This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

Inter-brain EEG Feature Extraction and Analysis for Continuous Implicit Emotion Tagging during Video Watching

Yue Ding, Xin Hu, Zhenyi Xia, Yong-Jin Liu, *Senior Member, IEEE,* Dan Zhang*, Member, IEEE*

Abstract— How to efficiently tag the emotional experience of multimedia contents is an important and challenging problem in the field of affective computing. This paper presents an EEG-based real-time emotion tagging approach, by extracting inter-brain features from a group of participants when they watch the same emotional video clips. First, the continuous subjective reports on both the arousal and valence dimensions of emotion were obtained by employing a three-round behavioral rating paradigm. Second, the inter-brain features were systematically explored in both spectral and temporal domain. Finally, regression analyses were performed to evaluate the effectiveness of inter-brain amplitude and phase features. The inter-brain amplitude feature showed significantly better prediction performance than the inter-brain phase feature, as well as another two conventional features (spectral power and inter-subject correlation). By combining the four types of features, regression values (R^2) were obtained for the prediction of arousal (0.61±0.01) and valence (0.70±0.01), corresponding to prediction errors of 1.01±0.02 and 0.78±0.02 (unit on 9-point scales), respectively. The contributions of different electrodes and frequency bands were also analyzed. Our results show promising potentials of inter-brain EEG features in real-time emotion tagging applications.

Key Words—emotion, inter-brain, EEG, implicit tagging

1 INTRODUCTION

THE ever-growing volume of multimedia resources available to us brings an increasing need for effective techniques to manage and retrieve multimedia contents. Among diverse approaches, tags have been widely used as an effective form. In industry, commercial web systems (e.g., Youtube, Last.fm and Flickr) have successfully introduced a variety of toolkits based on tags to assist different users in discovering, exploring and sharing media contents [1]. It is essential to annotate multimedia items with accurate semantic tags. In particular, tags involving emotion information play an important role, since multimedia contents, such as videos and music, are primarily valued for their ability to induce specific emotional experiences from people [2]. Emotion tagging of multimedia resources has been utilized to improve the retrieval and recommendation of multimedia items such as music (e.g.,[3], [4]), images (e.g., [5]–[7]), and videos (e.g.,[8], [9]).

Among affective retrieval or recommendation systems, there're mainly two kinds of multimedia emotion tagging methods: multimedia content analysis methods and human annotation methods. The multimedia content analysis methods extract low-level physical features from visual or audio streams (e.g., the color, motion and lighting features from videos. and the pitch, formant, and energy features from audios) and train machine-learning classifiers with certain ground truth emotion labels. Despite the explosive growth of content analysis methods, automatic tagging technologies still can hardly achieve satisfactory performance on real-world multimedia data that vary widely in genre, quality, and content [10]–[12].

Meanwhile, the human annotation methods have been demonstrated to be capable of generating tags which are directly derived from human behavioral or neural response [2], [13]. The human annotation methods can be further divided into two categories: explicit tagging and implicit tagging. The former means that users manually assign tags for the delivered multimedia contents, and the latter refers to autonomously generating tags from people's spontaneous responses when consuming multimedia contents. Although users' explicit tagging is widely used in some popular social media websites (e.g., Twitter, Instagram, YouTube, etc.) for data retrieving and recommending purposes, manual tagging is very time consuming and labor-intensive for the ever-expanding online multimedia contents [14], [15]. More importantly, when it comes to emotion tagging for those continuous media streams like audios and videos, the explicit tagging approach is inadequate in tracking the fast-changing temporal dynamics of human emotion experiences. Users usually tag their emotions in a post hoc manner. Performing real-time emotion tagging is difficult and very challenging, as it requires frequently reporting of emotion tags, which would severely interfere with users' experiencing behavior. Therefore, various implicit tagging methods have been proposed as

- *Y.Ding is with the Department of Biomedical Engineering, Tsinghua University, Beijing, P.R. China, 100084. E-mail: yueding.bme@gmail.com.*
- *X.Hu, Z. Xia and D. Zhang are with the Department of Psychology, Tsinghua University, Beijing, P.R. China, 100084. E-mail: huxin530@gmail.com, gene2788@163.com, dzhang@tsinghua.edu.cn.*
- *Y-J.L. is with the Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, P.R. China, 100084. E-mail: liuyongjin@tsinghua.edu.cn.*
- *Y.D. and X.H. contribute equally to the paper.*
- *D. Z. is the corresponding author.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

an alternative or complementary approach, such as autonomously generating affective tags from people's facial expression [16]–[18], eye gaze [19], [20], physiological signals [21], [22], and neural signals [23]–[25]. Among these possible measurements, neural signals have increasingly attracted interest in this field in recent years because they provide a direct measurement of human emotion responses [26], [27].

Electroencephalography (EEG) is one of the most popular and accessible neural signal measurement techniques. EEG has been widely used in the field of affective computing due to its fine temporal resolution, relatively low cost and high portability, as compared to functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) techniques. [13], [28]. To date, emotion experiences such as the valence and arousal dimensions, or typical discrete positive and negative emotions, have been reported to be effectively recognized using EEGs [28]–[30]. Notably, nearly all studies extracted EEG features on the basis of a single person's brain responses, and consequently emotion recognition models were trained and hereby applicable specifically for each individual person. These individualized designs are motivated by the high individual variability in emotion experience[31].

Despite of the popularity of individual-based feature extraction methods, some recent studies suggest analyzing inter-brain neural features, which serves as a promising approach to characterize a variety of cognitive functions including emotions [32]–[34]. The majority of the inter-brain studies have been focusing on fMRI responses and their inter-subject correlation (ISC), which is calculated as the average of pairwise Pearson correlations between the time courses of the voxel-wise BOLD responses from all possible pairs of individuals exposed to identical complex and continuous stimuli [32]. The ISC feature is considered to reflect the consistency of neural responses across a group of participants, rather than individual's response magnitude. fMRI studies have found that ISCs over sensory and higher-order cortices were significantly correlated with auditory and visual perception, speech engagement, as well as emotional dimensions of arousal and valence [33], [34]. However, the exploration on EEG-based inter-brain features is still very rare; more details are summarized in Section 2.1).

The aim of the present study is to investigate the effectiveness of different inter-brain EEG features for implicitly tagging continuous emotions during video watching. Continuous subjective reports on emotion experience were obtained by using a real-time behavioral rating paradigm. Inter-brain EEG features were systematically explored in both the spectral and the temporal domains, by filtering the signals into different frequency bands and decomposing into amplitude and phase components. In our study, the linear regression method was employed to evaluate the effectiveness of these features. Our results suggest that inter-brain EEG features are promising for affective computing applications.

2 Related Work

2.1 EEG-based inter-brain features

In single-brain emotion recognition models, EEG spectral powers and their spatial distributions are possibly the most widely used features. For instance, frontal asymmetry of alpha power and beta-to-alpha ratio associated with valence or approach-withdrawal level, occipital alpha power closely related to arousal level, etc. [35]–[37]. Whereas response magnitude is important for the features mentioned above, inter-brain analysis focus on a completely different aspect, i.e., consistency of neural responses across participants. Hereby, it is necessary to systematically evaluate possible contributions of different EEG features in the inter-brain context.

Researchers are beginning to employ EEG-based inter-brain analysis to study different cognitive functions. Most studies have followed the fMRI approach, i.e., by calculating the pairwise correlations between the time series of EEGs in a channel-wise manner. A few studies working on emotion have reported that such inter-brain correlations could predict the participants' preference and engagement [38], [39]. The rich spectral and temporal information, however, has not been fully utilized. It is reasonable to assume that inter-brain neural coupling at different frequency bands of EEG signals may carry distinct cognitive functions. In multi-person EEG studies on social interactions, it has been reported that inter-brain correlation of the amplitudes of theta and alpha over the right temporal-parietal junction, the amplitudes of alpha and beta over frontal regions were associated with the understanding of others' intention and high-level cooperative strategies [40], [41]. In another emotion-related study [42], the inter-brain consistency in the delta band was shown to have a primary contribution to behaviorally measured audience preference. Therefore, extracting inter-brain features from different frequency bands may help us to better reflect different cognitive components that are important for emotion experience.

