Investigating Emotion Dynamics and Its Association with Multimedia Content During Video Watching Through EEG Microstate Analysis

Wanrou Hu [a,b,1], Zhiguo Zhang [a,b,c,d,1], Huilin Zhao [a,b], Li Zhang [a,b], Linling Li [a,b],

Gan Huang [a,b], and Zhen Liang [a,b]

[a] School of Biomedical Engineering, Health Science Center, Shenzhen University,
Shenzhen 518060, China
[b] Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging,
Shenzhen 518060, China
[c] Peng Cheng Laboratory, Shenzhen 518055, China
[d] Marshall Laboratory of Biomedical Engineering, Shenzhen 518060, China

E-mails: huwanrou2019@email.szu.edu.cn, zgzhang@szu.edu.cn, 2018222041@email.szu.edu.cn, lzhang@szu.edu.cn, lilinling@szu.edu.cn, huanggan@szu.edu.cn, and janezliang@szu.edu.cn

Abstract

Emotions dynamically change in response to ever-changing environments. It is of great importance, both clinically and scientifically, to investigate the neural representation and evoking mechanism of emotion dynamics. But, there are many unknown places in this stream of research, such as consistent and conclusive findings are still lacking. In this work, we perform an in-depth investigation of emotion dynamics under a video-watching task by gauging the dynamic associations among evoked emotions, electroencephalography (EEG) responses, and multimedia stimulation. Here, we introduce EEG microstate analysis to study emotional EEG signals, which provides a spatial-temporal neural representation of emotion dynamics. To investigate the temporal characteristics of evoking emotions during video watching with its neural mechanism, we conduct two studies from the perspective of EEG microstates. In Study 1, the dynamic microstate activities under different emotion states and emotion levels are explored to identify EEG spatial-temporal correlates of emotion dynamics. In Study 2, the stimulation effects of multimedia content (visual and audio) on EEG microstate activities are examined to learn about the involved affective information and investigate the emotion-evoking mechanism. The results show that emotion dynamics could be well reflected by four EEG microstates (MS1, MS2, MS3, and MS4). Specifically, emotion tasks lead to an increase in MS2 and MS4 coverage but a decrease in MS3 coverage, duration, and occurrence. Meanwhile, there exists a negative association between valence and MS4 occurrence as well as a

[1] Equal contributions.

positive association between arousal and MS3 coverage and occurrence. Further, we find that MS4 and MS3 activities are significantly affected by visual and audio content, respectively. In this work, we verify the possibility to reveal emotion dynamics through EEG microstate analysis from sensory and stimulation dimensions, where EEG microstate features are found to be highly correlated to different emotion states (emotion task effect and level effect) and different affective information involved in the multimedia content (visual and audio). Our work deepens the understanding of the neural representation and evoking mechanism of emotion dynamics, which can be beneficial for future development in the applications of emotion decoding and regulation.

Keywords: Emotion dynamics; electroencephalography; microstate analysis; stimulation effect; video evoking.

1 Introduction

Emotion is a complex and dynamic physiological and psychological interaction that adaptively responds to internal and external changes (Scherer, 2009). As an essential component of subjective experience, emotion plays a crucial role in mental health maintenance and daily task manipulation, such as decision making, interaction, perception. However, due to the diversity and volatility of emotions (Shen and Lin, 2019), it is still a great challenge to fully understand the complex neural mechanism underlying emotion perception. In the existing emotion studies (Koelstra *et al.*, 2012; Zheng and Lu, 2015; Katsigiannis and Ramzan, 2018), videos are widely used for emotion induction in a laboratory environment, and the simultaneous brain responses during video watching are collected for data analysis and feature extraction. Classification models are built to map the electroencephalography (EEG) features to the given single emotion label for one video. Few studies concern the changing emotions in the time-variant brain responses during watching a video. In other words, current studies are still limited in the analysis and interpretation of emotion dynamics. Important questions, such as how emotions change dynamically during video watching and how the brain responds dynamically to the external video stimulation during emotion induction, are still open. To further investigate the brain activities of emotion dynamics and its association with multimedia content during video watching, in this work, we will focus on the exploration of the underlying relationship among the evoked emotion states, dynamic brain electrophysiological responses, and multimedia stimulation content.

