Journal of Neural Engineering

PAPER

Identifying similarities and differences in emotion recognition with EEG and eye movements among Chinese, German, and French People

To cite this article: Wei Liu *et al* 2022 *J. Neural Eng.* **19** 026012

View the article online for updates and enhancements.

Journal of Neural Engineering

RECEIVED
17 November 2021

REVISED
24 February 2022

ACCEPTED FOR PUBLICATION
10 March 2022

PUBLISHED
28 March 2022

PAPER

Identifying similarities and differences in emotion recognition with EEG and eye movements among Chinese, German, and French People

Wei Liu[1,4,5], Wei-Long Zheng[1,4,5,7], Ziyi Li[1,4,5], Si-Yuan Wu[1], Lu Gan[1] and Bao-Liang Lu[1,2,3,4,5,6,*]

[1] Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, People's Republic of China

[2] Clinical Neuroscience Center, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, 197 Ruijin 2nd Rd, Shanghai 200020, People's Republic of China

[3] RuiJin-Mihoyo Laboratory, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, 197 Ruijin 2nd Rd, Shanghai 200020, People's Republic of China

[4] The Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, People's Republic of China

[5] Brain Science and Technology Research Center, Shanghai Jiao Tong University, 800 Dongchuan Rd, Shanghai 200240, People's Republic of China

[6] Qing Yuan Research Institute, Shanghai Jiao Tong University, 800 Dongchuan Rd, Shanghai 200240, People's Republic of China

[7] Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02319, United States of America

[*] Author to whom any correspondence should be addressed.

E-mail: bllu@sjtu.edu.cn

Keywords: affective computing, cultural influence, multimodal emotion recognition, EEG, eye movement, cross-cultural emotion recognition, deep learning

Abstract

Objective. Cultures have essential influences on emotions. However, most studies on cultural influences on emotions are in the areas of psychology and neuroscience, while the existing affective models are mostly built with data from the same culture. In this paper, we identify the similarities and differences among Chinese, German, and French individuals in emotion recognition with electroencephalogram (EEG) and eye movements from an affective computing perspective. *Approach.* Three experimental settings were designed: intraculture subject dependent, intraculture subject independent, and cross-culture subject independent. EEG and eye movements are acquired simultaneously from Chinese, German, and French subjects while watching positive, neutral, and negative movie clips. The affective models for Chinese, German, and French subjects are constructed by using machine learning algorithms. A systematic analysis is performed from four aspects: affective model performance, neural patterns, complementary information from different modalities, and cross-cultural emotion recognition. *Main results.* From emotion recognition accuracies, we find that EEG and eye movements can adapt to Chinese, German, and French cultural diversities and that a cultural in-group advantage phenomenon does exist in emotion recognition with EEG. From the topomaps of EEG, we find that the γ and β bands exhibit decreasing activities for Chinese, while for German and French, θ and α bands exhibit increasing activities. From confusion matrices and attentional weights, we find that EEG and eye movements have complementary characteristics. From a cross-cultural emotion recognition perspective, we observe that German and French people share more similarities in topographical patterns and attentional weight distributions than Chinese people while the data from Chinese are a good fit for test data but not suitable for training data for the other two cultures. *Significance.* Our experimental results provide concrete evidence of the in-group advantage phenomenon, cultural influences on emotion recognition, and different neural patterns among Chinese, German, and French individuals.

1. Introduction

Emotions are universal biological human responses but interdependent with cultures, and people from different cultural backgrounds might show culture-specific variability in emotion generation and display. For example, people living in cultures with stronger social norms are more willing to regulate their negative emotions so that social harmony is not disrupted [1]. This culture-specific variability not only imposes a great challenge on affective computing [2, 3], but also influences some digital mental health treatments [4]. Therefore, cross-cultural emotion recognition is a fundamental research topic for psychology, neuroscience, computer science, and artificial intelligence, as the generalizability of emotion detection systems depends on the variability across cultures [3].

Many previous studies focused on cultural influences on emotions. Some researchers believe that emotion semantics are shaped by different social structures, beliefs, and other factors belonging to what we called 'culture,' so discrete emotion concepts such as 'anger' or 'fear' have different connotations in terms of culture [5, 6]. By analyzing 2474 spoken languages, Jackson and his colleagues demonstrated that emotion semantics vary across cultures while geographically closer cultures have more similar emotion concepts [7]. Cowen and colleagues collected 2168 music samples labeled by US and Chinese subjects, and found that people from different cultures have a higher degree of identification with discrete emotions despite variations in emotion semantics [8]. Facial expressions, which used to be considered universal across cultures [9], were questioned later [10] and were recently found to have evident cultural variations [11]. Researchers found evidence for the 'in-group advantage' phenomenon, namely, that emotion recognition is more accurate when judging emotional data from one's own cultural in-group compared to cultural out-groups [12, 13]. Cross-cultural affective neuroscience (CAN) was initiated in 2012 to investigate the influence of culture on the regulation of basic affective systems. Özkarar-Gradwohl reviewed the recent development of CAN and shared guidelines, clinical implications and ethical vision of CAN for future research [14]. CAN claims that cultural influence can be studied by observing the cultural variations from the following three aspects: (1) the level of emotional interdependency; (2) the types of reinforced or suppressed affects; and (3) the types of affects that accompany interdependent or independent self-construals.

As mentioned above, most of the existing research that examined cultural influences on emotions is in the fields of psychology and neuroscience. However, from the perspective of affective computing, few studies have been carried out, and many fundamental problems are still not fully explored. Can emotion recognition systems built with EEG and eye movements adapt to diverse cultural backgrounds? Can EEG and eye movements capture and reflect similarities and differences in aspects of emotion recognition accuracies and emotional neural patterns for subjects from different cultures? Is there an in-group advantage phenomenon for emotion recognition tasks with EEG and eye movements? What are the contributions of EEG and eye movements to emotion recognition for various cultures and experimental settings? If we want to build affective models that have good cultural generalization to accelerate the application of affective computing, what characteristics of emotion transferability might exist for different cultures, and to what should we pay attention? In this paper, to answer these questions, we build new datasets, design three different experimental settings, and adopt advanced deep learning algorithms to construct affective models.

Specifically, we extend our previous work on Chinese–German and Chinese–French cross-cultural emotion recognition [15, 16], and we comprehensively investigate the similarities and differences among Chinese, German, and French individuals on the task of recognizing positive, neutral, and negative emotions from EEG and eye movements. To fully identify the cultural similarities and influences, we carry out experiments under three different experimental settings: intraculture subject dependent (ICSD), intraculture subject independent (ICSI), and crossculture subject independent (CCSI). We compared unimodal and multimodal affective models to determine cultural similarities and influences. In addition, we systematically analyze the emotion recognition results, neural patterns, confusion matrices (CMs), and attentional weight distributions of trained affective models for three cultures. The main contributions of this paper are summarized as follows:

(a) By comparing the accuracies of 13 unimodal and multimodal emotion recognition models for native Chinese, German, and French subjects under the ICSD, ICSI, and CCSI experimental settings, we observe that a cultural in-group advantage phenomenon does exist with regard to emotion recognition from EEG data.
(b) From the topomaps of EEG, we observe that the γ and β bands exhibit obvious trends in different emotions for Chinese, while for German and French, θ and α bands exhibit common changes for different emotions.
(c) Our experimental results indicate that EEG and eye movements can adapt to Chinese, German, and French cultural diversities to achieve good emotion recognition performance and that EEG

and eye movements have complementary characteristics since multimodal methods outperform unimodal methods.

(d) By analyzing topomaps of EEG, attentional weights, and the emotion transferability chart, we conlcude that the Germans and French share more similarities within each other in neural patterns and attentional weight distributions compared with Chinese, and that the data from Chinese are a good fit for test data but not suitable for training data for the other two cultures.

The remainder of this paper is organized as follows. Section 2 introduces related work on cross-cultural emotion recognition and EEG-based emotion recognition. Section 3 describes the emotion recognition models used in this paper, including unimodal and multimodal approaches. Section 4 presents the datasets, features, and experimental settings. Section 5 describes the experimental results. Section 6 analyzes the results. Finally, conclusions and future work are presented in section 7.

2. Related work

2.1. Cross-culture emotion recognition

Cross-cultural studies on emotions have attracted interest in psychology and neuroscience for decades. Psychologists have found strong support that there are indeed cultural variabilities as well as universalities. Ekman and his colleagues found that facial expressions can be recognized across cultures at above-chance accuracy [17]. However, the level of accuracy varies, as European Americans achieve higher accuracy than Asians and Africans [9]. Additionally, people more easily understand emotions expressed by their in-group members who have the same cultural background [13]. In a study where people were asked to recall emotional events, European Americans demonstrated facial expressions more vividly than Hmong Americans, but their self-reports and heart rates were similar, suggesting that Western and Eastern cultures influence how people display their emotions [18]. Another study that investigated mixed emotions revealed that although North Americans and Japanese both feel similarly toward negative events (e.g. outperformed by other people), when they confront the same positive event (e.g. winning a competition), North Americans mostly feel good, while Japanese report feeling both happy and worried about other people's feelings at the same time [19].

Lomas proposed that positive psychology field would benefit from greater levels of cross-cultural engagement, awareness, and understanding. He created a lexicography of relevant 'untranslatable' words to discuss cross-cultural variation and the implications that such variation has for psychology [20]. Scherer and Fontaine analyzed a large-scale data set with ratings of affective features covering all components of the emotion process for 24 emotion words in 27 countries. They performed a series of hierarchical regression analysis. Their results are highly consistent with the claim that appraisal patterns determine the structure of the response components, which in turn predict central dimensions of the feeling component [21]. Many theories have proposed explaining the cultural variations, including the independent vs. interdependent self-concept [22]. These cultural factors influence people's affective valuations [23], emotion perception [24], regulation [25], and mental well-being [26].

Neuroscience complements the findings in psychology, trying to determine the fundamental mechanism in our brain that is related to cultural similarities and differences [27]. Compared with psychological studies that mainly depend on subjective self-report and behavioral observation, neuroscience uses new methodologies, such as brain imaging, EEG, or even genetic methods, to investigate this problem [28]. Many researchers have found substantial evidence of culturally specific neural patterns that differentiate emotional processes. For example, Murata and his colleagues designed an experiment for studying emotion process of different cultures. They asked Asian and European American participants to either attend or suppress expressions of emotion while exposing to either unpleasant or neutral pictures. They then compared parietal late positive potential (LPP) and found that Asians showed a significant decrease in parietal LPP while European Americans did not have such a decrease but exhibited increased activation in the frontal area [29].

Greck and colleagues adopted a task of empathy with anger to study how culture modulate brain activities. They collected fMRI data while Chinese and German subjects watching familiar angry, familiar neutral and unfamiliar neutral pictures. They found that Chinese tend to value more of the harmony and regulate their emotions accordingly, whereas German show more activation in reasoning, suggesting more emphasis on the inference of others' feelings [30]. When European Americans and Chinese individuals were asked to rate targets with excited or calm smiles, Chinese individuals showed stronger ventral striatal (VS) activity (related to the reward process in the brain) when they viewed calm smiles, whereas European Americans demonstrated stronger VS activity on excited smiles [31].