To further explore EEG signals in the temporal domain, it is necessary to consider its phase and amplitude decomposition. Separating amplitude and phase has been suggested to provide distinct insights into the neural mechanisms of a variety of cognitive functions [43]–[45]. Specifically, phase synchronization has already been suggested to be related to top-down control and reflecting the timing of neural populations for the exchange of information between the global and local neuronal networks [46]–[48], whereas amplitude responses have been related to the excitability of local neural assemblies [49], [50]. While a number of previous studies have suggested the functional importance of inter-brain phase synchrony in relatively low-frequency bands, such as delta, theta, alpha, in a variety of social interaction paradigms [51]–[53], the possible contributions of inter-brain phase and amplitude features for emotion recognition remain to be elucidated.

2.2 The ground truth for emotion tagging

Before developing an effective emotion tagging model for tracking the fast-changing temporal dynamics of our

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

emotion experiences during video watching, it is important to obtain reliable human annotated labels as the ground truth for later training purpose. These labels were usually obtained by recruiting a group of participants to manually report their subjective emotional experiences after the perception of given emotional stimuli [28], [29]. The reports could be multi-dimensional, including a set of discrete emotion categories (e.g., joy, anger, sadness, fear) and they were normally asked to give a numerical evaluation on each dimension, e.g., by using a 7-point Likert Scale. In practice, the majority of EEG-based emotion tagging studies have employed a simplified version of the obtained labels, e.g., by translating the numerical reports into two levels to represent the high and low state of one specific emotion [54]–[56]. While such a translation might be encouraged by the popularity of the binary classification methods, a few studies have adopted the regression methods to take the full advantage of the labels [57], [58].

Although the above-mentioned human annotation method has been widely used, it might not be sufficient to provide the ground truth labels for real-time emotion tagging scenarios. Specifically, the participants were required to report their emotional experiences in a post-hoc manner and therefore they were assumed to experience a constant magnitude of emotion throughout one elicitation procedure. While such an assumption might be valid for the perception of briefly presented emotional stimuli, complex and dynamically changing stimuli (e.g., videos) requires further development of the reporting method for obtaining the time course of emotion experiences. Alternatively, a dynamic annotation method has been proposed [34], [58], in which the participants were asked to go through the emotion elicitation procedure a second time and performed the report in a synchronized way with the stimuli. More repetition or more participants would be needed, if more emotion dimensions were to be labeled. While such a method could be time consuming and increase the burden of the participants, it is believed to provide more accurate labels for training emotion tagging models.

It is worth noting that the dynamic annotation method is not designed for application-oriented scenarios, but mainly for research purposes. The obtained real-time ratings can be used to investigate feature extraction and machine learning methods (as in the present study). The trained models can then be applied to implement advanced real-time and automatic emotion recognition systems, which do not need real-time subjective reports.

3 Method

3.1 Participants

Thirteen undergraduates from Tsinghua University were recruited (six females, mean age = 21 years, ranging 19 – 23 years) as paid volunteers. All of them had normal hearing, normal or corrected-to-normal vision. Informed consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki and approved by the local Ethics Committee of Tsinghua University.

3.2 Materials

Twenty clips of emotion eliciting videos from the MAHNOB-HCI dataset were used in the current study as the testing stimuli, including video clips selected from commercial movies like Hannibal, Mr. Bean's holiday, Love Actually, as well as from online resources like youtube.com and blip.tv. All of the video clips have been validated to be able to induce different emotional valence and arousal states (see more descriptions about those videos in [29]). The durations of those video clips ranged from 35 seconds to 117 seconds (mean duration=81.4±22.5 seconds). The emotional ratings of these 20 video clips by the participants in the present study are shown in Figure 1B.

3.3 Procedure

The experiment was carried out in a regular laboratory environment with ambient illumination from ceiling lights and without any electrical shielding. The stimuli were displayed on an LCD monitor (22-inch, DELL, USA) with a 60 Hz refreshing rate. Stereo speakers (DELL, USA) were used, and the sound volume was set at a fixed and comfortable level.

The experiment consisted of three rounds, each with the same 20 video clips presented as 20 trials (orders were independently randomized within each round). Before the first round, three practice trials were given to familiarize the participants with the procedure.

The first round aimed to collect the EEG signals during video watching. In each trial, the participants watched one

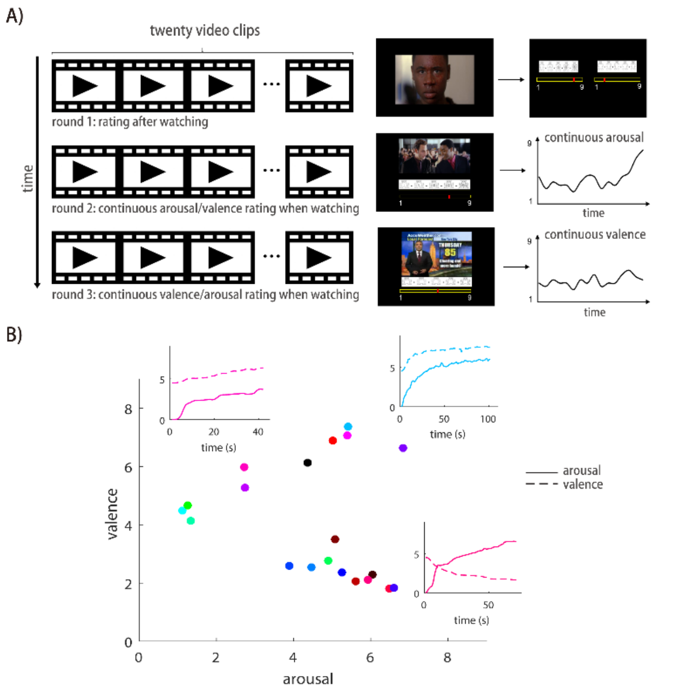

Fig. 1. Experimental procedure. (A) Left: an illustration of the three-round experiment; Right: the rating interfaces used in the three rounds. The order of round 2 and round 3 was randomized across participants. (B) An overview of the subjective ratings on arousal and valence. The colored dots represent the overall ratings of each individual video clips, and the three subplots depict the representative continuous ratings of the video clips shown in the corresponding colors. The solid and dashed lines indicate the continuous ratings on arousal and valence, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

clip of emotional videos, and reported their overall subjective emotion experiences afterward on both the valence and arousal dimensions on 9-point scales (0-8). Participants performed the reports by moving red vertical bars along the scales (Fig. 1A). Between every two sequential trials, the participants were asked to watch a 19-second neutral video (a color bar test pattern) in order to recover from the previously induced emotional state as much as possible. After the completion of all the 20 trials, participants took a rest for 5 minutes.

The second and third rounds were designed to collect people's continuous real-time emotional experiences to the same video stimuli presented in the first round. The participants performed a real-time report by moving a red vertical bar while watching replays of videos presented in the first round, using a computer mouse (Fig. 1A). They could move the red bar freely in all trials and the bar would stop at one of the two ends when the mouse moved beyond the range. The bar could be moved continuously in real-time, and it could be kept in a specific position as well, when reporting a period of constant emotion experiences. They were instructed to recall their real-time emotion experiences of watching the video stimuli for the first time, and the replays served as the memory clue to facilitate their recall and provided the temporal information for recording their real-time reports. The subjective reports were finalized into 1-second resolution time series for further EEG analysis, as a second-level temporal resolution was deemed sufficient to capture the temporal dynamics of human emotional experiences. The participants reported their real-time emotion on valence and arousal in the second and third round, respectively. The order of the emotional dimensions was counter-balanced across participants. There was a 5-minute rest between the two rounds.