EEG (Jenke *et al.*, 2014) provides an efficient and effective tool to investigate emotion-related brain dynamics, benefitting from its high time resolution, easy operation, low cost, and high portability (Alarcão and Fonseca, 2019; Hu and Zhang, 2019; Hu *et al.*, 2019). EEG data offer a more objective and alternative approach to characterize the diversity and variability of emotions than subjective self-reporting and external behaviors such as facial expression, gesture, and tone (Liu *et al.*, 2018; Shu *et al.*, 2018). Recently, EEG-based emotion analysis has achieved remarkable performance in both emotion mechanism exploration and practical applications. For example, Zheng *et al.* (Zheng *et al.*, 2019) found that distinctive spatial patterns of EEG activities were observed under different emotion states (positive, negative, and neutral). Their cross-day experiments consistently showed that high-frequency EEG oscillations in the beta and gamma bands over the lateral temporal region were activated under positive emotions. A higher alpha-band EEG response was observed in the parietal and occipital regions for neutral emotions. Also, Liu *et al.* (Liu *et al.*, 2019) found that EEG activities in the frontal and temporal lobes contributed to multimedia-evoked emotion representation through investigating the time-varying functional connectivity among 62-electrode EEG signals. These dynamic connectivity topographies exhibited the small-world property (Sporns, 2011) and covered both local and global spatial information of brain activities during emotion induction. However, due to the complex and non-stationary nature of EEG oscillations, the existing EEG analysis methods may fail to precisely characterize the dynamic spatial-temporal characteristics of spontaneous brain activities over the whole brain scalp under different emotion states.

EEG microstate analysis is an effective method for brain dynamics analysis, which is capable of characterizing the topographical distributions of instantaneous scalp electric potentials and reflecting the whole-brain EEG dynamics along the time (Koenig *et al.*, 2002). There exists a close electrophysiological relevance between EEG microstates (which mainly refer to four explored canonical microstates, termed as **MS1, MS2, MS3, MS4** below) and specific large-scale functional brain networks (e.g., the **auditory, visual, default mode, and dorsal attention networks**). Hence, microstate analysis has been demonstrated as a promising EEG analysis method for functional brain state estimation (Britz *et al.*, 2010; Musso *et al.*, 2010; Yuan *et al.*, 2012). Gianotti *et al.* (Gianotti *et al.*, 2008) adopted EEG microstate analysis to explore the temporal dynamics of emotion-related event-related potential (ERP). Their work was the first to introduce EEG microstate analysis into affective computing study and showed the possibility for valid emotion-related dynamic EEG microstate detection. Besides, Shen *et al.* (Shen *et al.*, 2020) discovered that EEG microstate features were discriminative for emotion recognition modeling. In Shen *et al.*'s work, traditional microstate features were extracted based on the public microstate templates, and random forest regression was used for binary emotion classification. These two works consistently show that EEG microstate analysis is an effective method for EEG-based emotion dynamics investigation.

On the other hand, multimedia content with rich audiovisual information offers realistic and vivid scenarios for evoking various human emotions. In current researches, multimedia content is commonly used as emotion induction materials in electrophysiological experiments (Zheng and Lu, 2015; Zheng *et al.*, 2019) or treated as the feature input in content-based affective modeling (Koelstra *et al.*, 2012; Horikawa and Kamitani, 2017). Few studies have comprehensively examined the relationships among evoked emotion states, dynamic brain electrophysiological responses, and multimedia stimulation content. An exploration of the underlying associations among these three aspects will be greatly beneficial for obtaining informative and reliable affective information to characterize the neural mechanism of emotion dynamics.

The present work investigates the associations among dynamic emotion states, dynamic brain responses, and multimedia stimulation content by using EEG microstate analysis as the main tool. The dynamic changes of emotions during video watching will be analyzed. The corresponding EEG signals will be acquired and analyzed to identify spatial-temporal EEG microstate patterns related to emotion dynamics. Further, the association between multimedia stimulation content and emotional EEG responses will be separately characterized on visual and audio content. In summary, two studies will be conducted as follows (Fig. 1).

- **Study 1: Emotion dynamics analysis (Section 2.3)**. EEG microstate dynamics will be first characterized to describe spatial-temporal changes of EEG activities under various emotion states during video watching. Subsequently, the activation patterns of EEG microstate dynamics will be analyzed to characterize emotion-related neural dynamics from three perspectives. (1) **Task effect** (Section 2.3.1): to quantify the differences of EEG microstate activities between pre-stimulus and post-stimulus stages. (2) **Level effect** (Section 2.3.2): to measure the differences of EEG microstate activities between low-level and high-level groups on different emotion dimensions. (3) **Evoking dynamics** (Section 2.3.3): to evaluate the temporal dynamics of emotion processing in the brain, in which the moment-to-moment changes of EEG microstate activities during video watching will be examined.

- **Study 2: Multimedia stimulation effect analysis (Section 2.4)**. The multimedia content used for emotion induction will be described as low-level attribute features and high-level semantic features in terms of **visual content** (Section 2.4.1) and **audio content** (Section 2.4.2). The corresponding associations between visual/audio content and emotional EEG responses will be measured. The **timing effect** (Section 2.4.3) of visual and audio content on EEG dynamics will be separately studied, and the temporal correlations between the changing multimedia content and dynamic EEG microstate activities will be examined.