Özkarar-Gradwohl reviewed gender effects of the affective neuroscience personality scales (ANPS) in 15 countries, and the results showed that gender differences on the ANPS were variable for different classes of basic emotions. Besides, the results were consistent with gender effects reported in the Big Five personality literature, including a trend of gender differences increasing when moving from 'East' to 'West' [32]. Tompson and colleagues tested the hypothesis that the carriers of 7- or 2-repeat allele of the dopamine D4 receptor gene (DRD4) may be more likely to show culturally typical response patterns than non-carriers. They let 194 European Americans and 204 east Asians rated the frequency of actually experiencing various positive and negative emotions in a typical week, and they found a significant culture × DRD4 interaction for emotional experience, east Asian carriers reported experiencing greater emotional balance than non-carriers, while European Americans showed a stronger positivity bias [33].

Lin and colleagues studied the problem that how culture plays a role in the neural mechanisms involved in intergroup perception. They recruited European Americans and Chinese participants in an emotion perspective-taking task where they viewed images of ingroup and outgroup members while undergoing an fMRI scan. They found culture-specific patterns of neural activation in the fusiform gyrus when perceiving ingroup and outgroup members and fusiform and amygdala showed different functional connectivity for different cultures [34].

Cross-cultural emotion recognition has been a highlighted topic in affective computing in recent years. Researchers have investigated this problem mostly from the perspective of human communication systems such as facial expressions [11, 35, 36], acoustic or lexical information [37, 38], body posture [39], or multimodal analysis [40–42]. Overall, these studies presented a certain level of universality across cultures but also revealed cultural specificity, as the systems were observed to have a better performance within the same cultural group. Incorporating physiological data gives researchers a chance to explore whether emotion is influenced by cultural display rules, which reflected in overt behaviors such as facial expressions or speech, or is deeply a state that brings us physiological changes [43].

2.2. EEG-based emotion recognition

To build a good emotion recognition system, our first target is to obtain a high recognition accuracy. Traditionally, researchers have proposed various features to capture emotion characteristics and tried many machine learning methods to achieve good results. Zheng and colleagues extracted six different EEG features in five frequency bands to examine their emotion recognition performance and found stable patterns for different emotions [44]. Garía-Martínez *et al* focused on EEG nonlinear features for emotion recognition, and gave a good summary of recent work using nonlinear features [45]. With the rapid development of deep learning methods, deep learning models such as the deep belief networks [46], convolutional neural networks [47–49], long-short term memory networks [50, 51], and deep graph neural networks [52, 53] have been applied to EEG-based emotion recognition and have made significant improvements compared with traditional machine learning approaches.

EEG data are nonstationary signals, meaning that signals from different subjects might have significant variability, which causes trouble for emotion recognition systems. Zheng and Lu applied transfer component analysis and transductive parameter transfer methods to address this problem and achieved a great improvement compared with nontransfer methods [54]. Li and colleagues proposed a multisource transfer learning method that reduced the reliance on the labeled data amount by treating existing persons as sources and new persons as the target [55]. Zhao and colleagues proposed a plug-and-play domain adaptation framework where the model is adjusted with unlabeled data for calibration so that it can be applied to a new person [56].

Recently, many researchers applied generative adversarial networks (GANs) to address data augmentation problems since EEG data are challenging to collect massively. Hartmann *et al* proposed the EEG-GAN framework to generate EEG data, and they evaluated the generated data with the inception score, Frechet inception distance, and sliced Wasserstein distance. Their experimental results indicated that the proposed framework can generate naturalistic EEG samples [57]. Luo and colleagues applied conditional Wasserstein GAN (cWGAN), selective variational autoencoder, and selective cWGAN to generate new EEG features and evaluated the performance with and without data augmentation. Their results indicated the effectiveness of the proposed methods [58].

2.3. Multimodal emotion recognition

Since emotions are complex cognitive processes, multimodal signals, which capture more aspects of emotions, are better than unimodal signals. Many previous studies adopt EEG and eye movements for emotion recognition tasks because EEG can reflect emotional changes in the central neural system and eye movements reflect periphysiological changes [59]. By fusing EEG and eye movements, researchers improved the performance of emotion recognition models.

The first way to fuse multiple modalities is at the feature level, and the most common method is to concatenate features from different modalities into a new feature [60]. In addition to feature-level fusion, there are decision-level fusion strategies in which we build classifiers for each modality and fuse the decision values with some mathematical operations. Lu and colleagues applied MAX fusion, SUM fusion, and fuzzy integral fusion strategies to fuse EEG and eye movement features, and they found complementary properties between EEG and eye movement data [60]. Another way to fuse multimodal signals is to build a deep learning model to fuse them, and various network structures have been proposed. Baltrušaitis and colleagues summarized these fusion structures and classified these multimodal deep learning models into multimodal joint representation methods and multimodal coordinate representation methods [61]. Zheng and colleagues proposed EmotionMeter which is a multimodal fusion framework for emotion recognition [62]. Liu and colleagues compared both joint-representation-based models and coordinated-representation-based models on several multimodal emotion datasets [63].

3. Methods

3.1. Unimodal models

For unimodal affective models, we examined support vector machines (SVMs) with a linear kernel, k-nearest neighbor (KNN), logistic regression (LR), and a deep neural network (DNN). To reduce the cost of training time, we first evaluated traditional machine learning methods (i.e. SVM, KNN, and LR) with various EEG features to determine the best EEG feature and EEG frequency band in terms of accuracy. We also evaluated the significance of the selected EEG features with three-way analysis of variance (ANOVA). Then, we evaluated the DNN with the selected EEG features and eye movement features. The DNN model used in this paper contains three hidden layers and a three-dimensional output layer corresponding to three emotion categories.

3.2. Multimodal models

For multimodal affective models, we evaluated three traditional feature fusion approaches and two deep learning methods. For traditional fusion approaches, we examined concatenation fusion, MAX fusion, and fuzzy integral fusion [63]. For deep learning fusion methods, we examined the bimodal deep autoencoder (BDAE) and deep canonical correlation analysis with an attention mechanism (DCCA-AM).

3.2.1. Bimodal deep autoencoder

BDAE was proposed by Ngiam and colleagues [64]. In our previous work, we adopted BDAE to multimodal emotion recognition [65]. The BDAE training procedure includes encoding and decoding. In the encoding phase, we trained two restricted Boltzmann machines (RBMs) for EEG and eye movement features. These two hidden layers are concatenated together, and the concatenated layer is used as the visual layer of a new upper RBM. In the decoding stage, we unfolded the stacked RBMs to reconstruct the input features. Finally, we used a back-propagation algorithm to minimize the reconstruction error.

3.2.2. Deep canonical correlation analysis with attention mechanism

The original DCCA was proposed by Andrew and colleagues [66]. It computes representations of multiple modalities by passing them through multiple stacked layers of nonlinear transformations. In this paper, we extend the original DCCA framework by adding an attention-based fusion module. Figure 1 depicts the framework of DCCA with the attention mechanism (DCCA-AM) used in this paper.

Let $X_1 \in \mathbb{R}^{N\times d_1}$ and $X_2 \in \mathbb{R}^{N\times d_2}$ be the instance matrices for two modalities. Here, N is the number of instances, and d_1 and d_2 are the dimensions of the extracted features for these two modalities. We build two DNNs for the two modalities to transform the raw features nonlinearly as follows:

$$O_1 = f_1(X_1; W_1), \tag{1}$$

$$O_2 = f_2(X_2; W_2), \tag{2}$$

where W_1 and W_2 denote parameters for the nonlinear transformations, $O_1 \in \mathbb{R}^{N\times d}$ and $O_2 \in \mathbb{R}^{N\times d}$ are the outputs of the neural networks, and d denotes the output dimension of DCCA.

The goal of DCCA is to jointly learn the parameters W_1 and W_2 for both neural networks so that the correlation of O_1 and O_2 is as high as: possible:

$$(W_1^*, W_2^*) = \underset{W_1, W_2}{\arg\max}\; corr(f_1(X_1; W_1), F_2(X_2; W_2)). \tag{3}$$

We use the backpropagation algorithm to update W_1 and W_2. Let $\bar{O}_1 = O_1' - \frac{1}{N}O_1'1$ be the centered output matrix (similar to $\bar{O}_2$). We define $\hat{\Sigma}_{12} = \frac{1}{N-1}\bar{O}_1\bar{O}_2'$, $\hat{\Sigma}_{11} = \frac{1}{N-1}\bar{O}_1\bar{O}_1' + r_1\mathbf{I}$. Here, r_1 is a regularization constant (similar to $\hat{\Sigma}_{22}$). The total correlation of the top k components of O_1 and O_2 is the sum of the top k singular values of matrix $T = \hat{\Sigma}_{11}^{-1/2}\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1/2}$. In this paper, we take $k = d$, and the total correlation is the trace of T:

$$corr(O_1, O_2) = \left(tr(T'T)\right)^{1/2}. \tag{4}$$

The CCA loss is the negative of total correlation:

$$L_{cca} = -corr(O_1, O_2). \tag{5}$$

Finally, we calculate the gradients with the singular decomposition of: $T = UDV'$,

$$\frac{\partial corr(O_1, O_2)}{\partial O_1} = \frac{1}{N-1}(2\nabla_{11}\bar{O}_1 + \nabla_{12}\bar{O}_2), \tag{6}$$

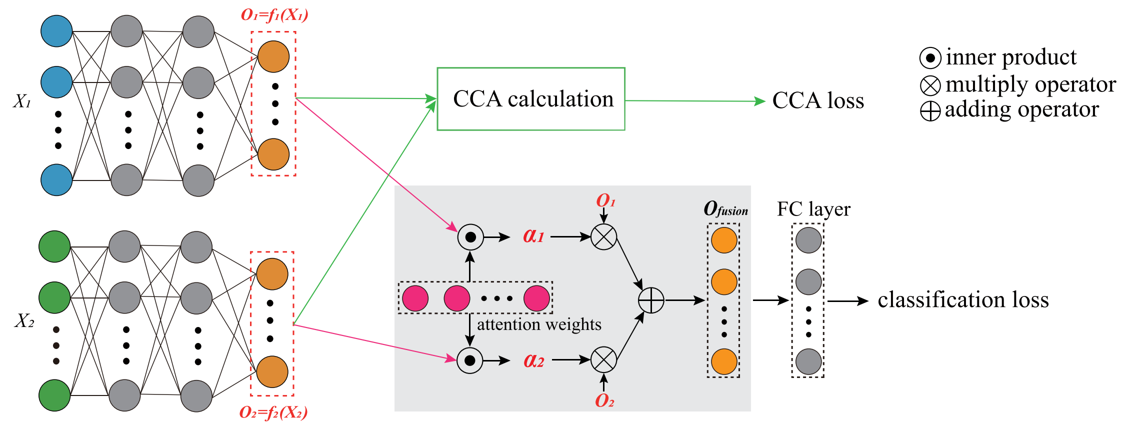

Figure 1. Framework of DCCA with attention mechanism. Different modalities are transformed by corresponding deep neural networks separately. Outputs (O_1, O_2) are regularized by CCA constraint. Attention-based fusion strategy (with gray background) is adopted to fuse O_1 and O_2, and fused features are used for emotion recognition. We update parameters to minimize negative CCA metric and classification loss simultaneously.

where

$$\nabla_{11} = -\frac{1}{2}\hat{\Sigma}_{11}^{-1/2}UDU'\hat{\Sigma}_{11}^{-1/2}, \tag{7}$$

$$\nabla_{12} = \hat{\Sigma}_{11}^{-1/2}UV'\hat{\Sigma}_{22}^{-1/2}, \tag{8}$$

and $\partial corr(O_1, O_2)/\partial O_2$ has a symmetric expression. With O_1 and O_2, we build an attentional module to fuse the transformed features.