The purpose of such a design was to obtain more accurate labels for evaluating the inter-brain features. The design of our EEG experiment followed a previous fMRI study [34] for the collection of continuous real-time subjective reports. An overview of the experimental procedure is shown in Figure 1A. Presentation of the stimuli and the rating procedure were programmed in MATLAB (The Mathworks, USA) using the Psychophysics Toolbox 3.0 extensions [59].

3.4 EEG Recordings

EEG signals were recorded with the NeuSen.Uamp, system (Neuracle, China) at a sampling frequency of 1000 Hz. Fifteen electrodes were arranged according to the international 10-20 system (F3/4, Fz, C3/4, T3/4, T5/6, P3/4, Pz, O1/2, Oz), with reference at Cz and a forehead ground at FPz. Electrode impedances were kept below 10 kOhm for all electrodes.

3.5 Signal Preprocessing and Feature Extraction

Due to personal issues, two participants' behavioral data were partly lost, therefore the continuous emotional rating was averaged over the remaining 11 participants. For all the 13 participants, their EEG signals were first downsampled to 200 Hz, notch filtered to remove the 50 Hz powerline noise, and bandpass filtered to 0.5-50 Hz. Artifacts related to eye-movement, muscle movement, and other possible environmental noises were removed using independent component analysis (ICA). On average, 1-2 independent components were excluded per participant. The remaining ICs were then back-projected onto the scalp EEG channels, reconstructing the artifact-free EEG signals.

The preprocessed data were then band-pass filtered into four bands, namely theta (4-7 Hz), alpha (8-13 Hz), beta (14-29 Hz) and gamma (30-47 Hz). Similar with previous EEG studies focusing on emotion dynamics [60]–[62], here delta (1-4 Hz) band was not included as it could not be effectively preserved and extracted with the chosen 1-second temporal dynamics (see below), according to the signal processing theory. The data were further segmented into trials according to the 20 video clips. All the band-pass filtered segments were subjected to a Hilbert transform

$$\xi^i(t) = x^i_{m,n}(t) + jh(x^i_{m,n}(t)) \tag{1}$$

where t is time, $x^i_{m,n}(t)$ represents EEG data from the m-th channel, n-th trial (video clip) and the i-th participant and $h(x^i_{m,n}(t))$ represents its Hilbert transform, and j is the imaginary unit. Hereby, the instantaneous amplitude $A^i_{m,n}(t)$ and the instantaneous phase $\phi^i_{m,n}(t)$ of the i-th participant are

$$A^i_{m,n}(t) = \sqrt{(x^i_{m,n}(t))^2 + (h(x^i_{m,n}(t))^2} \tag{2}$$

$$\phi^i_{m,n}(t) = tan^{-1}(h(x^i_{m,n}(t))/x^i_{m,n}(t)) \tag{3}$$

The above formulae are applied to the EEG data of all channels from all participants, at all the four frequency bands, for all video clips, respectively.

The inter-brain EEG features were then collected as follows. The inter-brain phase feature (termed as *i-Phase*) was formed by averaging the phase-locking values (*PLVs*) over a certain period of time, in which *PLVs* were calculated as the consistency of phases across participants

$$PLV_{m,n}(t) = \sum_{i=1}^{N} \exp(\phi^i_{m,n}(t))/N \tag{4}$$

where N is the number of participants. PLV varies between 0 and 1: When PLV = 1, the phases are completely locked to a fixed phase angle; when the phases are uniformly distributed between 0 and 2π, PLV = 0. Note that PLV is an amplitude-independent measurement as it assigns constant unit amplitude on each trial.

Inter-brain amplitude feature (termed as *i-Amplitude*), however, was obtained by averaging the instantaneous amplitudes across all participants over a certain period of time, in which the averaged amplitude feature at one time point is calculated as

$$A_{m,n}(t) = \sum_{i=1}^{N} A^i_{m,n}(t)/N \tag{5}$$

A large amplitude value indicates a neural response of a high magnitude. As phase information is separated out, the *i-Amplitude* feature reflects a genuine response intensity. Here a time period of 1 second was selected to extract both the *i-Phase* and *i-Amplitude* features, following the selection in previous studies [16], [57]

To compare the proposed inter-brain features to existing methods, we also extracted two typical features which were widely used in previous studies. The first feature was the EEG spectral power feature (termed as *Power*),

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

which was probably the most popular individual-based feature [63], [64]. In our study, we extracted the spectral power at the four frequency bands (theta, alpha, beta, and gamma) as used in the inter-brain features by computing the square of the intensity of the bandpass filtered signal. The power feature used in the present study was the averaged spectral power across the participants (therefore also an inter-brain feature). The second feature was the inter-subject correlation feature (termed as *ISC*), which is the most widely used inter-brain feature to date [34], [45]. Here we calculated *ISC* for each frequency band respectively, as follows

$$ISC(k) = \frac{1}{N(N-1)/2} \sum_{i=1}^{N} \sum_{j=2, j>i}^{N} r_{ij} \qquad (6)$$

where N is the number of participants, and r_{ij} represents the temporal (Pearson's) correlation between participant i and j, given one pair of epoched data at a specific channel from the *k-th* 1-second epoch.

All of those extracted features (*i-Phase, i-Amplitude, Power,* and *ISC*) had a maximal dimensionality of 60 (4 frequency bands × 15 channels). In order to obtain maximal emotional responses (following the procedure in [28]), the EEG signals from the first round were used: the total number of 1611-second long video duration resulted in 1611 data samples (1-second non-overlapping time window). Correspondingly, the continuous subjective emotion experiences on both valence and arousal dimensions of these time points were obtained as well, by calculating the averages of the individual reports.

Before performing the regression analysis, the bivariate correlations between each EEG feature and the subjective emotion experiences of valance and arousal were calculated. Then the extracted feature data were used for the linear regression analysis, with continuous experiences on either valence or arousal as the dependent variable and one type of EEG features as the independent variables (4 frequency bands × 15 channels for each type). Both the fitness and the prediction error of the regression models were taken as the indexes for evaluation, where the fitness was the R^2 statistic of linear regression, and the difference between the actual real-time tagging and the regressed rating was the prediction error. To explore the possible best performance and the most significant features, a LASSO (least absolute shrinkage and selection operator) regression was employed, taking all types of features as input and set the number of output non-zero features as 60 (to make the results best comparable with the other regression models). When examining the influence of frequency bands and spatial distribution, linear regression models were also built with features from each frequency band as the independent variables (4 extracted features × 15 channels). In order to perform statistical tests on the prediction capability of the EEG features, a bootstrap procedure was applied for all regression analyses, with 1000 repetitions, and two-sample t-tests were applied to compare the independently bootstrapped samples. To control the influence of multiple comparisons, a false discovery rate (FDR) correction was conducted on p values of all statistical tests [65].

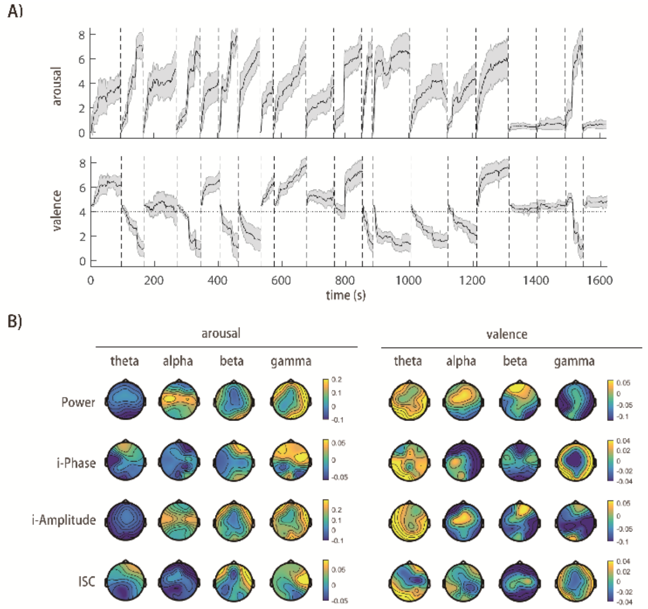

Fig. 2. Continuous explicit tagging and its correlation with inter-brain EEG features. (A) Time series of explicit emotion tagging with the 95% confidence interval of the arousal (upper) and valence (lower) scores across the 11 participants. Vertical lines denote breaks between the video clips. In the valence plot, the horizontal line at 4 shows the neutral valence. (B) Bivariate correlations between the inter-brain EEG features and the emotion experiences. The correlations for arousal and valence are shown in the left and right panels, respectively.

