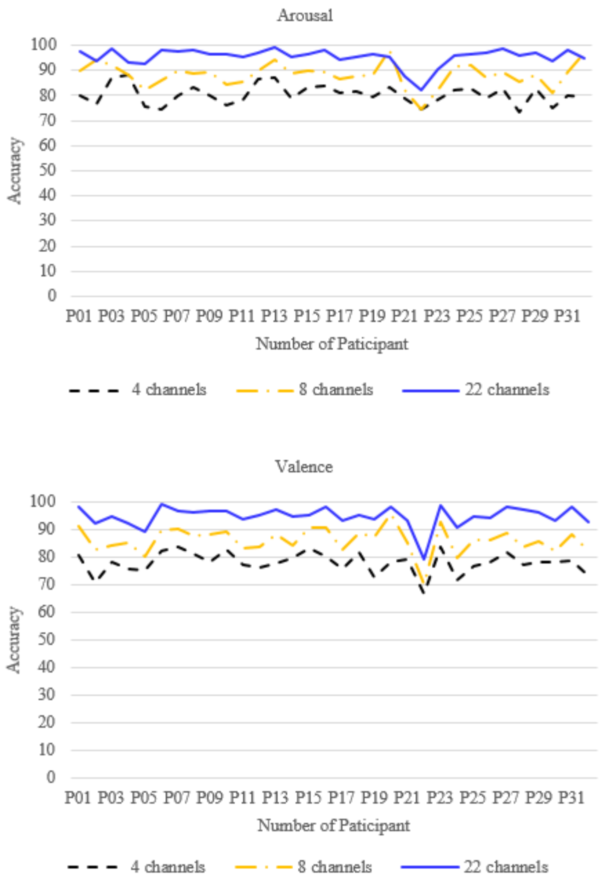

Fig. 5. Classification accuracy of different channels.

4.5. Comparison of Classification Accuracy

The continuous CNN emotion recognition model is compared with other EEG emotion recognition methods using the same dataset to confirm its effectiveness further. The EEG signal's Lempel-Ziv (LZ) complexity and wavelet detail coefficient are calculated in the literature [41]. It inputs the data into the LIBSVM classifier for classification after computing the average approximate entropy. Literature [42] combined LSTM and RNN models to learn features directly from the original data, and the average accuracy on arousal and valence was 85.56% and 85.54%. Literature [43] uses 3D CNN to classify the data of a single frequency band and its combined frequency bands. Literature [44] segments the EEG data, extracts the differential entropy features, and then forms a feature cube as the input of the deep learning model of the joint graph CNN and LSTM. The final average classification accuracy of literature [44] is 90.54% and 90.06%. The proposed method has a better effect on emotion recognition on DEAP. It can be used in vocal performance integrated with psychology, as shown in Table 3, where the average accuracy of the emotion recognition method proposed in this paper on arousal and valence is higher than that of the other four benchmarks.

By accurately modeling a vocalist's psychological state during performance using the continuous CNN model, specific insights can be provided on their emotional engagement and reactivity over time. For example, low arousal or valence classification periods may indicate moments where the performer is not fully immersed or emotionally connected. Alternatively, sustained high engagement across metrics may reveal sections that resonate most powerfully. Through EEG feedback, vocal coaches can pinpoint areas

Table 3. Comparison of classification accuracy with four benchmarks %.

	Arousal	Valence
Chen et al. [41]	74.88	82.36
Alhagry et al. [42]	85.56	85.54
Yang et al. [43]	90.42	89.54
Yin et al. [43]	90.54	90.06
Proposed method	95.36	95.31

needing improvement and provide personalized training to boost engagement and musical expressivity. For performers, real-time biofeedback allows self-monitoring to maintain an ideal psychological flow. Combining EEG recognition and vocal pedagogy can thus enhance training efficiency, motivation, and creative interpretation.

V. CONCLUSION

This paper used EEG to recognize vocal performers' psychological changes, enhancing their musical expression. We performed feature extraction and classification on EEG data and finally selected a 22-channel electrode distribution through baseline subtraction interference and channel selection. The experimental results showed that combining a rest-state baseline signal before each trial significantly improved accuracy. The average classification accuracy of the proposed method on arousal and valence was 95.36% and 95.31%, which proved that combining the baseline signal effectively improved the classification accuracy. The continuous CNN model also further extracted adequate information from the features. Simultaneously, selecting some channels that had a more significant impact on emotion recognition to replace all channels significantly reduced the data volume and reduced computational complexity. This study extracted differential entropy features, and the extracted feature types were relatively single. In future work, we will consider combining different EEG signal feature extraction methods to extract deeper features. Meanwhile, this study did not use the remaining 8-channel physiological signals in the 32-channel EEG signal. In later studies, other physiological signals will be combined with EEG signals to enhance feature extraction and emotion classification methods.

REFERENCES

[1] G. F. Welch, E. Himonides, J. Saunders, I. Papageorgi, and M. Sarazin, "Singing and social inclusion," *Frontiers in Psychology*, vol. 5, 2020.