For the attention-based fusion module (layers with gray background in figure 1), first, we initialize an attention layer with parameters W_{attn}, and then we calculate the inner product of attentional weights and outputs of different modalities and apply softmax to normalize the results to obtain attentional weights α_1 and α_2, respectively:

$$\begin{aligned}\hat{\alpha}_1 &= < O_1, W_{attn} >,\\ \hat{\alpha}_2 &= < O_2, W_{attn} >,\\ \alpha_1, \alpha_2 &= softmax(\hat{\alpha}_1, \hat{\alpha}_2),\end{aligned} \tag{9}$$

where W_{attn} is the hyperparameter to compute attentional weights. After calculating the attentional weights, we extract the fused features as follows:

$$O = \alpha_1\, O_1 + \alpha_2\, O_2. \tag{10}$$

Next, a fully connected layer is added as a classifier with which we can calculate the classification loss. Under attention-based fusion settings, all updates can be calculated with backpropagation, and we optimize the CCA loss and cross-entropy classification loss simultaneously:

$$L = \gamma_1\, L_{cca} + \gamma_2\, L_{classification}, \tag{11}$$

where L is the total loss, and γ_1 and γ_2 are hyperparameters.

4. Datasets, features, and experimental settings

4.1. Datasets

4.1.1. Chinese dataset

For Chinese, we use the multimodal version of the SEED dataset, which contains EEG and eye movements. The SEED dataset was developed by Zheng and Lu [67]. A total of 15 Chinese film clips of three emotions (positive, neutral, and negative) were chosen from a pool of materials as stimuli used in the experiments, and every participant took part in the experiment three times. In this paper, we adopted a multimodal version of SEED where EEG and eye movement data of 36 sessions from 12 native Chinese participants (6 males and 6 females; MEAN: 23.08, Std: 2.02) are included.

4.1.2. German dataset (SEED-GER)[8]

The German emotion dataset we used in this paper is the same as the dataset used in our previous work [15]. Eighteen stimuli materials of three emotion categories (i.e. positive, neutral, and negative) were selected from the dataset developed by Scharfer and colleagues [68]. All stimuli materials are in English since there is no German version of the materials. Eight native German participants (7 males, 1 female; MEAN: 22.25, Std: 1.98) took part in our experiments three times. The subjects watched these materials during the experiment, and EEG and eye movements were acquired simultaneously. Due to equipment failure and subjects quitting, we collected 20 sessions of multimodal data in which four subjects completed the three sessions, and the other four subjects completed only two sessions.

[8] The German dataset (SEED-GER) used in this paper will be freely available to the academic community as a subset of SEED.

4.1.3. French dataset (SEED-FRA)[9]

In our previous study [16], French stimuli materials were selected from the dataset developed by Scharfer and colleagues [68]. Note that the same stimuli dataset as that for the German dataset were used, but the movie clips used were different for German and French. Twenty-one film clips of three emotions (i.e. positive, neutral, and negative) were used in our experiment. We recruited 8 native French subjects (5 males, 3 females; MEAN: 22.50, Std: 2.78). Each took part in the experiments three times and their EEG and eye movement data were collected simultaneously.

All participants were undergraduate/graduate/exchange students from Shanghai Jiao Tong University. They are right-handed and have normal or corrected-to-normal vision and normal hearing without any mental diseases for the emotion experiments, and the participants got paid for the participation in the experiments.

The subjects were informed about the experiment procedure in advance, and they were instructed to sit comfortably, watch the movie clips attentively, and refrain as much as possible from overt movements. At the end of each movie clip, the subject was asked to fill in a self-assessment form immediately to record his/her true emotion (positive, negative, neutral, or others) and its intensity (ranges from 1 to 5, 1 means no intensity and 5 means strong intensity) during watching the movie clip. We only used the data where the correct emotion was elicited, and the emotion intensity was greater than or equal to 3.

The data acquisition devices for Chinese, German, and French are all the same. We use Neuroscan SynAmps[10] amplifier to acquire EEG signals. The sampling rate is 1000 Hz and there are 62 channels. The eye movement signals are acquired with SMI eye tracking glasses[11].

The main reasons for choosing native German and French subjects instead of subjects with other cultural backgrounds are that our laboratory has exchange students from Germany who can recruit native German subjects, and we have graduate students from The SJTU-ParisTech Elite Institute of Technology who can recruit native French subjects.

The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements. All subjects gave their informed consent for inclusion before they participated in the study, and they are fully informed about the experimental procedure. This study does not involve identifiable human subjects.

4.2. Feature extraction

4.2.1. EEG feature extraction

Before extracting EEG features, we preprocessed the raw EEG signals with the Curry 7 software[12]. The raw EEG signals are filtered with a 0.2–50 Hz bandpass filter, and eye blinking artifacts are removed with a threshold algorithm, and finally the processed EEG signals are down sampled from 1000 Hz to 200 Hz.

According to the existing work [44, 52], the following features are efficient for EEG-based emotion recognition: power spectral density (PSD), differential entropy (DE), differential asymmetry (DASM), rational asymmetry (RASM), and asymmetry (ASM). In this paper, we extracted these five kinds of features from raw EEG, and the feature extraction procedures were the same as those used in [67]. Under unimodal conditions, we calculated EEG features with sliding windows of 1 s. For multimodal settings, 4-s sliding windows were used to make sure there are same number of samples as eye movement features.

4.2.2. Eye movement features

The eye movements in the Chinese, French, and German multimodal datasets contain statistical information such as pupil diameter and blink duration, and computational statistics such as temporal and frequency features. We extracted the same 33 eye movement features as in our previous study [62].

We extract eye movement features with sliding windows of 4 s. The reasons for 4 s sliding window are as follows: (1) Eye movement signals change slowly during the movie watching task. For example the eye movement tracker needs several seconds to detect blinking and fixation. If we use 1 s sliding window, a lot of the features extracted are zero. When using 4 s sliding window, there is a good balance for both sample number and feature quality. (2) In previous studies, 4 s sliding window performs well for multimodal emotion recognition [62, 63]. Therefore, we used the same setting in this paper.

4.3. Expeirmental settings

In this paper, we design three different experimental settings to investigate the similarities and differences among Chinese, German, and French populations. For the intraculture setting, we construct unimodal and multimodal affective models in both subject-dependent and subject-independent conditions. For the cross-cultural setting, we build unimodal and multimodal affective models in subject-independent conditions since the subject-dependent condition does not exist. The relationships among the three different experimental settings used in this paper are illustrated in figure 2.

[9] The French dataset (SEED-FRA) used in this paper will be freely available to the academic community as a subset of SEED.

[10] https://compumedicsneuroscan.com/product/synamps-rt-64-channel-eeg-erp-ep-amplifier/.

[11] https://en.wikipedia.org/wiki/SensoMotoric_Instruments.

[12] https://www.compumedics.com.au/en/products/curry/.

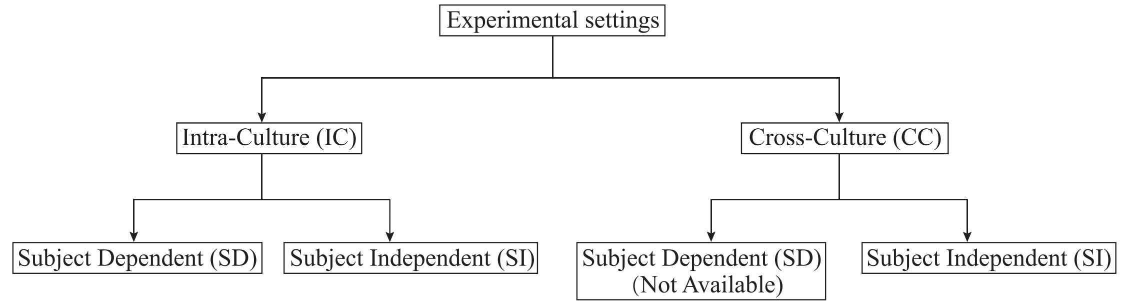

Figure 2. Three experimental settings used in this paper. Subject-dependent situation in cross cultural setting is not available because the subjects from different cultures will never be the same.

- Intraculture subject-dependent (ICSD). For the ICSD setting, we separate data from the same session into training and test samples. Specifically, for the Chinese dataset, we use the samples from the first nine clips as training samples and the samples from the remaining six clips as test samples. For the German and French datasets, the training and test ratios are 12:6 and 12:9, respectively.
- Intraculture subject-independent (ICSI). For the ICSI setting, we use a leave-one-subject-out cross-validation (CV) scheme to evaluate the performance of emotion recognition models. For every subject from the datasets, we use the data from his/her three sessions (or two sessions for the German dataset) as testing samples and data of the other subjects as training samples.
- Cross-culture subject-independent (CCSI). For the CCSI setting, we use samples from one culture as training data and samples of the other two cultures as test data.

4.4. Parameter tuning

For the KNN classifier, we tuned the number of neighbors n in the range of $[3,10]$ to find the best hyperparameter. For the LR model, we use the default function provided by the scikit-learn module. For the SVM classifier, we adopted the function in the scikit-learn module with a linear kernel, and we tuned the parameter C with grid searching from the sets of $[2^{-10}, 2^{-9}, \cdots, 2^{10}]$ and $[0.1, 20]$ with a step size of 0.5 for the large-step and small-step situations, respectively.

For the DNN model used in this paper, there are three hidden layers with 128, 64, and 32 hidden units. The output layer has three units corresponding to three emotions. The nonlinear activation functions used are ReLU or LeakyReLU according to different cultures. In addition, we added a dropout layer and a batchnorm layer for Germany. The optimization algorithm was RMSProp, and we set the epoch number to 15 000. The best learning rate was searched from 0.00 001 to 0.5 for three cultures.

For the BDAE method, training epochs and hidden units for all RBMs were searched from sets $[1000, 700, 500, 300, 200, 100]$ and $[700, 500, 200, 170, 150, 130, 110, 90, 70, 50]$, respectively. We used the RMSProp optimizer, and the learning rate was set to 0.0001. Since the BDAE outputs the fused samples, we use the SVM classifier for emotion recognition tasks.

For the DCCA-AM method, the hidden units for two nonlinear transform networks f_1 and f_2 were randomly searched from ranges $[100, 200]$ and $[20, 50]$, respectively. The hyperparameters O_1 and O_2 was set to 12, $\gamma_1 = 0.1$ and $\gamma_2 = 1.0$ according to our previous paper [63]. We also used the RMSProp optimizer and set the epoch number to 100 and the learning rate to 0.0001.

The source code of this paper can be downloaded at the following link: https://github.com/csliuwei/CrossCultureCode.

5. Experimental results

5.1. Intraculture subject dependent (ICSD) results

In section 5.1.1, we first train KNN, LR, and SVM with various EEG features and eye movement features to determine the best features. In section 5.1.2, we compare DE features and eye movement features for unimodal models trained by both traditional machine learning and deep learning algorithms. In section 5.1.3, we report the performance of five multimodal fusion strategies.

For the ICSD setting, we separate samples from the same session of one subject into training set and test set. For Chinese, the subject watched 15 film clips during one session and we use the samples from the first 9 trials as training data and the samples from the last 6 trials as test data. For German, the subject watched 18 movie clips during one session and samples from the first 12 trials are as training set and samples from the last 6 trials are as test set. For French, there are 21 trials during one session, and the ratio for training data and test data are 12:9.

5.1.1. Feature and classifier selection

We first trained KNN, LR, and SVM classifiers with five EEG features (DASM, RASM, ASM, PSD, and DE) and eye movement features for Chinese, German,

Table 1. Comparison of average accuracy (%) and standard deviation (%) of three classifiers with five different features in ICSD setting. The best results are in bold.