Another question that we need to answer is whether inter-brain features can outperform the best performing individual's single-brain features, rather than the grand-average. Hereby, we performed a similar regression analysis for each individual's EEG data, using both the *Power* and *i-Amplitude* feature (*i-Phase* and *ISC* features cannot be calculated for individuals) and compared their effectiveness to the inter-brain results.

Lastly, although we only had 13 participants' data, we preliminarily explored the influence of the number of participants on the regression results. For each given number of participants, all possible participant combinations were included.

4 RESULT

4.1 Subjective emotional experience and correlations with EEG features

The overall scores of the video clips were shown in the valence/arousal space in Figure 1B, where the continuous explicit tagging traces of three videos from distinguished three clusters were given as the examples for three classes: neutral (average valence: 4.49±0.63, low arousal: 0.55±0.54), positively excited (high valence: 6.10±1.44, high arousal: 3.56±2.52) and negatively excited (low valence: 2.73±1.58, high arousal: 4.05±2.63).

Participants' continuous explicit tagging was shown in Figure 2A, in which the 95% confidence interval (shadow range in Fig. 2A) indicating the reliability of the participants' explicit tags. The range of the ratings also confirmed that the video clips we used were able to elicit strong and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

temporally dynamic emotional reactions, with valence ranging from 0.01 to 7.86 and arousal ranging from 0.86 to 7.80. The bivariate correlations were conducted to reveal the relationship between the four extracted EEG features *(Power, i-Amplitude, i-Phase* and *ISC*) and the real-time subjective emotion experiences during video viewing. As shown in Figure 2B, for arousal, *Power* and *i-Amplitude* features showed stronger correlations than the other two, especially in alpha and gamma frequency bands (left panel, Fig. 2B). For valence, the four features had a similar magnitude of correlation strength, with stronger values for lower frequency bands such as theta and alpha (right panel, Fig. 2B).

4.2 Regression results

Figure 3 shows the results of regression with the arousal ratings as the dependent variable. Among all inter-brain EEG features, *i-Amplitude* feature achieved the best fitted model (R^2 = 0.59±0.02, Mean ± Standard Deviation calculated from bootstrapping), and correspondingly the smallest prediction error (1.04±0.02, unit on the 9-point scales, see green bar in Fig. 3A). The classical *Power* feature also obtained a good performance (R^2 = 0.56±0.02, prediction error = 1.08±0.03, see red bar in Fig. 3A). The prediction capabilities of *i-Phase* (R^2 = 0.09±0.01, prediction error = 1.65±0.03, see yellow bar in Fig. 3A) and *ISC* (R^2 = 0.09±0.01, prediction error = 1.66±0.03, see blue bar in Fig. 3A), however, were significantly worse. When combing all the inter-brain features, the best performance (R^2 = 0.61±0.01) with the smallest error (prediction error = 1.01±0.02) was obtained (see black bar in Fig. 3A). The performance of the combination of all the features significantly surpassed all the individual features in terms of both R2 (V.S. *Power*: $t(1998)=-70.95$, $p<0.001$; V.S. *i-Phase*: $t(1998)=-896.88$, $p<0.001$; V.S. *i-Amplitude*: $t(1998)=-26.23$, $p<0.001$; V.S. *ISC*: $t(1998)=-894.98$, $p<0.001$) and prediction error (V.S. *Power*: $t(1998)=61.98$, $p<0.001$; V.S. *i-Phase*: $t(1998)=582.91$, $p<0.001$; V.S. *i-Amplitude*: $t(1998)=21.76$, $p<0.001$; V.S. *ISC*: $t(1998)=583.64$, $p<0.001$). The performance of *i-Amplitude* was significantly better than the other three features in terms of both R2 (V.S. *Power*: $t(1998)=-45.71$, $p<0.001$; V.S. *i-Phase*: $t(1998)=-833.13$, $p<0.001$; V.S. *ISC*: $t(1998)=-831.75$, $p<0.001$) and prediction error (V.S. *Power*: $t(1998)=40.86$, $p<0.001$; V.S. *i-Phase*: $t(1998)=552.55$, $p<0.001$; V.S. *ISC*: $t(1998)=553.39$, $p<0.001$). There was no significant difference between *i-Phase* and *ISC* (R2: $t(1998)=2.65$, $p>0.05$; prediction error: $t(1998)=-2.28$, $p>0.05$), and both of the two features' performance was significantly worse than *Power* in terms of R2 (V.S. *i-Phase*: $t(1998)=701.59$, $p<0.001$; V.S. *ISC*: $t(1998)=701.15$, $p<0.001$) and prediction error (V.S. *i-Phase*: $t(1998)=-484.87$, $p<0.001$; V.S. *ISC*: $t(1998)=485.93$, $p<0.001$). When combining all the features, among the chosen 60 features in the LASSO model, *Power* accounted for the largest portion (40.00%), followed by *i-Amplitude* (30.00%), while *ISC* contributed least (10.00%) (Figure 3B). The beta parameters of the regression model were taken to evaluate the contribution of each individual inter-brain feature: as listed in Table 1, the top 10 significant features mainly came from *i-Amplitude* and *Power* features in relatively high frequency bands, with half of them from parietal-occipital areas. The predicted real-time ratings by different types of inter-brain features are shown in Figure 3C, in which the models using

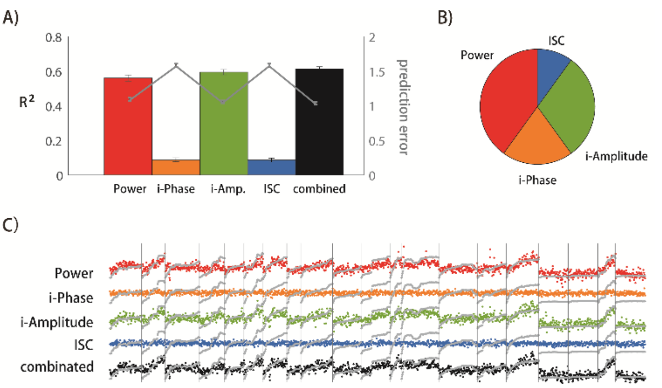

Fig. 3. Regression results for arousal. (A) Model fitting R^2 (bars with the left scale) and model prediction error (gray lines with the right scale) of the regression using the four inter-brain features and their combination. Error bars indicate standard deviation. (B) The percentages of each type of features in the chosen 60 features in the combined regression model. (C) The predicted ratings using different types of features for all 20 video clips. The gray lines represent the behavioral ratings.