[2] T. DeGroot, S. Valcea, and M. Hamdani, "Examining the impact of vocal attractiveness on team performance," *Current Psychology*, vol. 42, no. 17, pp. 14147-14158, 2022.

[3] D. E. Tan, F. M. Diaz, and P. Miksza, "Expressing emotion through vocal performance: Acoustic cues and the effects of a mindfulness induction," *Psychology of Music*, vol. 48, no. 4, pp. 495-512, 2020.

[4] H. Li, "Fuzzy control based stage phobia analysis system for vocal performers," *International Journal on Artificial Intelligence Tools*, vol. 30, 2022.

[5] Y. Zhang, "Cultivation and interpretation of students' psychological quality: Vocal psychological model," *Frontiers in Public Health*, vol. 10, 2022.

[6] K. Latham, B. Messing, M. Bidlack, S. Merritt, X. Zhou, and L. M. Akst, "Vocal health education and medical resources for graduate-level vocal performance students," *Journal of Voice*, vol. 31, no. 2, 2017.

[7] Q. Zhang, "Analysis of the effect of intervention on performance anxiety among students majoring in vocal music performance," *Psychiatria Danubina*, vol. 34, pp. S903-S908, 2022.

[8] V. Gynnild, "Assessing vocal performances using analytical assessment: A case study," *Music Education Research*, vol. 18, no. 2, pp. 224-238, 2016.

[9] Y. Wang, "Multimodal emotion recognition algorithm based on edge network emotion element compensation and data fusion," *Personal and Ubiquitous Computing*, vol. 23, pp. 383-392, 2019.

[10] D. Han, Y. Kong, J. Han, and G. Wang,"A survey of music emotion recognition," *Frontiers of Computer Science*, vol. 16, no. 6, p. 166335, 2022.

[11] X. Yang, Y. Dong, and J. Li, "Review of data features-based music emotion recognition methods," *Multimedia Systems,* 24, pp. 365-389, 2018.

[12] C. P. Polizzi and S. J. Lynn, "Regulating emotionality to manage adversity: A systematic review of the relation between emotion regulation and psychological resilience," *Cognitive Therapy and Research*, vol. 45, pp. 577-597, 2021.

[13] B. A. Jaso, S. E. Hudiburgh, A. S. Heller, and K. R. Timpano, "The relationship between affect intolerance, maladaptive emotion regulation, and psychological symptoms," *International Journal of Cognitive Therapy*, vol. 13, pp. 67-82, 2020.

[14] C. L. Kim and B .G. Kim, "Few-shot learning for facial expression recognition: A comprehensive survey," *Journal of Real-Time Image Processing*, vol. 20, no. 3, pp. 1-18, 2023.

[15] S. J. Park, B. G. Kim, and N. Chilamkurti, "A robust facial expression recognition algorithm based on multi-rate feature fusion scheme," *Sensors*, vol. 21, no. 7, pp. 1-26, 2021.

[16] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion recognition from multimodal physiological signals for emotion aware healthcare systems," *Journal of Medical and Biological Engineering*, vol. 40, pp. 149-157, 2020.

[17] R. Xu, J. Chen, J. Han, L. Tan, and L. Xu, "Towards emotion-sensitive learning cognitive state analysis of big data in education: deep learning-based facial expression analysis using ordinal information," *Computing*, vol. 102, pp. 765-780, 2020.

[18] Q. Liu and H. Liu, "Criminal psychological emotion recognition based on deep learning and EEG signals,"

Neural Computing & Applications, vol. 33, no. 1 pp. 433-447, 2021.

[19] D. Jeong, B. G. Kim, and S. Y. Dong, "Deep joint spatio-temporal network (DJSTN) for efficient facial expression recognition," *Sensors,* vol. 20, no. 7, p. 1936, 2020.

[20] A. Gupta, G. Siddhad, V. Pandey, P. P. Roy, and B. G. Kim, "Subject-specific cognitive workload classification using EEG-based functional connectivity and deep learning," *Sensors*, vol. 21, no. 20, p. 6710, 2021.

[21] L. Deng, X. Wang, F. Jiang, and R. Doss, "EEG-based emotion recognition via capsule network with channel-wise attention and LSTM models," *CCF Transactions on Pervasive Computing and Interaction*, vol. 3, pp. 425-435, 2021.

[22] D. Huang, S. Zhou, and D. Jiang, "Generator-based domain adaptation method with knowledge free for cross-subject EEG emotion recognition," *Cognitive Computation*, vol. 14, no. 4, pp. 1316-1327, 2022.

[23] S. Gannouni, A. Aledaily, K. Belwafi, and, H. Aboalsamh, "Emotion detection using electroencephalography signals and a zero-time windowing-based epoch estimation and relevant electrode identification," *Scientific Reports*, vol. 11, no. 1, p. 7071, 2021.

[24] R. Zatarain Cabada, H. Rodriguez Rangel, M. L. Barron Estrada, and H. M. Cardenas Lopez, "Hyperparameter optimization in CNN for learning-centered emotion recognition for intelligent tutoring systems," *Soft Computing*, vol. 24, no. 10, pp. 7593-7602, 2020.