		KNN	LR	SVM
		Acc, Std	Acc, Std	Acc, Std
Chinese	DASM	62.63, 15.69	71.85, 14.63	74.99, 13.62
	RASM	62.62, 15.56	72.58, 14.91	74.92, 13.68
	ASM	63.05, 15.66	71.85, 14.63	74.50, 13.33
	PSD	67.05, 15.15	74.56, 12.70	77.96, 11.74
	DE	72.03, 11.85	80.51, **10.13**	**83.44**, 11.10
German	DASM	48.34, 16.79	57.53, 17.14	60.44, 16.66
	RASM	49.54, 17.06	57.63, 16.42	60.47, 16.17
	ASM	49.18, 16.06	57.96, 16.42	61.08, 16.17
	PSD	51.84, **14.37**	61.00, 18.76	59.46, 17.56
	DE	53.97, 17.71	62.74, 19.34	**65.47**, 16.93
French	DASM	44.03, 11.39	49.97, 12.28	55.62, 14.32
	RASM	44.23, 12.28	49.85, 12.37	55.83, 14.89
	ASM	43.89, 11.84	50.02, 12.72	57.79, 14.86
	PSD	45.23, **9.67**	54.60, 10.23	62.13, 12.10
	DE	46.09, 11.98	56.98, 12.78	**64.84**, 13.64

Table 2. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of the SVM classifiers with DE features of five individual frequency bands and their direct concatenation (Total) under ICSD setting. The best results are in bold.

		Delta	Theta	Alpha	Beta	Gamma	Total
Chinese	Acc	63.27	67.66	68.56	77.63	76.63	**83.44**
	Std	11.82	13.50	13.35	13.54	14.36	**11.10**
German	Acc	46.61	53.99	59.36	63.56	62.25	**65.47**
	Std	**10.77**	16.11	14.63	19.62	18.36	16.93
French	Acc	41.93	52.43	53.74	61.08	56.37	**64.84**
	Std	15.84	11.73	**10.99**	12.51	15.10	13.64

Table 3. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of different classifiers with DE features of ICSD setting. The best results are in bold.

	Chinese		German		French	
	Acc	Std	Acc	Std	Acc	Std
KNN	72.03	11.85	53.97	17.71	46.09	11.98
SVM	83.44	11.10	65.47	16.93	64.84	13.64
LR	80.53	10.13	62.74	19.34	56.98	12.78
DNN	**86.53**	**9.51**	**70.87**	**9.54**	**67.52**	**10.69**

and French. Table 1 lists the emotion recognition results, and we observe that

- The DE features have the best performance for all three cultures and the best classification accuracies for Chinese, German and French cultures are 83.44%, 65.47%, and 64.84%, respectively.
- SVM has the best performance under all conditions for three cultures.

In general, it is clear that the DE features outperform the other four EEG features and that SVM outperforms the other two classifiers for all three cultures.

We performed a three-way ANOVA with cultures (three levels), classifiers (three levels), and features (five levels) as factors. There were significant main effects of features ($p < 0.001$), classifiers ($p < 0.001$), and cultures ($p < 0.001$), and there were no significant interactions. The main effect of features is significant which means that the DE features have better performance than other features in general. The main effect of classifiers is significant indicating that the SVM classifier has better recognition accuracies than the other classifiers.

In addition, we investigated how different frequency bands affect the emotion recognition results. We trained the SVM classifiers by using the DE features of the δ band, θ band, α band, β band, γ band, and the direct concatenation of all five bands, namely, X_{total}:

$$X_{total} = [X_\delta, X_\theta, X_\alpha, X_\beta, X_\gamma], \tag{12}$$

where $X_\delta, X_\theta, X_\alpha, X_\beta$, and X_γ represent DE features of individual bands.

The emotion recognition performance of individual frequency bands is given in table 2. Two conclusions can be drawn from table 2: (1) For all three cultures, high-frequency bands (i.e. β and γ bands) have better performance than those of lower-frequency bands, and this finding is consistent with the existing work [67]. (2) For all three cultures, the best results are achieved with X_{total}, suggesting that features from lower-frequency bands contain emotional information that complements higher-frequency bands.

We performed a two-way ANOVA with cultures (three levels) and bands (six levels) as factors. There were significant main effects of bands ($p < 0.001$) and cultures ($p < 0.001$), and the interaction effect of bands × cultures is not significant ($p = 0.901$). The main effect of bands indicated that total bands of the DE features performed better than the other features.

The experimental results of tables 1 and 2 indicate that the DE feature is the best feature among these five features, and the SVM classifier performs best among the three traditional classifiers. In the following sections, we adopt the DE features to build unimodal and multimodal affective models by using deep learning algorithms.

5.1.2. Performance of unimodal models

In addition to shallow models, we trained a five-layer DNN with the DE features. The experimental results are listed in table 3. The DNN classifier outperforms the SVM classifier for all three cultures with emotion recognition accuracies of 86.53% for Chinese, 70.87% for German, and 67.52% for French.

For table 3, we carried out two-way ANOVA with cultures (three levels) and classifiers (four levels) as factors. The main effects of cultures ($p < 0.001$) and classifiers ($p < 0.001$) were significant, and there was no significant interaction effect ($p = 0.703$).

Table 4. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of unimodal classifiers with eye movement features in ICSD setting. The best results are in bold.

	Chinese		German		French	
	Acc	Std	Acc	Std	Acc	Std
KNN	65.35	21.52	64.80	30.09	45.06	17.49
SVM	75.49	17.00	78.72	21.67	51.26	15.26
LR	73.19	18.47	75.15	25.73	46.60	**11.92**
DNN	**77.45**	**15.88**	**79.87**	**18.22**	**64.52**	15.61

Table 5. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of different multimodal fusion methods in ICSD setting. The best results are in bold.

		Concat	MAX	Fuzzy	BDAE	DCCA-AM
Chinese	Acc	82.43	80.81	84.22	90.58	**92.79**
	Std	13.76	13.78	12.08	10.26	**8.21**
German	Acc	74.54	78.10	83.45	88.05	**88.63**
	Std	20.48	16.04	16.27	14.94	**10.87**
French	Acc	69.07	62.29	68.31	80.32	**80.71**
	Std	17.32	16.73	17.53	**11.04**	13.09

For eye movement features, we also evaluated the performance of different classifiers. The experimental results are given in table 4. Similar to the performance of EEG-based emotion recognition, DNN performs best among four different classifiers for all three cultures. Specifically, the best recognition accuracies for Chinese, German, and French countries are 77.45%, 79.87%, and 64.52%, respectively. We then performed a two-way ANOVA test with cultures (three levels) and classifiers (four levels) as factors. The main effects of cultures ($p < 0.001$) and classifiers ($p < 0.001$) were significant, and there was no significant interaction effect ($p = 0.513$).

From tables 3 and 4, the EEG modality is better at emotion recognition than eye movement modality for Chinese and French individuals, but eye movement features outperform EEG features for German individuals.

5.1.3. Performance of multimodal models

We evaluated five multimodal fusion strategies: concatenation fusion, MAX fusion, fuzzy integral fusion, the BDAE method, and DCCA with an attention mechanism. From table 5, DCCA-AM had the best performance among these five strategies with 92.79%, 88.63%, and 80.71% recognition accuracies for Chinese, German, and French individuals, respectively. We then performed a two-way ANOVA test with cultures (three levels) and classifiers (five levels) as factors. The main effects of cultures ($p < 0.001$) and classifiers ($p < 0.001$) were significant, and there was no significant interaction effect ($p = 0.611$).

From table 5, we see that the deep-learning-based multimodal fusion methods perform better than both unimodal models and traditional fusion methods. Therefore, we conclude that multimodal signals improve emotion recognition performance and that deep learning models capture multimodal information more effectively than traditional fusion strategies.

5.2. Intraculture subject independent (ICSI) results

In the ICSD setting, both training data and test data are from the same subject. However, we also want to investigate the performance of emotion recognition models trained and tested with different subjects.

In the ICSI setting, we used a leave-one-subject-out cross-validation scheme where training data and test data are from different subjects: samples from one subject are used as test data, and samples from the other subjects are used as training data.

5.2.1. Performance of unimodal models

We evaluate the emotion recognition performance of KNN, SVM, LR, and DNN with DE features and eye movement features. The recognition accuracies are listed in table 6. For EEG features, DNN performs best among all four classifiers, and the recognition accuracies for Chinese, German, and French classifiers are 82.81%, 65.87%, and 64.18%, respectively. For eye movement features, DNN also has the best performance and achieves 80.26%, 84.28%, and 79.85% accuracies. We performed a three-way ANOVA test with cultures (three levels), classifiers (four levels), and modalities (two levels) as factors. The main effects of cultures ($p < 0.001$), classifiers ($p < 0.001$), and modalities ($p < 0.001$) are all significant. However, the main effects of modalities and cultures were qualified by a significant interaction effect of modalities × cultures ($p < 0.001$). In addition, there were no other significant interaction effects.

According to table 6, we find that eye movements have a better emotion transferability than EEG for German and French individuals and have comparable performance for Chinese individuals. Here, we use the term 'emotion transferability' to represent the emotion recognition ability of an emotion model where the training subject and test subject are different.

5.2.2. Performance of multimodal models

For multimodal fusion evaluation, we adopt three traditional fusion strategies (concatenation, MAX, and fuzzy integral) and two deep-learning-based fusion strategies (BDAE and DCCA-AM). As we can see from table 7, DCCA-AM has the best performance for Chinese (84.04%), French (79.57%) and German (82.07%). We performed a two-way ANOVA test with cultures (three levels) and classifiers (five levels) as factors. The main effects of cultures ($p < 0.001$) and classifiers ($p < 0.001$) are significant. However, these main effects were qualified by a significant interaction effect of classifiers × cultures ($p < 0.001$). In addition, we find that the advantages of deep-learning-based models over traditional

Table 6. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of DE and eye movement features with different classifiers in ICSI setting. The best results are in bold.

		KNN		SVM		LR		DNN	
		EEG	Eye	EEG	Eye	EEG	Eye	EEG	Eye
Chinese	Acc	54.09	52.10	72.63	69.94	68.36	60.00	**82.81**	80.26
	Std	8.71	14.90	10.50	16.15	11.71	16.59	**7.52**	10.15
German	Acc	40.94	56.11	55.64	77.91	50.39	64.30	65.87	**84.28**
	Std	7.42	12.94	12.17	10.69	10.88	12.96	10.05	**6.74**
French	Acc	37.21	49.18	50.10	71.78	47.18	63.71	64.18	**79.85**
	Std	**6.76**	9.77	10.29	8.33	12.20	10.74	8.56	7.32

Table 7. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of different multimodal fusion methods in ICSI setting. The best results are in bold.

		Concat	MAX	Fuzzy	BDAE	DCCA-AM
Chinese	Acc	83.00	78.43	80.92	83.89	**84.04**
	Std	8.69	10.95	8.32	8.79	**7.35**
German	Acc	77.30	66.45	74.26	81.90	**82.07**
	Std	10.43	12.54	8.45	8.08	**7.81**
French	Acc	71.89	67.06	73.04	71.27	**79.57**
	Std	**7.50**	10.64	9.59	8.56	7.80

models between tables 7 and 6 are not as apparent as those between tables 5 and 3. This might be bacause the subject-independent scheme is more complex than the subject-dependent scheme, which limits the improvements of multimodal deep learning models.