TABLE 1
THE TOP 10 SIGNIFICANT FEATURES OF THE REGRESSION MODEL FOR AROUSAL

Beta	Feature	Frequency band	Electrode
0.90	*i-Amplitude*	Gamma	F3
0.79	*i-Amplitude*	Gamma	Oz
-0.66	*i-Amplitude*	Gamma	Fz
-0.53	*Power*	Gamma	P4
0.48	*i-Amplitude*	Gamma	C4
-0.47	*i-Amplitude*	Alpha	T4
0.46	*i-Amplitude*	Beta	Oz
-0.44	*Power*	Gamma	O1
0.38	*Power*	Gamma	O2
0.34	*Power*	Alpha	C3

TABLE 2
THE TOP 10 SIGNIFICANT FEATURES OF THE REGRESSION MODEL FOR VALENCE

Beta	Feature	Frequency band	Electrode
-0.88	*i-Amplitude*	Gamma	C4
-0.70	*i-Amplitude*	Gamma	O2
0.65	*i-Amplitude*	Gamma	T4
0.59	*Power*	Beta	Oz
-0.54	*i-Amplitude*	Gamma	P3
0.48	*i-Amplitude*	Gamma	P4
0.47	*i-Amplitude*	Gamma	O1
0.45	*i-Amplitude*	Theta	Oz
-0.41	*i-Amplitude*	Theta	F4
0.40	*Power*	Gamma	P4

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

i-Amplitude and *Power* features effectively followed the behavioral ratings.

Figure 4 shows the result of regression models with valence as the dependent variable. Similarly, the *i-Amplitude* and *Power* features showed better performances than the other two features (Fig. 4A, *Power*: $R^2 = 0.62\pm0.02$, prediction error = 0.87±0.02; *i-Amplitude*: $R^2 = 0.67\pm0.01$, prediction error = 0.82±0.02; *i-Phase*: $R^2 = 0.08\pm0.01$, prediction error = 1.41±0.03; *ISC*: $R^2 = 0.09\pm0.01$, prediction error = 1.41±0.02). The model with combined feature also achieved the smallest error and highest R^2 ($R^2 = 0.70\pm0.01$, prediction error = 0.78±0.02), with more contribution from *Power* (45.00%), and *i-Amplitude* (33.33%) (Fig. 4B). As listed in Table 2, the top 10 significant features mainly came from the *i-Amplitude* feature, with more than half of them from parietal-occipital areas. The predicted real-time ratings by different types of inter-brain features are shown in Figure 4C. Again, the models using *i-Amplitude* and *Power* features effectively followed the behavioral ratings.

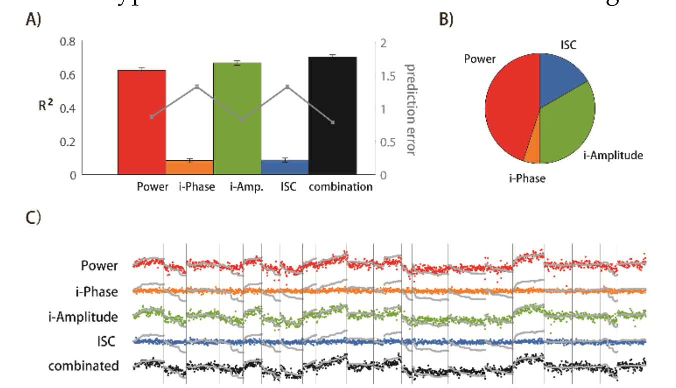

Fig. 4. Regression results for valence. (A) Model fitting R^2 (bars with the left scale) and model prediction error (gray lines with the right scale) of the regression using the four inter-brain features and their combination. Error bars indicate standard deviation. (B) The percentages of each type of features in the chosen 60 features in the combined regression model. (C) The predicted ratings using different types of features for all 20 video clips. The gray lines represent the behavioral ratings.

A complete overview of the contributions of each electrode and frequency band within each inter-brain feature is shown in Figure 5. The contributions were reflected by the beta coefficients from the regression models using the four inter-brain features. In general, *Power* and *i-Amplitude* (shown in the scale ranging from -1 to 1, see Fig. 5) had larger beta coefficients than *i-Phase* and *ISC* (shown in the scale ranging from -0.2 to 0.2, see Fig. 5). Distinct spectral and spatial patterns were observed for the four features. When regressing two emotional ratings and the *Power* feature, the beta band contributed the most, while the gamma band contributed the most in the regression of two emotional ratings and *i-Amplitude*, as well as the regression of arousal ratings and *Power*. For the rest regression analyses, the four bands contributed roughly similar. In general, a larger contribution of the higher frequency bands (beta and gamma) than the lower frequency bands was observed from the topographic result as well as the linear regression (Arousal: theta, $R^2 = 0.31\pm0.02$; alpha, $R^2 = 0.30\pm0.02$; beta, $R^2 = 0.45\pm0.02$; gamma, $R^2 = 0.51\pm0.02$. Valence: theta, $R^2 = 0.37\pm0.02$; alpha, $R^2 = 0.29\pm0.02$; beta, $R^2 = 0.55\pm0.02$; gamma, $R^2 = 0.57\pm0.02$).

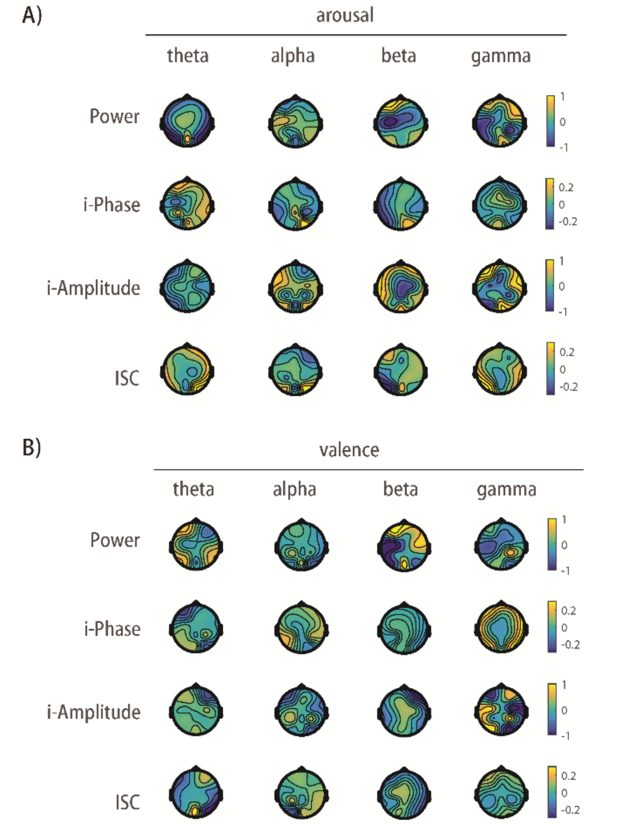

Fig. 5. The topographies of regression beta coefficients for the representation of the contributions of each electrode and frequency band. The beta coefficients were taken from regression models based on the four EEG features separately. The results for arousal and valence are shown on the top and bottom panels.

4.3 Inter-brain features vs. single-brain features

The regression R^2 for arousal and valence based on the individual-based single-brain amplitude (Fig. 6A) and power (Fig. 6B) features are plotted together with the results based on inter-brain features. The R^2 was significantly larger for the inter-brain features (stars in Fig. 6), as compared to those from all the single-brain features (circles in Fig. 6) by one-tail one-sample t-test, for both Power

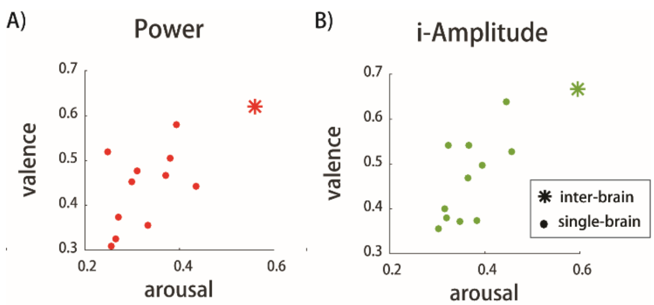

Fig. 6. Regression results for arousal and valence based on the single-brain *Power* (A) and *i-Amplitude* (B) features, each dot representing the results of one individual participant. The results from inter-brain features are also shown for comparison (shown as stars).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

(arousal: t(10)=-12.39, p<0.001; valence: t(10)=-7.09, p<0.001) and *i-Amplitude* (arousal: t(10)=-14.82, p<0.001; valence: t(10)=-7.24, p<0.001). Notably, inter-brain features outperformed the single-brain features from the best individual participant.