[25] K. U. Devi and R. Gomathi, "Retraction note to: Brain tumour classification using saliency driven nonlinear diffusion and deep learning with convolutional neural networks (CNN)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, p.475, 2022.

[26] C. B. Hancock and H. E. Price, "First citation speed for articles in psychology of music," *Psychology of Music*, vol. 44, no. 6, pp. 1454-1470, 2017.

[27] S. Swaminathan and E. G. Schellenberg, "Current emotion research in music psychology," *Emotion Review*, vol. 7, no. 2, pp. 189-197, 2015.

[28] Z. Ning, "Research on the psychological course and development trend of vocal performance," in *Proceedings of the 2017 3rd International Conference on Economics, Social Science, Arts, Education and Management Engineering (ESSAEME 2017)*, 2018, vol. 119, pp. 2002-2006.

[29] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from EEG-based brain-computer interface using machine learning: A comprehensive review," *Neural Computing & Applications*, vol. 34, no. 15, pp. 12527-12557, 2022.

[30] K. Kambleand J. Sengupta, "A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals," *Multimedia Tools and Applications*, vol. 82, no. 18, pp.27269-27304, 2023.

[31] J. Izquierdo-Reyes, R. A. Ramirez-Mendoza, M. R. Bustamante-Bello, J. L. Pons-Rovira, and J. E. Gonzalez-Vargas, "Emotion recognition for semi-autonomous vehicles framework," *International Journal of Interactive Design and Manufacturing*, vol. 12, pp. 1447-1454, 2018.

[32] M. R. Elkobaisi, F. Al Machot, and H. C. Mayr, "Human emotion: A survey focusing on languages, ontologies, datasets, and systems," *SN Computer Science*, vol. 3, no. 4, p. 282, 2022.

[33] Q. Cai, G. C. Cui, and H. X. Wang, "EEG-based Emotion recognition using multiple kernel learning," *Machine Intelligence Research,* vol. 19, no. 5, pp. 472-484, 2022.

[34] S. Tiwari, S. Goel, and A. Bhardwaj, " EEG signals to digit classification using deep learning-based one-dimensional convolutional neural network, *Arabian Journal for Science and Engineering*, vol. 48, no. 8, pp. 9675-9697, 2022.

[35] S. Hwang, K. Hong, G. Son, and H. Byun, "Learning CNN features from de features for eeg-based emotion recognition," *Pattern Analysis and Applications*, vol. 23, pp. 1323-1335, 2020.

[36] A. Garg, V. Chaturvedi, A. B. Kaur, V. Varshney, and A. Parashar, "Machine learning model for mapping of music mood and human emotion based on physiological signals," *Multimedia Tools and Applications*, vol. 81, no. 4, pp. 5137-5177, 2022.

[37] Y. Kakisaka, R. Alkawadri, Z. I. Wang, R. Enatsu, J. C. Mosher, and A. S. Dubarry, et al, "Sensitivity of scalp 10−20 EEG and magnetoencephalography," *Epileptic disorders*, vol. 15, pp. 27-31, 2013.

[38] E. S. Pane, A. D. Wibawa, and M. H. Purnomo, "Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters," *Cognitive processing*, vol. 20, pp. 405-417, 2019.

[39] L. D. Duy and P. D. Hung, "Adaptive graph attention network in person re-identification," *Pattern Recognition and Image Analysis*, vol. 32, no. 2, pp. 384-392, 2022.

[40] N. Thammasan, K. Moriyama, K. I. Fukui, and M. Nomao, "Familiarity effects in EEG-based emotion recognition," *Brain Informatics*, vol. 4, pp. 39-50, 2017.

[41] T. Chen, S. Ju, F. Ren, M. Fan, and Y. Gu, "EEG emotion recognition model based on the LIBSVM classifier," *Measurement*, vol. 164, 2020.

[42] S. Alhagry, A. Aly, and A. Reda, "Emotion recognition

based on eeg using lstm recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[43] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3D input for EEG-based emotion recognition," in *Neural Information Processing (ICONIP 2018) PT VII*, 2018, vol. 11307, pp. 433-443.

[44] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Applied Soft Computing*, vol. 100, 2021.

AUTHORS

Xiaochan Li received her B.S. degree from Hunan Institute of Science and Technology in 2005, received her M.S. degree from Hunan Normal University in 2011, and received her Ph.D. degree from University of Perpetual Help System DALTA in 2023. She is currently an associate professor at Huaihua Normal College. Her main research interests include musical expression, computer application, etc.

Yi Shi received her bachelor degree from Hunan Normal University in 2002 and received her M.S. degree from Hunan University of Science and Technology in 2016. She is currently an associate professor at Huaihua Normal College. Her main research interests include musical expression, computer application, etc.

Daohua Pan was born in Harbin, Heilongjiang, P.R. China, in 1981. She received Ph.D. degree in computer science and technology from Harbin Institute of Technology in 2021. Now, she works in Department of Electronic and Information Engineering, Heilongjiang Vocational College for Nationalities. Her research interests include AI, pattern recognition, pervasive computing and wearable computer and system evaluation theory and technology