5.3. Cross-culture subject independent (CCSI) results

In the ICSD and ICSI settings, we examined emotion recognition performance trained and tested with both the same and different subjects from the same culture. In the CCSI setting, we evaluated emotion recognition model performance with training samples and test samples from different cultures. The cross-culture scheme is always subject-independent since the subjects in the training set and test set are from different cultures (i.e. subjects in the training set and test set will never be the same).

In the CCSI setting, we used samples from one culture as training data and samples from one of the other two cultures as test data, and there are 6 situations in total, namely Chinese as training and German as test, Chinese as training and French as test, German as training and Chinese as test, German as training and French as test, French as training and Chinese as test, and French as training and German as test.

5.3.1. Performance of unimodal models

Classifiers used in this section are KNN, SVM, LR, and DNN. Table 8 presents the recognition accuracies and standard deviations, and we performed statistical significant tests on the results of these classifiers. When using Chinese as training data, we find that DNN achieves the best results for both German (55.32% for EEG and 62.44% for eye movements) and French (which are 57.28% for EEG and 62.38% for eye movements). When German is the training set, we achieve the best results with DNN for both Chinese (64.34% for EEG and 56.69% for eye movements) and French (58.93% for EEG and 67.12% for eye movements). When French is used as training data, DNN again has the best performance: for Chinese, the EEG and eye movement recognition accuracies are 66.19% and 62.99%, respectively; for German, the recognition accuracies for EEG and eye movements are 60.10% and 65.39%, respectively.

We performed a three-way ANOVA test with experimental settings (six levels), classifiers (four levels), and modalities (two levels) as factors. The main effects of classifiers ($p < 0.001$), settings ($p < 0.001$), and modalities ($p < 0.001$) are all significant. However, these main effects were qualified by significant interaction effects of classifiers × settings ($p = 0.002$), classifiers × modalities ($p = 0.016$), settings × modalities ($p < 0.001$), and classifiers × settings × modalities ($p = 0.005$).

5.3.2. Performance of multimodal models

Table 9 shows the emotion recognition performance in terms of accuracies and standard deviations. It is obvious that the deep-learning-based fusion methods perform better than traditional fusion strategies that is consistent with our previous findings that the deep learning models fuse multimodal signals better than traditional models. Specifically, when Chinese is used as training data, French is best predicted by DCCA-AM with 73.84% recognition accuracy, and German is also best recognized by DCCA-AM with 73.63%. When we use German as training samples, the DCCA-AM method outperforms BDAE and traditional fusion methods: for Chinese test samples, we obtain 84.28% recognition accuracy, and for the French test set, the recognition accuracy is 77.06%. When we use French data as the training set, DCCA-AM achieves the best test recognition accuracies for Chinese (82.09%) and German (77.92%) cultures. We then carried out a two-way ANOVA test with experimental settings (six levels) and classifiers (five levels) as factors. The main effects of settings ($p < 0.001$) and classifiers ($p < 0.001$) are significant. However, these main effects were qualified by

Table 8. Performance of mean accuracies (Acc (%)) and standard deviations (Std (%)) of DE and eye movement features in CCSI setting. The best results are in bold.

			KNN		SVM		LR		DNN	
Training	Test		EEG	Eye	EEG	Eye	EEG	Eye	EEG	Eye
Chinese	French	Acc	38.44	43.63	46.81	57.60	45.89	50.58	57.28	**62.38**
		Std	7.28	11.00	7.66	9.21	7.90	9.93	**7.12**	9.35
	German	Acc	35.82	38.21	47.88	51.76	45.33	53.81	55.32	**62.44**
		Std	**6.20**	10.62	8.87	10.40	10.88	12.07	7.32	9.79
German	Chinese	Acc	40.02	37.49	47.39	51.38	41.99	43.48	**64.34**	56.69
		Std	**7.43**	13.69	10.13	12.73	10.70	12.26	9.85	11.45
	French	Acc	38.97	44.91	47.69	65.41	46.70	60.31	58.93	**67.12**
		Std	**6.38**	13.68	11.08	8.09	11.11	8.82	8.25	7.69
French	Chinese	Acc	34.94	46.35	52.30	51.31	42.80	46.29	**66.19**	62.99
		Std	**7.39**	15.21	10.96	15.77	12.48	17.04	7.44	15.47
	German	Acc	34.70	43.58	48.88	59.74	47.53	57.26	60.10	**65.39**
		Std	**6.30**	13.34	8.90	7.63	9.47	9.64	7.86	9.37

Table 9. Comparison of mean accuracies (Acc (%)) and standard deviations (Std (%)) of different multimodal fusion methods in CCSI setting. The best results are in bold.

Training	Testing		Concat	MAX	Fuzzy	BDAE	DCCA-AM
Chinese	French	Acc	63.91	60.19	62.47	71.58	**73.84**
		Std	10.63	9.11	9.47	8.57	**8.22**
	German	Acc	58.22	54.10	59.38	69.50	**73.63**
		Std	12.61	8.65	10.17	**6.87**	10.75
German	Chinese	Acc	52.57	50.78	55.70	69.15	**84.28**
		Std	13.50	15.12	13.17	8.39	**8.19**
	French	Acc	63.62	61.27	65.28	74.42	**77.06**
		Std	13.28	11.19	9.67	**6.29**	6.45
French	Chinese	Acc	59.75	57.21	64.14	79.26	**82.09**
		Std	14.22	15.95	14.27	8.59	**8.41**
	German	Acc	62.51	56.80	59.91	75.37	**77.92**
		Std	9.66	11.15	11.19	**6.77**	8.76

a significant interaction effect of settings × classifiers ($p < 0.001$).

From the emotion recognition results in unimodal and multimodal situations under the ICSD, ICSI, and CCSI settings, we draw two conclusions:

- EEG and eye movements can be applied to emotion recognition tasks of different cultural backgrounds since all experimental results are higher than random guess results.
- EEG and eye movements have complementary information leading to a better description of emotion since the results from multimodal fusion settings are higher than the results from unimodal settings.

5.4. In-group advantage

In-group advantage is a phenomenon in which emotion recognition is more accurate when judging emotional data from one's own cultural in-group compared to the cultural out-group. In our settings, the ICSD and ICSI settings can be seen as cultural in-groups and the CCSI setting is cultural out-groups. We plot the average accuracies of all 13 classifiers (4 unimodal EEG classifiers, 4 unimodal eye movement classifiers, and 5 multimodal classifiers) to compare the average performance in the ICSD, ICSI, and CCSI settings, and the results are shown in figures 3(a)–(c). Besides, we also depicted the average performace of Chinese, German, and French subjects in in-group and out-group settings in figure 3(d).

It is clear that the average performance of in-group settings is higher than the average performance of the out-group settings, which is consistent with the 'in-group advantage' phenomenon. We adopted a two-way ANOVA test with cultures (three levels) and experimental settings (two levels) as factors. There was a significant main effect for experimental settings ($p < 0.001$). Overall, in-group settings achieved higher emotion recognition accuracies than out-group settings. Additionally, there was a significant main effect of cultures ($p = 0.049$). However, these main effects were qualified by a significant culture × settings interaction ($p = 0.002$). The in-group settings brought significant improvements for Chinese ($p < 0.001$) and German ($p < 0.001$) subjects; the improvement was not significant for French people ($p = 0.149$) though we observed a higher performance in in-group settings. Besides, Chinese, German, and French subjects have

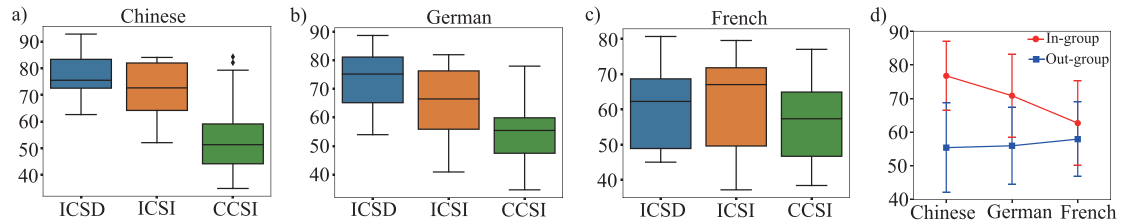

Figure 3. In-group advantage analysis. Subfigures (a)–(c) show average performance of various classifiers in ICSD, ICSI, and CCSI settings for Chinese, German and French people, respectively. Subfigure (d) shows average emotion recognition accuracies for Chinese, German, and French people in in-group and out-group settings.

comparable emotion recognition accuracies in out-group settings ($p = 0.723$). However, the accuracies changed significantly for different cultures in in-group settings ($p < 0.001$).

For Chinese and German populations, the average performance of the ICSD setting has the best performance among the three settings, and the average performance of the CCSI setting is the worst. Intuitively, this result is consistent with our experiences: from ICSD to CCSI, the divergences between training data and test data become increasingly prominent, leading to a decrease in average performance.

For French, a different phenomenon from both Chinese and German cultures is that the ICSI setting achieves better average performance than the ICSD setting. This phenomenon might be caused by a compact distribution of French samples from different subjects, leading to a bigger training set for the ICSI settings than the ICSD setting resulting to a better average recognition performance.

6. Result analysis and discussion

6.1. Summary of findings

Neural patterns for Chinese, German and French people indicate there are different responses to positive, negative, and neutral emotions. For Chinese people, γ and β bands show decreasing activities for positive, neutral, and negative emotions, while for German and French individuals, θ and α bands share increasing activities for positive, neutral, and negative emotions.

For ICSD and ICSI settings, we show that DCCA-AM model fuses EEG and eye movement features effectively and improves recognition accuracies for individual emotions. In addition, from the distribution of attentional weights, we observe that German and French have similar attentional distribution which is different from that of Chinese subjects.

For the CCSI setting, our main finding is that the data from Chinese are a good fit for test data but not suitable for training data for the other two cultures. This finding might be helpful in building an affective model with good cultural generalization.

6.2. Neural patterns analysis

As depicted in figure 4, we calculate the average DE features of five frequency bands for each emotion category and culture to investigate the neural patterns for different emotions and different cultures.

For Chinese, the γ band has the most apparent trend: the areas of high energy decrease in the order of positive, neutral, and negative emotions. Specifically, the positive emotion has the most prominent high-energy regions, containing temporal lobes and some locations of the prefrontal lobe; the neutral emotion has a smaller high-activation area located in the prefrontal lobe; and the negative emotion has the smallest high-energy distribution. For the β band, the high-energy areas are concentrated in the prefrontal areas for positive emotions, the prefrontal and occipital areas for neutral emotions, and the prefrontal midline areas for negative emotions. The θ and α bands have high energy in nearly all the brain for all three emotions. For the δ band, negative emotions have the lowest activation compared with positive and neutral emotions.

German and French have the most obvious trends appearing in the θ and α bands, where positive emotion has the lowest activation and negative emotion has the highest activation. Another common characteristic for German and French is that for the β band, the activation is low for positive and neutral emotions, while the activations have a sudden boost for negative emotions. Like Chinese, the average energies for the θ and α bands are higher than those for the other frequency bands.

From the perspective of cross-cultural emotion recognition, for positive and neutral emotions, German has the lowest activation level across all frequency bands among the three cultures, the activation level of French follows, and Chinese has the most potent activation levels. For negative emotion, Chinese has the lowest energy across five frequency bands, and French has the highest energy.