4.4 The number of participants

The possible influence of regression results by the number of participants was also explored. As shown in Figure 7, there were gradual increases of the regression R^2 with increasing number of participants, for both arousal (Fig. 7A) and valence (Fig. 7B). Overall, the increases of the regression R^2 were significant when the number of participants gradually increasing from 1 to 2, 2 to 3, …, 8 to 9, but the increase became non-significant, when the number of participants exceeded 10 (one-tail two-sample t-test, p>0.05, FDR corrected), for both *Power* and *i-Amplitude* features.

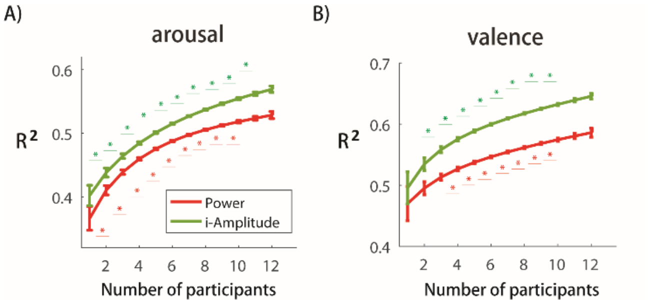

Fig. 7. Performance (regression R^2) as a function of participant number for arousal (A) and valence (B). Error bars indicate standard error of the mean derived from different samplings of the participants given a certain participant number. Stars indicate significant differences at p=0.05 level with FDR correction.

5 Discussion

The present study investigated the effectiveness of different inter-brain EEG features for implicitly tagging real-time emotions during video watching. Inter-brain EEG features were calculated in different frequency bands, and the possible contributions from amplitude and phase components were separately explored. The inter-brain amplitude consistency (*i-Amplitude*) and the averaged spectral power (*Power*) showed significantly better prediction performances when compared to inter-brain phase consistency (*i-Phase*) and inter-subject correlation (*ISC*). The regression results were further enhanced by combining all features. The significant electrodes and frequency bands were also investigated and reported.

In contrast to most previous EEG-based emotion tagging studies, the exploration of these inter-brain features in the present study was conducted on the basis of continuous behavioral ratings rather than stimulus-wise ratings. As shown in Figure 1B, the continuous ratings provided rich temporal information about human emotion experiences, and fast changes of the emotion magnitudes were observed. These data argued against the general assumption of a constant emotion magnitude throughout the emotional video clips, as used in most previous studies [14], [29]. Together with the high consistency of the real-time ratings across participants (Fig. 2A), these ratings were likely to be a better choice as the labels of the multimedia contents, especially for those with complex and mixed emotions.

By decomposing the inter-brain features into amplitude and phase components, we found the inter-brain amplitude consistency predicted the real-time emotion experiences very well, while the inter-brain phase-locking yielded much worse results. Such an observation may be quite surprising at first glance, as single-brain inter-trial or inter-regional phase-locking has been demonstrated as an effective indicator for a variety of cognitive functions [66]–[68]. The substantially stronger predictive power of amplitude rather than phase in the inter-brain context, however, may be explained by the high inter-participant temporal variability in emotion perception. Hereby, the EEG activities from different participants might not be precisely time locked as reflected by their phase angles. Instead, there should be still a certain degree of inter-brain 'synchronization' at a coarser time scale, revealed by the amplitude responses with phase information discarded. Nevertheless, phase responses also provided useful information, as the regression performance based on combined features were significantly better than using the amplitude feature alone. Whether our findings can be extended beyond the emotion tagging context remain to be elucidated.

Notably, the performance of the popular inter-brain *ISC* feature only reached a similar level as the inter-brain phase-locking features. *ISC* and *i-Phase* were calculated using quite different formulae: *ISC* captured the linear correlations between pairs of EEG activities, in which their phase synchrony might play an important role; *i-Phase* took the neural activities from all participants simultaneously. The similar poor performance of *ISC* and *i-Phase* may due to the common source of phase information being less functional in identifying continuous emotion. Indeed, similar spatial patterns of the bivariate correlation between the two features and the emotion ratings were observed (Fig. 2B).

The complete overview of the frequency band and spatial distribution revealed the major contribution of beta and gamma bands, especially for the regression of *Power* and *i-Amplitude*. The beta-power asymmetry at the parietal regions [69], and the gamma spectral changes [70], [71] has been reported as the indicators of emotional states in facial expression perception, while the frontal electrical activity was found to contribute much to musical emotions [72]. As video clips consisted of both visual and auditory emotion elements, it is reasonable to see a distributed brain region involvement. As the inter-brain features has rarely been investigated in previous studies, our findings provide possibly the first piece of evidence on the neural response patterns in the context of emotional experiences.

The inter-brain features outperformed the single-brain features from the best-performing participant (Fig. 6), suggesting its promising application potential for real-time emotion tagging. The collective nature of the inter-brain features could facilitate emotion tagging in continuous and complex stimulation scenarios, probably in the following manners. First, calculating the inter-brain features may serve to enhance the signal-to-noise ratio of the features, as all participants were perceiving the same audiovisual stimulation [73]. Hereby data from each individual brain were supposed to share similar stimulation related responses, with different random noises that could be suppressed by inter-brain integration. Second, the intra-participant variation in the participants' mental states, such as fatigue, distraction, etc., may deteriorate the quality of the single-brain neural data. Collecting neural activities from multiple participants may help to alleviate this problem, as the mental state variations are likely to be randomly distributed over time for different participants. Third and most importantly, the inter-brain features were constructed from a collective perspective, which was substantially different from the conventional single-brain features. Therefore, completely new information may be extracted by calculating the inter-brain features, with more brains providing more robust information. These explanations were preliminarily supported by our regression analyses as the function of the number of participants (Fig. 7). However, as the performance enhancement did not reach a significant level when having more than 10 participants, further investigations with more participants and more varied stimulation materials are needed to estimate the necessary number of participants precisely.

Although the difference in methodology (regression vs. classification) made it difficult to have a direct comparison between our results and those in previous studies, the high R^2 values (>0.6) and the low prediction errors (approx. 1 in the 9-point scales) imply a good potential for practical applications. Compared to the popular classification models (normally binary for high vs. low levels), our regression models can yield fine-scale emotional experiences in real-time, thereby providing a better description of human emotion experiences.

Admittedly, there was a possible confounding factor of time for the regression results for arousal, as there was a positive correlation between arousal and time in the majority of the video stimuli (Figure 2A). Controlling the factor of time in the regression analysis, however, would severely deteriorate the subjective reports on arousal, and inevitably but unnecessarily weaken the regression results. As our regression model also showed a good prediction performance for the few video clips with no increase of arousal with time (Figure 3C), such a confounding might have limited impact on our results. Nevertheless, studies with more diversified stimuli (i.e. video clips with decreasing arousal with time) are necessary to further clarify this issue.

Moreover, while the present study investigated the basic units of inter-brain features from the spectral and temporal perspectives, further explorations are needed, at least from the following perspectives. First, while we empirically chose a 1-second duration for data analysis, the optimal temporal resolution for recognizing emotion experiences remains to be examined. Delta band EEG activities, for instance, could be actively involved, if a longer duration were found to be effective in capturing the emotion dynamics. Second, the inter-brain features could be constructed using more complex and advanced strategies. For example, when classifying different emotion types, some advanced classification approaches have been proposed, including using local temporal correlation common spatial pattern to extract neural features from response- and stimulus-locked EEG signals [74], and using diverse ways to combine neural features, such as optimal linear combination of neural decision variables, neural majority decision rule [73], and weighted majority or extreme vote [73]–[75]. Third, investigations on discrete emotions beyond the valence-arousal model are necessary as well, to validate the generalization of the present findings. Explorations in these directions are necessary and are expected to further enhance the performance of real-time emotion tagging.