6.3. Intraculture subject dependent analysis

We compared the CMs of EEG and eye movement features with the SVM classifier, concatenation fusion method, and DCCA-AM method for Chinese,

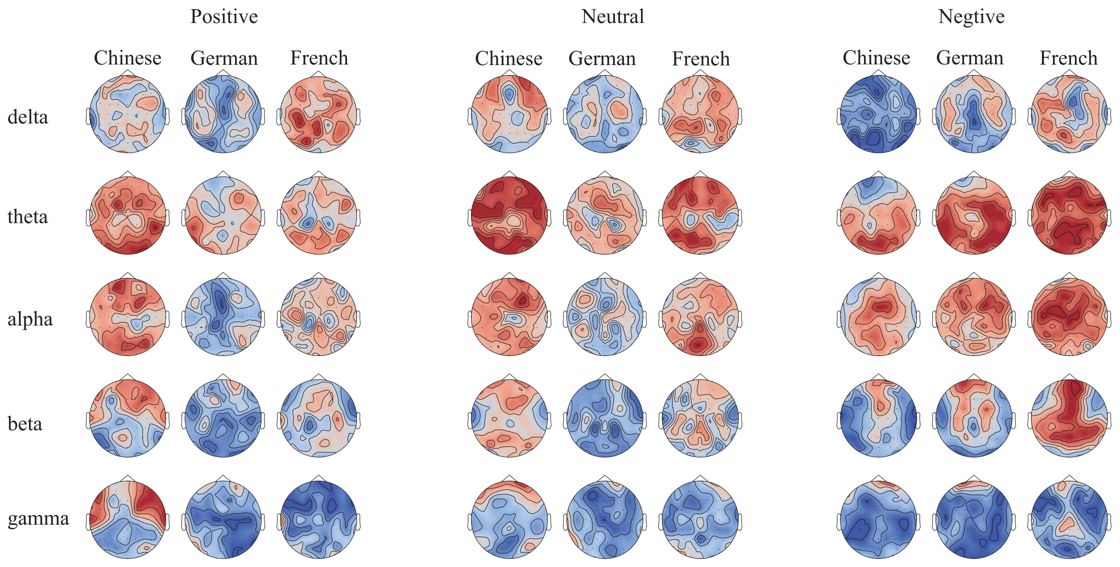

Figure 4. The average neural patterns of the DE features of five frequency bands for different emotions and cultures. Blue colors indicate low activation/energy levels, and red colors indicate high activation/energy levels.

German, and French individuals. We calculated 12 CMs (three cultures with four CMs for each culture) in total. We present the CMs as a nested pie plot in figure 5(a). In figure 5(a), twelve CMs are depicted as twelve circles, where the outer four circles, middle four circles, and inner four circles represent Chinese, German, and French, respectively. For each culture, the four circles represent EEG features, eye movement features, concatenation fusion, and DCCA-AM from outer to inner, respectively. True test emotion categories are shown in three fan-shaped sectors, and the predicted emotions are in different colors.

For Chinese individuals, EEG features have better performance than eye movement features for positive, neutral, and negative emotions. The concatenation fusion strategy improves positive emotion recognition performance, achieves similar results to that of EEG features for neutral emotions, and performs worse than both EEG features and eye movement features for negative emotions. In contrast, DCCA-AM has the best performance on three individual emotion recognition tasks, indicating that the DCCA-AM method best fuses multimodal features.

For Germans, eye movements generally perform better than EEG for individual emotion recognition, and the concatenation method might cause a performance decrease (for neutral and negative emotions) compared with unimodal situations. However, similar to the Chinese, the DCCA-AM method outperforms both EEG and eye movement unimodal situations.

For the French, EEG features have better performance than eye movement features for positive and negative emotions, while eye movement features perform better than EEG features for neutral emotion. In addition, DCCA-AM has the best performance for the three emotions.

In addition, we find that for all three cultures, positive and neutral emotions are more likely to be misclassified into negative emotions. Negative emotions are more likely to be misclassified into neutral emotions for Chinese and German individuals. For French individuals, negative emotions are inclined to be misclassified into positive emotions.

The DCCA-AM method can fuse EEG and eye movement features automatically with an attention mechanism. We examine the attentional weight distributions as shown in Figure 5(c), which can reflect the importance of different modalities. EEG features have higher weights than eye movement features for all three cultures, indicating that the EEG features contribute more than eye movement features in emotion recognition tasks. From figures 5(a) and (c), for individual emotion in Chinese, EEG features always have a better performance than eye movement features consistent with the attentional weights. For German individual emotions, we notice that when eye movement features perform better than EEG features for neutral and negative emotions, the attentional weight gaps between EEG and eye movement modalities become narrower than those in positive emotions. For French, trends are similar to German that when eye movements perform better than EEG in neutral emotions, the gap becomes narrower than positive and negative emotions.

6.4. Intraculture subject independent analysis

In the ICSI setting, we also compare the CMs and attentional weights for different classifiers, emotions, and cultures, as shown in figures 5(b) and (d), respectively.

For Chinese, DCCA-AM outperforms unimodal situations for positive and neutral emotions, while for negative emotion, DCCA-AM only has a similar

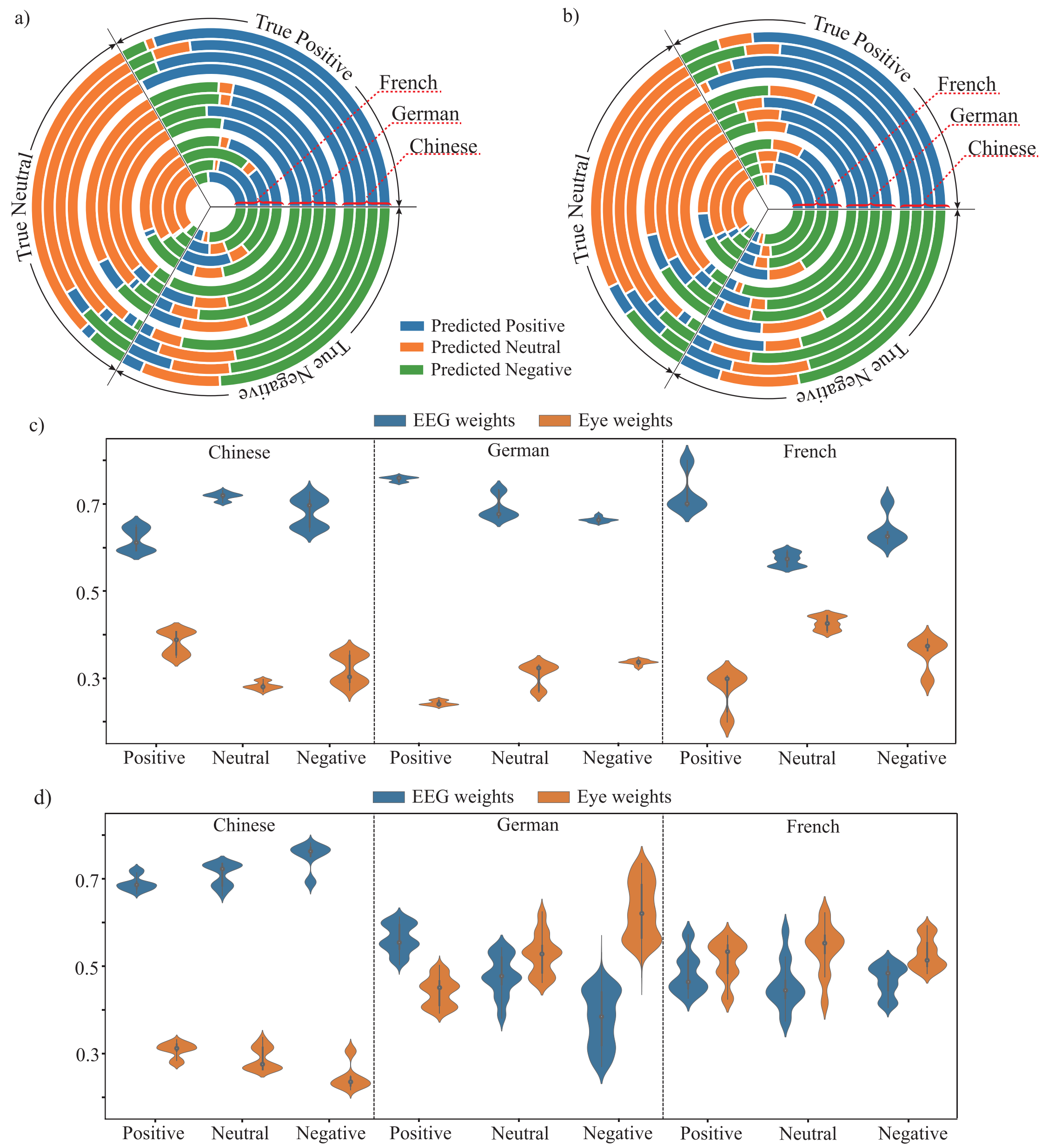

Figure 5. Confusion matrices (CMs) and attentional weights in ICSD (a), (c) and ICSI (b), (d) settings. (a), (b) Confusion matrices are shown in a nested pie plot. Outer four circles represent CMs of Chinese, middle four circles represent CMs of German, and inner four circles represent French. For each culture, the four circles are CMs of EEG features, eye movement features, concatenation fusion, and DCCA-AM methods. True test positive, neutral, and negative labels are shown as three fan-shaped sectors, and predicted emotions are depicted in different colors. (c), (d) Violin plots for attentional weights of EEG and eye movement modalities for various cultures and emotions.

performance as eye movement features. EEG achieves higher recognition accuracies for positive emotion and negative emotion than eye movements, where the gaps between EEG weights and eye movement weights are large. For neutral emotions, eye movements outperform EEG, corresponding to the trends to narrow the gap between two modalities.

For German and French individuals, eye movements outperform EEG in positive, neutral, and negative emotions, leading to different distributions: (1) The ICSI distributions in figure 5(d) for German and French individuals are much more compact than the ICSD distributions in figure 5(c), and there are no obvious gaps between EEG and eye movement attentional weights. (2) Eye movements have larger average weights than EEG for almost all three emotions except positive emotion in Germany.

6.5. Cross-culture analysis

In the cross-culture setting, we first analyze CMs and attentional weights as shown in figures 6 and 7, respectively. Since when one culture is used as test data, the other two cultures are used as training data separately, we plot 24 CMs (3 test cultures × 2 training cultures for each test culture × 4 classifiers) in figure 6. In addition, the emotion transfer chart in

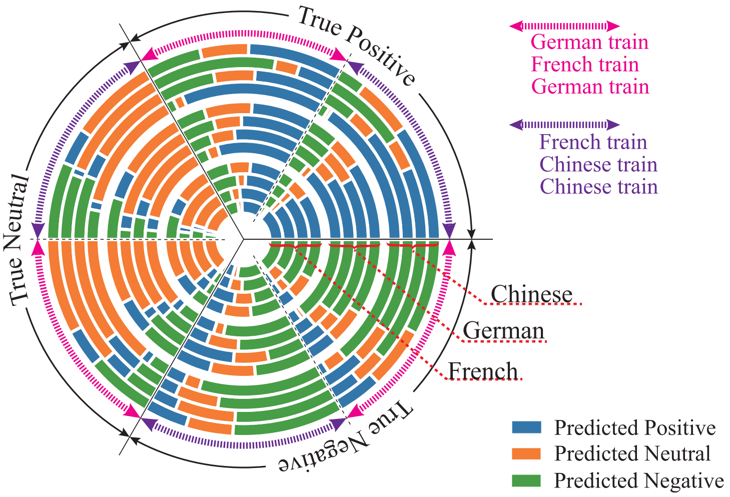

Figure 6. CCSI confusion matrices (CMs). Confusion matrices are shown as nested pie plot. Outer four circles represent test CMs of Chinese, middle four circles represent test CMs of German, and inner four circles represent test French. For each culture bundle, four circles are CMs of EEG features, eye movement features, concatenation fusion, and DCCA-AM method, respectively. True test positive, neutral, and negative labels are shown as three fan-shaped emotion sectors (separated by solid lines), and predicted labels are depicted in different colors. Each emotion sector is split into two small sectors (separated by dotted lines, pink sector and purple sector), which indicate different training cultures for three cultures: In pink sector from outer to inner, German, French, and German features are used as training and Chinese, German, and French as test correspondingly. In purple sector from outer to inner, French, Chinese, and Chinese samples are used training data, and Chinese, German, and French samples are used as test data.