6 Conclusion

The present study for the first time systematically investigated the possible contribution of different inter-brain features for continuous emotion tagging. The inter-brain features were explored in both the spectral and temporal domain. The regression results suggest superior performance of the inter-brain amplitude feature and the combination of all inter-brain features outperformed the single-brain features from the best participant. Our results advocate for the usage of inter-brain features for emotion tagging and provide evidence to guide selection of inter-brain features.

Acknowledgment

This work is supported by National Key Research and Development Plan (2016YFB1001200), National Science Foundation of China (U1736220, 61725204), National Social Science Foundation of China (17ZDA323), MOE (Ministry of Education, China) Project of Humanities and Social Sciences (17YJA190017), and Tsinghua University Initiative Scientific Research Program (2014z21043). We thank Dr. Xiaoqin Wang for proofreading an earlier version of this manuscript.

References

[1] J. Shen, M. Wang, S. Yan, and X.-S. Hua, "Multimedia Tagging: Past, Present and Future," in *Proceedings of the 19th ACM International Conference on Multimedia*, 2011, pp. 639–640.

[2] M. Soleymani and M. Pantic, "Human-centered implicit tagging: Overview and perspectives," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, pp. 3304–3309, 2012.

[3] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," *9th Int. Symp. Comput. Music Model. Retr.*, no. June, pp. 19–22, 2012.

[4] M. K. Shan, F. F. Kuo, M. F. Chiang, and S. Y. Lee, "Emotion-based music recommendation by affinity discovery from film music," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7666–7674, 2009.

[5] M. Tkalčič, A. Odić, A. KoTkalšičir, and J. Tasič, "Affective labeling in a content-based recommender system for images," *IEEE Trans. Multimed.*, vol. 15, no. 2, pp. 391–400, 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

[6] S. J. Westerman and S. Kaur, "Supporting creative product/commercial design with computer-based image retrieval," in *Proceedings of the 14th European conference on Cognitive ergonomics*, 2007, vol. 250, p. 75.

[7] S. Cho and J. Lee, "A human-oriented image retrieval system using interactive genetic algorithm," *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans*, vol. 32, no. 3, pp. 452–458, 2002.

[8] S. Zhao, H. Yao, and X. Sun, "Video classification and recommendation based on affective analysis of viewers," *Neurocomputing*, vol. 119, pp. 101–110, 2013.

[9] C. H. Chan and G. J. F. Jones, "Affect-based indexing and retrieval of films," *Proc. 13th Annu. ACM Int. Conf. Multimed. - Multimed. '05*, p. 427, 2005.

[10] Y.-H. Yang and H. H. Chen, "Machine Recognition of Music Emotion," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, pp. 1–30, 2012.

[11] M. Wang, B. Ni, X. Hau, and T.-S. Chua, "Assistive tagging: A Survey of Multimedia Tagging with Human-Computer Joint Exploration," *ACM Comput. Surv.*, vol. 44, no. 4, pp. 1–24, 2012.

[12] Yao Wang, Zhu Liu, and Jin-Cheng Huang, "Multimedia content analysis-using both audio and visual clues," *IEEE Signal Process. Mag.*, vol. 17, no. 6, pp. 12–36, 2000.

[13] M.-K. M. Kim, M.-K. M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG.," *Comput. Math. Methods Med.*, vol. 2013, p. 573734, 2013.

[14] S. Koelstra and I. Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image Vis. Comput.*, vol. 31, no. 2, pp. 164–174, 2013.

[15] A. Yazdani, E. Skodras, N. Fakotakis, and T. Ebrahimi, "Multimedia content analysis for emotional characterization of music video clips," *Eurasip J. Image Video Process.*, vol. 2013, pp. 1–10, 2013.

[16] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals," *IEEE Trans. Affect. Comput.*, pp. 1–1, 2017.

[17] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, and J. Kissler, "Mapping the emotional face. How individual face parts contribute to successful emotion recognition," *PLoS One*, vol. 12, no. 5, pp. 1–15, 2017.

[18] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," *ICMI'03 Fifth Int. Conf. Multimodal Interfaces*, pp. 258–264, 2003.

[19] S. Vrochidis, I. Patras, and I. Kompatsiaris, "Gaze movement-driven random forests for query clustering in automatic video annotation," *Multimed. Tools Appl.*, vol. 76, no. 2, pp. 2861–2889, 2017.

[20] S. N. Hajimirza, M. J. Proulx, and E. Izquierdo, "Reading users' minds from their eyes: A method for implicit image annotation," *IEEE Trans. Multimed.*, vol. 14, no. 3 PART 2, pp. 805–815, 2012.

[21] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, "Affective Characterisation of Movie Scenes based on Content Analysis and Physiological Changes," *Int. J. Semant. Comput.*, vol. 3, no. 2, pp. 235–254, 2009.

[22] S. Chen, S. Wang, C. Wu, Z. Gao, X. Shi, and Q. Ji, "Implicit hybrid video emotion tagging by integrating video content and users' multiple physiological responses," in *Proceedings - International Conference on Pattern Recognition*, 2017, pp. 295–300.

[23] S. Koelstra, C. Mühl, and I. Patras, "EEG analysis for implicit tagging of video data," in *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, 2009.

[24] M. Soleymani, "Multimedia Implicit Tagging Using Eeg Signals," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on. IEEE*, 2013, vol. 231287, no. 231287.

[25] R. Gupta, K. ur Rehman Laghari, and T. H. Falk, "Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization," *Neurocomputing*, vol. 174, pp. 875–884, 2016.

[26] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, 2010.

[27] M. Soleymani and M. Pantic, "26 Multimedia Implicit Tagging," *Soc. Signal Process.*, p. 368, 2017.

[28] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[29] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, 2012.

[30] X. Hu *et al.*, "EEG Correlates of Ten Positive Emotions," *Front. Hum. Neurosci.*, vol. 11, no. January, 2017.

[31] K. R. Scherer, "What are emotions? and how can they be measured?," *Social Science Information*, vol. 44, no. 4. pp. 695–729, 2005.

[32] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach, "Intersubject Synchronization of Cortical Activity during Natural Vision," *Science (80-.).*, vol. 303, no. 5664, pp. 1634–1640, 2004.

[33] I. P. Jääskeläinen *et al.*, "Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing.," *Open Neuroimag. J.*, vol. 2, pp. 14–9, 2008.

[34] L. Nummenmaa, E. Glerean, M. Viinikainen, I. P. Jaaskelainen, R. Hari, and M. Sams, "Emotions promote social interaction by synchronizing brain activity across individuals," *Proc. Natl. Acad. Sci.*, vol. 109, no. 24, pp. 9599–9604, 2012.

[35] D. O. Bos, "EEG-based emotion recognition," *Influ. Vis. Audit. Stimuli*, pp. 1–17, 2006.

[36] T. Baumgartner, M. Esslen, and L. Jäncke, "From emotion perception to emotion experience: Emotions evoked by pictures and classical music," *Int. J. Psychophysiol.*, vol. 60, no. 1, pp. 34–43, 2006.

[37] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V Friesen, "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I.," *J. Pers. Soc. Psychol.*, vol. 58, no. 2, p. 330, 1990.

[38] C.-H. Han, J.-H. Lee, J.-H. Lim, Y.-W. Kim, and C.-H. Im,

"Global Electroencephalography Synchronization as a New Indicator for Tracking Emotional Changes of a Group of Individuals during Video Watching," *Front. Hum. Neurosci.*, vol. 11, 2017.

[39] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement?," *Front. Hum. Neurosci.*, vol. 6, no. May, pp. 1–9, 2012.

[40] M. Kawasaki, Y. Yamada, Y. Ushiku, E. Miyauchi, and Y. Yamaguchi, "Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction," *Sci. Rep.*, vol. 3, 2013.

[41] N. Sinha, T. Maszczyk, W. Zhang, J. Tan, and J. Dauwels, "EEG hyperscanning study of inter-brain synchrony during cooperative and competitive interaction," in *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, 2017, pp. 4813–4818.

[42] D. A. Bridwell, C. Roth, C. N. Gupta, and V. D. Calhoun, "Cortical response similarities predict which audiovisual clips individuals viewed, but are unrelated to clip preference," *PLoS One*, vol. 10, no. 6, 2015.

[43] E. Salinas and T. J. Sejnowski, "Correlated neuronal activity and the flow of neural information," *Nature Reviews Neuroscience*, vol. 2, no. 8. pp. 539–550, 2001.

[44] P. G. Schyns, G. Thut, and J. Gross, "Cracking the code of oscillatory activity," *PLoS Biol.*, vol. 9, no. 5, 2011.

[45] D. Zhang, B. Hong, S. Gao, and B. Röder, "Exploring the temporal dynamics of sustained and transient spatial attention using steady-state visual evoked potentials," *Exp. Brain Res.*, vol. 235, no. 5, pp. 1575–1591, 2017.

[46] O. David, L. Harrison, and K. J. Friston, "Modelling event-related responses in the brain," *Neuroimage*, vol. 25, no. 3, pp. 756–770, 2005.

[47] P. Sauseng, W. Klimesch, W. R. Gruber, S. Hanslmayr, R. Freunberger, and M. Doppelmayr, "Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion," *Neuroscience*, vol. 146, no. 4. pp. 1435–1444, 2007.

[48] P. Sauseng and W. Klimesch, "What does phase information of oscillatory brain activity tell us about cognitive processes?," *Neuroscience and Biobehavioral Reviews*, vol. 32, no. 5. pp. 1001–1013, 2008.

[49] W. Klimesch, P. Sauseng, and S. Hanslmayr, "EEG alpha oscillations: The inhibition-timing hypothesis," *Brain Research Reviews*, vol. 53, no. 1. pp. 63–88, 2007.

[50] K. E. Mathewson, A. Lleras, D. M. Beck, M. Fabiani, T. Ro, and G. Gratton, "Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing," *Frontiers in Psychology*, vol. 2, no. MAY. 2011.

[51] A. Pérez, M. Carreiras, and J. A. Duñabeitia, "Brain-To-brain entrainment: EEG interbrain synchronization while speaking and listening," *Sci. Rep.*, vol. 7, no. 1, 2017.

[52] J. Jahng, J. D. Kralik, D. U. Hwang, and J. Jeong, "Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game," *Neuroimage*, vol. 157, pp. 263–274, 2017.

[53] C. Szymanski *et al.*, "Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation," *Neuroimage*, vol. 152, pp. 425–436, 2017.

[54] N. Jatupaiboon, S. Pan-Ngum, and P. Israsena, "Real-time EEG-based happiness detection system," *Sci. World J.*, vol. 2013, 2013.

[55] V. H. Anh, M. N. Van, B. B. Ha, and T. H. Quyet, "A real-time model based Support Vector Machine for emotion recognition through EEG," in *2012 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2012, pp. 191–196.

[56] Y. P. Lin, Y. H. Yang, and T. P. Jung, "Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening," *Front. Neurosci.*, no. 8 MAY, 2014.

[57] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, 2016.

[58] N. Thammasan, K. Moriyama, K. I. Fukui, and M. Numao, "Continuous music-emotion recognition based on electroencephalogram," *IEICE Trans. Inf. Syst.*, vol. E99D, no. 4, pp. 1234–1241, 2016.

[59] D. H. Brainard, "The Psychophysics Toolbox," *Spat. Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

[60] S. Koelstra, C. Muhl, and M. Soleymani, "Deap: A database for emotion analysis; using physiological signals," *IEEE Trans.*, vol. 3, no. 1, pp. 1–15, 2012.

[61] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement?," *Front. Hum. Neurosci.*, vol. 6, 2012.

[62] W. T. Chang *et al.*, "Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing," *Neuroimage*, vol. 114, pp. 49–56, 2015.

[63] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, 2003.

[64] C. M. Michel and M. M. Murray, "Towards the utilization of EEG as a brain imaging tool," *NeuroImage*, vol. 61, no. 2. pp. 371–385, 2012.

[65] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Stat.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[66] C. E. Schroeder, P. Lakatos, Y. Kajikawa, S. Partan, and A. Puce, "Neuronal oscillations and visual amplification of speech," *Trends Cogn. Sci.*, vol. 12, no. 3, pp. 106–113, 2008.

[67] M. Siegel, T. H. Donner, and A. K. Engel, "Spectral fingerprints of large-scale neuronal interactions," *Nature Reviews Neuroscience*, vol. 13, no. 2. pp. 121–134, 2012.

[68] A. M. Bastos and J.-M. Schoffelen, "A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls," *Front. Syst. Neurosci.*, vol. 9, 2016.

[69] D. J. L. G. Schutter, P. Putman, E. Hermans, and J. Van Honk, "Parietal electroencephalogram beta asymmetry and selective attention to angry facial expressions in healthy human subjects," *Neurosci. Lett.*, vol. 314, no. 1–2, pp. 13–16,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2018.2849758, IEEE Transactions on Affective Computing

2001.

[70] M. Balconi and C. Lucchiari, "Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis," *Int. J. Psychophysiol.*, vol. 67, no. 1, pp. 41–46, 2008.

[71] M. Li and B. L. Lu, "Emotion classification based on gamma-band EEG," in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 2009, pp. 1323–1326.

[72] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cogn. Emot.*, vol. 15, no. 4, pp. 487–500, 2001.

[73] M. P. Eckstein *et al.*, "Neural decoding of collective wisdom with multi-brain computing," *Neuroimage*, vol. 59, no. 1, pp. 94–108, 2012.

[74] D. Valeriani, R. Poli, and C. Cinel, "Enhancement of group perception via a collaborative brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 6, pp. 1238–1248, 2017.

[75] Y. Wang and T.-P. Jung, "A Collaborative Brain-Computer Interface for Improving Human Performance," *PLoS One*, vol. 6, no. 5, p. e20422, 2011.

Yue Ding is a Ph.D. candidate in School of Medicine, Tsinghua University, Beijing, China. She received a B.E. degree in Biomedical Engineering from Dalian University of Technology and Science in 2010. Her research interests include neuroeconomics, and the neural mechanism of auditory emotion and memory.

Xin Hu is a Ph.D. student in Department of Psychology, Tsinghua University. She received an LL.B. degree in International Politics from University of International Relations in 2014 and an M.Ed degree in Psychology from Tsinghua Univeristy in 2017. Her research interests include the neural mechanism of emotion and neuromarketing.

Zhenyi Xia is a game designer in Netease Games. He received a bachelor degree in Psychology from Tsinghua University in 2015 and an M.P.S degree in Information Science from Cornell University in 2017. His research interest focuses on neural mechanism of emotion in digital media and entertainment.

Yong-Jin Liu is an Associate Professor with the TNList, Department of Computer Science and Technology, Tsinghua University. He received the B.Eng degree from the Tianjin University, China, in 1998, and the Ph.D degree from the Hong Kong University of Science and Technology, Hong Kong, China, in 2004. His research interests include computational geometry, computer graphics, pattern analysis, and computer-aided design. He is a senior member of the IEEE.

Dan Zhang is an Associate Professor at Department of Psychology, Tsinghua University, Beijing, China. He received a B.E. degree in automation in 2005 and a Ph.D. degree in biomedical engineering in 2011, both from Tsinghua University. He was a postdoctoral fellow in School of Medicine, Tsinghua University from 2011 to 2013. His research interests include social neuroscience, engineering psychology, and brain-computer interfaces. He is a member of the IEEE.