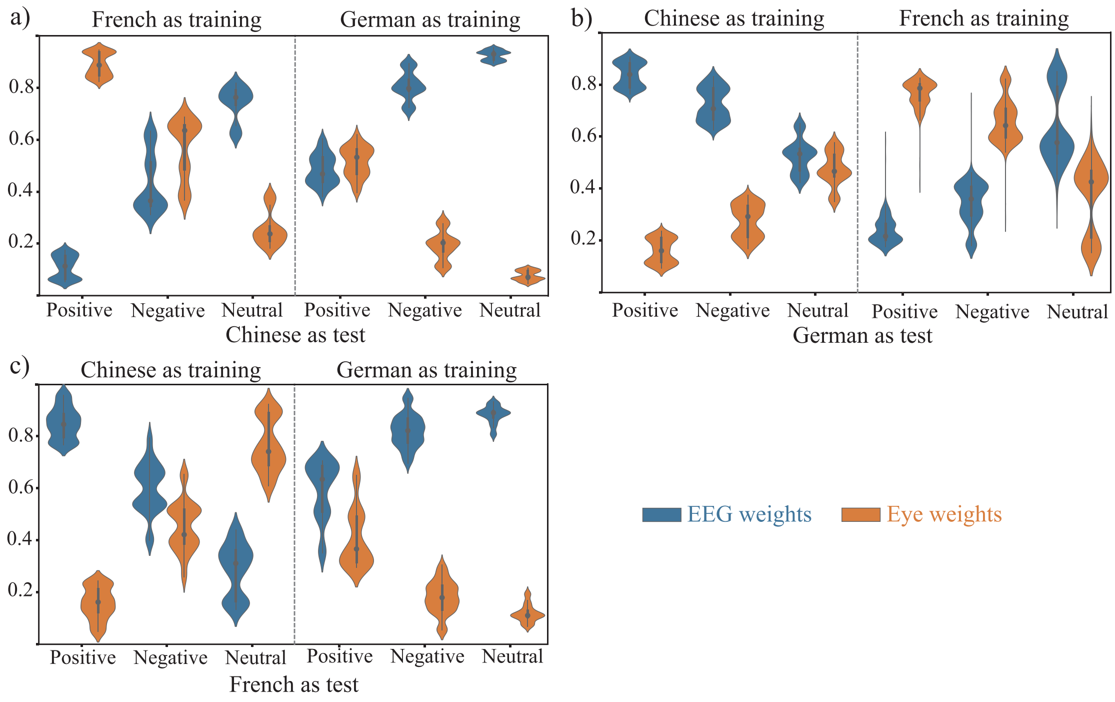

Figure 7. Attentional weight distributions in CCSI setting. Panels (a)–(c) represent situations where Chinese, German, and French are used as test sets, and the other two cultures are used as training sets, respectively.

figure 8 shows the accuracies when every culture is used for both training and test.

When Chinese is test data, from figure 6, for different classifiers, it is evident that the DCCA-AM method has the best performance for three emotions. For positive emotions, EEG has higher recognition accuracies than eye movements regardless of training cultures indicating that EEG is easily transferred from the training culture to Chinese. For neutral emotions, eye movements perform better than EEG, which suggests that eye movements play a more important role in neutral emotion. For negative emotions, when French is the training culture, EEG and eye movements have comparable performance, while when

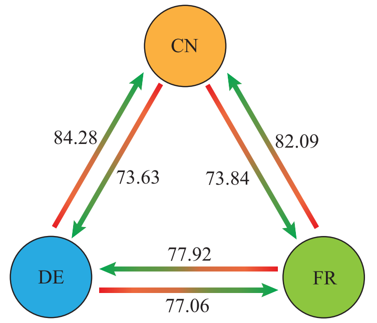

Figure 8. Emotion transferability chart in CCSI setting: arrow red side means the culture is used as training set, arrow green side means the culture is used as test set, and number next to the arrow is the best emotion recognition accuracy on test set (refer to table 9). This chart shows results when every culture is used as both training and test. CN: Chinese; FR: French; DE: German.

German is the training culture, eye movements outperform EEG. From figures 6 and 7(a), it seems that the attentional weights have a reversed relationship with the CMs: DCCA-AM assigns higher weights to modalities that have worse performance. Taking neutral emotion as an example, we find that eye movement features perform better than EEG features in both German and French training situations; however, DCCA-AM gives EEG features higher weights, fuses these two modalities, and achieves improvements. In addition, the attentional weights calculated by German and French training data show similar trends, namely, EEG plays an increasingly important role in positive, negative, and neutral emotions. These results might indicate that German and French individuals share similar culture-related emotional cognitive patterns and that the reversal attentional weight relationship might also indicate that the culture-related emotional cognitive patterns shared by German and French individuals are different from Chinese patterns.

When German and French are used as test data, from figures 6, 7(b) and (c), we find that (1) when Chinese are used as training data, the EEG weights decrease for positive, negative, and neutral emotions regardless of the cultural background of the test set. (2) When French is used as training data and German is used as test data, and when German is used as training data and French is used as test data, the EEG weights become increasingly larger, changing reversely to when Chinese is used as the training set. These two observations might again suggest that Germans and French people share culture-related emotional cognitive patterns within each other and that these patterns are different from those of Chinese people. However, we did not observe a consistent researsal relationship in attentional weight distributions when German and French are used as test data.

We then depict an emotion transfer chart as shown in figure 8. We focus on two situations: (1) a specific culture is the test data, and (2) a specific culture is the training data. The first situation reflects emotional transferability and shows how easily this culture is transferred from other cultures, and the second situation can reflect how suitable this culture is as the training set for cultural generalization, where researchers might build emotion models with a specific culture and generalize the model to other cultures.

For the first situation, when French is used as test data, the test emotion recognition accuracies when Chinese and German individuals as training data are 73.84% and 77.06%, respectively. When Chinese is used as test data, the test emotion recognition accuracies when German and French as training data are 84.28% and 82.09%, respectively. When German is used as test data, the test emotion recognition accuracies when Chinese and French as training data are 69.50% and 77.92%, respectively. It is obvious that (1) when used as test data, Chinese has higher test recognition accuracies (above 82%) than German and French (below 78%), and (2) when used as test data, Chinese has the smallest performance gap (2.19% between 84.28% and 82.09%), and German and French have larger performance gaps (4.29% between 77.92% and 73.63% for German, and 3.22% between 77.06% and 73.84% for French). These experimental results indicate that (1) German, not Chinese, transfers more easily to French and that French, not Chinese, transfers more easily to German, indicating that there might be shared culture-related emotional patterns between French and German; and (2) German and French have similar emotion transferaility on the Chinese test set.

For the second situation, it is worth noting that accuracies in the second situation cannot be used to evaluate emotion transferability since the test data are different. However, this situation might provide some insights for cultural generalization. When German and French are used as training data, the test accuracies are all above 77%, and the performance gaps between the test sets are 7.22% and 4.19%, respectively. When Chinese is used as training data, the test accuracies for German and French are below 74%, and the performance gap between the German and French test sets is 0.21%. As the training set, Chinese has the smallest test recognition accuracy gap, but at the same time, the test accuracies are also the lowest compared with when German and French are used as the training set. This suggests that Chinese might not be a good training set for cultural generalization.

When the data from Chinese are used as test data and the data from German or French are used as training data, we call information used to predict Chinese emotion states is transferred from the models built with the data from German or French. When the data from Chinese are used as training data and the data from German or French are used as test data, we call the models trained with the data from Chinese transfer Chinese emotional information to German or French. We can obtain the following two findings: (a) Chinese is much more easily transferred from other cultures than transferred to other cultures. In other words, the data from Chinese are a good fit for test data but not suitable for training data for the other two cultures. And (2) German and French seem to be more helpful in building emotion recognition models with better cultural generalization. However, more research is needed.

7. Conclusions and future work

In this paper, we have systematically evaluated the relationship between cultures and emotions with EEG and eye movements from an affective computing perspective. We believe that our findings could deepen the understanding of cultural influences on emotions, affective brain-computer interface, and emotion recognition models with good cultural generalization. First, we have collected EEG and eye movement data for native Chinese, German, and French individuals, and we have examined five EEG features. We have found that DE features performed better than other features. Second, we have evaluated unimodal and multimodal emotion recognition models in the ICSD, ICSI, and CCSI settings, and we have found that an in-group advantage exists in EEG-based emotion recognition according to average emotion recognition accuracies in in-group and out-group settings. Third, we have analyzed neural patterns by visualizing DE features, and we have found that the γ and β bands exhibit decreasing activities for positive, neutral, and negative emotions for Chinese, while for German and French, the θ and α bands exhibit increasing activities for positive, neutral, and negative emotions. Fourth, with CMs and attentional weights, we have analyzed intra- and intercultural influences on emotion recognition and we have found that French and German individuals might share culture-related emotional patterns that are different from those of Chinese individuals. Fifth, by analyzing cross-cultural emotion recognition accuracies, we have found that the data from Chinese are a good fit for test data but not suitable for training data for the other two cultures. This can help us build an emotion recognition system with good cultural generalization performance.

The field of cultural influences on emotions includes many topics, and we merely discussed some typical topics in this paper. In the future, we will work on this topic from the following aspects: (1) The amount of data from these cultures is not large enough, and the number of subjects is not the same for each culture. We will recruit more subjects with more cultural backgrounds to improve the dataset and build a culture-balance dataset. (2) For the ICSI and CCSI settings, we will use transfer learning methods to build subject-independent models, which might achieve better recognition accuracies by reducing subject differences.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://bcmi.sjtu.edu.cn/home/seed/.

Acknowledgments

Bao-Liang Lu was supported in part by grants from the National Natural Science Foundation of China (Grant No. 61976135), SJTU Trans-Med Awards Research (WF540162605), the Fundamental Research Funds for the Central Universities, the 111 Projects, and GuangCi Professorship Program of RuiJin Hospital Shanghai Jiao Tong University School of Medicine.

ORCID iDs

Wei Liu https://orcid.org/0000-0002-3840-1980
Wei-Long Zheng https://orcid.org/0000-0002-9474-6369
Ziyi Li https://orcid.org/0000-0002-8944-741X

References

[1] Ford B-Q and Mauss I-B 2015 Culture and emotion regulation *Curr. Opin. Psychol.* **3** 1–5

[2] Picard R-W 2000 *Affective Computing* (Cambridge, MA: MIT Press)

[3] D'mello S-K and Kory J 2015 A review and meta-analysis of multimodal affect detection systems *ACM Comput. Surv.* **47** 1–36

[4] Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H and Sander L-B 2021 Cultural adaptation of internet-and mobile-based interventions for mental disorders: a systematic review *npj Digit. Med.* **4** 1–18

[5] Ratner C 2000 A cultural-psychological analysis of emotions *Cult. Psychol.* **6** 5–39

[6] Thompson W-F and Balkwill L-L 2010 Cross-cultural similarities and differences *Handbook of Music and Emotion: Theory, Research, Applications* (Oxford: Oxford University Press) pp 755–88

[7] Jackson J-C, Watts J, Henry T-R, List J-M, Forkel R, Mucha P-J, Greenhill S-J, Gray R-D and Lindquist K-A 2019 Emotion semantics show both cultural variation and universal structure *Science* **366** 1517–22

[8] Cowen A-S, Fang X, Sauter D and Keltner D 2020 What music makes us feel: at least 13 dimensions organize subjective experiences associated with music across different cultures *Proc. Natl Acad. Sci.* **117** 1924–34

[9] Ekman P *et al* 1987 Universals and cultural differences in the judgments of facial expressions of emotion *J. Personality Soc. Psychol.* **53** 712

[10] Russell J-A 1994 Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies *Psychol. Bull.* **115** 102

[11] Cordaro D-T, Sun R, Keltner D, Kamble S, Huddar N and McNeil G 2018 Universals and cultural variations in 22 emotional expressions across five cultures *Emotion* **18** 75–93

[12] Laukka P and Elfenbein H-A 2021 Cross-cultural emotion recognition and in-group advantage in vocal expression: a meta-analysis *Emot. Rev.* **13** 3–11

[13] Elfenbein H-A and Ambady N 2002 On the universality and cultural specificity of emotion recognition: a meta-analysis *Psychol. Bull.* **128** 203

[14] Özkarar-Gradwohl F-G 2019 Cross-cultural affective neuroscience *Front. Psychol.* **10** 794

[15] Wu S-Y, Schaefer M, Zheng W-L, Lu B-L and Yokoi H 2017 Neural patterns between Chinese and Germans for EEG-based emotion recognition *2017 8th Int. IEEE/ Conf. on Neural Engineering (NER)* pp 94–7

[16] Gan L, Liu W, Luo Y, Wu X and Lu B-L 2019 A cross-culture study on multimodal emotion recognition using deep learning *Int. Conf. on Neural Information Processing* pp 670–80

[17] Ekman P and Friesen W-V 1971 Constants across cultures in the face and emotion *J. Personality Soc. Psychol.* **17** 124

[18] Tsai J-L, Chentsova-Dutton Y, Freire-Bebeau L and Przymus D-E 2002 Emotional expression and physiology in european americans and Hmong Americans *Emotion* **2** 380

[19] Miyamoto Y, Uchida Y and Ellsworth P-C 2010 Culture and mixed emotions: co-occurrence of positive and negative emotions in Japan and the United States *Emotion* **10** 404

[20] Lomas T 2021 Towards a cross-cultural lexical map of wellbeing *J. Posit. Psychol.* **16** 622–39

[21] Scherer K-R and Fontaine J-R-J 2019 The semantic structure of emotion words across languages is consistent with componential appraisal models of emotion *Cogn. Emot.* **33** 673–82

[22] Markus H-R and Kitayama S 1991 Culture and the self: implications for cognition, emotion and motivation *Psychol. Rev.* **98** 224

[23] Tsai J-L, Knutson B and Fung H-H 2006 Cultural variation in affect valuation *J. Personality Soc. Psychol.* **90** 288

[24] Park G, Lewis R-S, Wang Y-C, Cho H-J and Goto S-G 2018 Are you mad at me? Social anxiety and early visual processing of anger and gaze among Asian American biculturals *Cult. Brain* **6** 151–70

[25] Matsumoto D, Yoo S-H and Nakagawa S 2008 Culture, emotion regulation and adjustment *J. Personality Soc. Psychol.* **94** 925

[26] Soto J-A, Perez C-R, Kim Y-H, Lee E-A and Minnick M-R 2011 Is expressive suppression always associated with poorer psychological functioning? A cross-cultural comparison between European Americans and Hong Kong Chinese *Emotion* **11** 1450

[27] Tsai J-L and Qu Y 2018 The promise of neuroscience for understanding the cultural shaping of emotion and other feelings *Cul. Brain* **6** 99–101

[28] Han S, Northoff G, Vogeley K, Wexler B-E, Kitayama S and Varnum M-E 2013 A cultural neuroscience approach to the biosocial nature of the human brain *Ann. Rev. Psychol.* **64** 335–59

[29] Murata A, Moser J-S and Kitayama S 2013 Culture shapes electrocortical responses during emotion suppression *Soc. Cogn. Affect Neurosci.* **8** 595–601

[30] de Greck M, Shi Z, Wang G, Zuo X, Yang X, Wang X, Northoff G and Han S 2012 Culture modulates brain activity during empathy with anger *NeuroImage* **59** 2871–82

[31] Park B, Qu Y, Chim L, Blevins E, Knutson B and Tsai J-L 2018 Ventral striatal activity mediates cultural differences in affiliative judgments of smiles *Cult. Brain* **6** 102–17

[32] Özkarar-Gradwohl F and Turnbull O 2021 Gender effects in personality: a cross-cultural affective neuroscience perspective *Cult. Brain* **9** 79–96

[33] Tompson S-H, Huff S-T, Yoon C, King A, Liberzon I and Kitayama S 2018 The dopamine d4 receptor gene (DRD4) modulates cultural variation in emotional experience *Cult. Brain* **6** 118–29

[34] Lin L-C, Qu Y and Telzer E-H 2018 Cultural influences on the neural correlates of intergroup perception *Cult. Brain* **6** 171–87

[35] Corneanu C-A, Simón M-O, Cohn J-F and Guerrero S-E 2016 Survey on RBG, 3D, thermal and multimodal approaches for facial expression recognition: history, trends and affect-related applications *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 1548–68

[36] Srinivasan R and Martinez A-M 2021 Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild *IEEE Trans. Affective Comput.* **12** 707–21

[37] Sagha H, Deng J, Gavryukova M, Han J and Schuller B 2016 Cross lingual speech emotion recognition using canonical correlation analysis on principal component subspace *2016 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (ICASSP) pp 5800–4

[38] Zhang B, Provost E-M and Essl G 2017 Cross-corpus acoustic emotion recognition with multi-task learning: seeking common ground while preserving differences *IEEE Trans. Affect. Comput.* **10** 85–99

[39] Kleinsmith A, De Silva P-R and Bianchi-Berthouze N 2006 Cross-cultural differences in recognizing affect from body posture *Interact. Comput.* **18** 1371–89

[40] Ringeval F *et al* 2019 AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI and cross-cultural affect recognition *Proc. 9th Int. on Audio/Visual Emotion Challenge and Workshop* pp 3–12

[41] Liang J-J, Chen S-Z, Zhao J-M, Jin Q, Liu H-B and Lu L 2019 Cross-culture multimodal emotion recognition with adversarial learning *ICASSP 2019–2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (ICASSP) pp 4000–4

[42] Han J, Zhang Z-X, Pantic M and Schuller B 2021 Internet of emotional people: towards continual affective computing cross cultures via audiovisual signals *Future Gener. Comput. Syst.* **114** 294–306

[43] Knapp R-B, Kim J and André E 2011 Physiological signals and their use in augmenting emotion recognition for human-machine interaction *Emotion-Oriented Systems* (Berlin: Springer) pp 133–59

[44] Zheng W-L, Zhu J-Y and Lu B-L 2017 Identifying stable patterns over time for emotion recognition from EEG *IEEE Trans. Affect. Comput.* **10** 417–29

[45] García-Martínez B, Martinez-Rodrigo A, Alcaraz R and Fernández-Caballero A 2018 A review on nonlinear methods using electroencephalographic recordings for emotion recognition *IEEE Trans. Affect. Comput.* **12** 1

[46] Zheng W-L, Zhu J-Y, Peng Y and Lu B-L 2014 EEG-based emotion classification using deep belief networks *2014 IEEE Int. Conf. on Multimedia and Expo* (ICME) pp 1–6

[47] Kwon Y-H, Shin S-B and Kim S-D 2018 Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system *Sensors* **18** 1383

[48] Chen J-X, Zhang P-W, Mao Z-J, Huang Y-F, Jiang D-M and Zhang Y-N 2019 Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks *IEEE Access* **7** 44317–28

[49] Yang Y, Wu Q, Qiu M, Wang Y and Chen X 2018 Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network *2018 Int. Conf. on Neural Networks* (IJCNN) pp 1–7

[50] Ma J, Tang H, Zheng W-L and Lu B-L 2019 Emotion recognition using multimodal residual LSTM network *Proc. 27th ACM Int. Conf. on Multimedia* pp 176–83

[51] Alhagry S, Fahmy A-A and El-Khoribi R-A 2017 Emotion recognition based on EEG using LSTM recurrent neural network *Emotion* **8** 355–8

[52] Zhong P, Wang D and Miao C 2020 EEG-based emotion recognition using regularized graph neural networks *IEEE Trans. Affect. Comput.* 1–1

[53] Song T, Zheng W, Song P and Cui Z 2018 EEG emotion recognition using dynamical graph convolutional neural networks *IEEE Trans. Affect. Comput.* **11** 532–41

[54] Zheng W-L and Lu B-L 2016 Personalizing EEG-based affective models with transfer learning *Proc. 25th Int. Conf. on Artificial Intelligence* pp 2732–8

[55] Li J, Qiu S, Shen Y-Y, Liu C-L and He H 2020 Multisource transfer learning for cross-subject EEG emotion recognition *IEEE Trans. Cybern.* **50** 3281–93

[56] Zhao L-M, Yan X and Lu B-L 2021 Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition *Proc. 35th AAAI Conf. on Artificial Intelligence* pp 863–70

[57] Hartmann K-G, Schirrmeister R-T and Ball T 2018 EEG-GAN: generative adversarial networks for electroencephalograhic (EEG) brain signals (arXiv:1806.01875)

[58] Luo Y, Zhu L-Z, Wan Z-Y and Lu B-L 2020 Data augmentation for enhancing EEG-based emotion recognition with deep generative models *J. Neural Eng.* **17** 056021

[59] Zheng W-L, Dong B-N and Lu B-L 2014 Multimodal emotion recognition using EEG and eye tracking data *2014 36th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 5040–3

[60] Lu Y-F, Zheng W-L, Li B-B and Lu B-L 2015 Combining eye movements and EEG to enhance emotion recognition *24th Int. Conf. on Artificial Intelligence* pp 1170–6

[61] Baltrušaitis T, Ahuja C and Morency L-P 2018 Multimodal machine learning: a survey and taxonomy *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 423–43

[62] Zheng W-L, Liu W, Lu Y-F, Lu B-L and Cichocki A 2019 Emotionmeter: a multimodal framework for recognizing human emotions *IEEE Trans. Cybern.* **49** 1110–22

[63] Liu W, Qiu J-L, Zheng W-L and Lu B-L 2021 Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition *IEEE Trans. on Cognitive and Developmental Systems* p 1

[64] Ngiam J, Khosla A, Kim M, Nam J, Lee H and Ng A-Y 2011 Multimodal deep learning *Int. Conf. on Machine Learning* pp 689–96

[65] Liu W, Zheng W-L and Lu B-L 2016 Emotion recognition using multimodal deep learning *Int. Conf. on Neural Information Processing* pp 521–9

[66] Andrew G, Arora R, Bilmes J and Livescu K 2013 Deep canonical correlation analysis *Int. Conf. on Machine Learning* pp 1247–55

[67] Zheng W-L and Lu B-L 2015 Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks *IEEE Trans. on Auton. Mental Dev.* **7** 162–75

[68] Schaefer A, Nils F, Sanchez X and Philippot P 2010 Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers *Cogn. Emot.* **24** 1153–72