Two main contributions of this work are summarized as follows. (1) The relationships among emotion states, dynamic brain responses in terms of EEG microstate dynamics, and multimedia stimulation content are investigated to reveal the neural mechanism of emotion dynamics. (2) Emotion-related EEG dynamics patterns across different subjects (32 subjects), experimental conditions (pre-stimulus, video-stimulated, and post-stimulus stages), and brain states (40 videos) are fully investigated by an efficient EEG microstate analysis pipeline. Two main observations of this work are found as follows. (1) An emotion-evoked brain activity difference is observed on different emotion dimensions. It is found that arousal changes mainly affect MS3 activities and valence states are mainly related to MS4 activities. (2) A stimulation perception difference is observed in dynamic EEG microstate features. Visual content mainly leads to the changes in MS4 activities and audio content is mainly related to the changes in MS3. Overall, by examining the dynamic characteristics of EEG during emotion induction, this work helps us to gain a deeper insight into the neural processing mechanism of emotion dynamics.

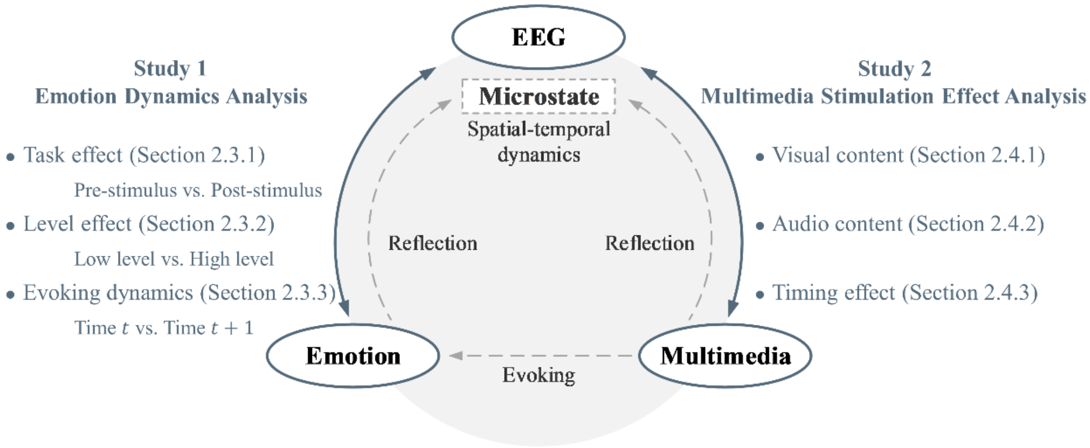

EEG
Microstate
Spatial-temporal dynamics
Study 1
Emotion Dynamics Analysis
• Task effect (Section 2.3.1)
Pre-stimulus vs. Post-stimulus
• Level effect (Section 2.3.2)
Low level vs. High level
• Evoking dynamics (Section 2.3.3)
Time t vs. Time $t+1$
Study 2
Multimedia Stimulation Effect Analysis
• Visual content (Section 2.4.1)
• Audio content (Section 2.4.2)
• Timing effect (Section 2.4.3)
Reflection
Reflection
Emotion
Multimedia
Evoking

Fig. 1 The framework of multimedia-evoked emotional EEG dynamics analysis. In Study 1, based on EEG microstate analysis, the relationships among evoked emotion states, dynamic EEG activities, and multimedia stimulation content are characterized. The emotion-related EEG dynamics analysis is conducted under the exploration of task effect, level effect, and evoking dynamics. In Study 2, the multimedia stimulation effect on emotion-related EEG microstate dynamics is conducted, and the timing effect of visual and audio content (content-based temporal shift) is evaluated.

2 Methods

2.1 EEG data and preprocessing

A well-known DEAP database (a database for emotion analysis using physiological signals) (Koelstra *et al.*, 2012) with 32 subjects' EEG recordings during video watching is used for emotion dynamics investigation. In this database, 40 emotional music videos (corresponding to 40 trials below) with a fixed length of 60 s were randomly presented for emotion induction. Simultaneously, 32-electrode EEG signals were recorded at a sampling rate of 512 Hz. To investigate the emotion-related brain dynamics from the recorded EEG data, we first perform a standard EEG preprocessing procedure (including filtering, common average re-reference, and independent component analysis) for noise removal and signal quality enhancement. Then, the preprocessed EEG data are divided into **pre-stimulus** (3 s), **video-stimulated** (60 s), and **post-stimulus** (3 s) segments for further analysis. More details about the DEAP database and EEG preprocessing procedure are provided in Appendix I of the Supplementary Materials.

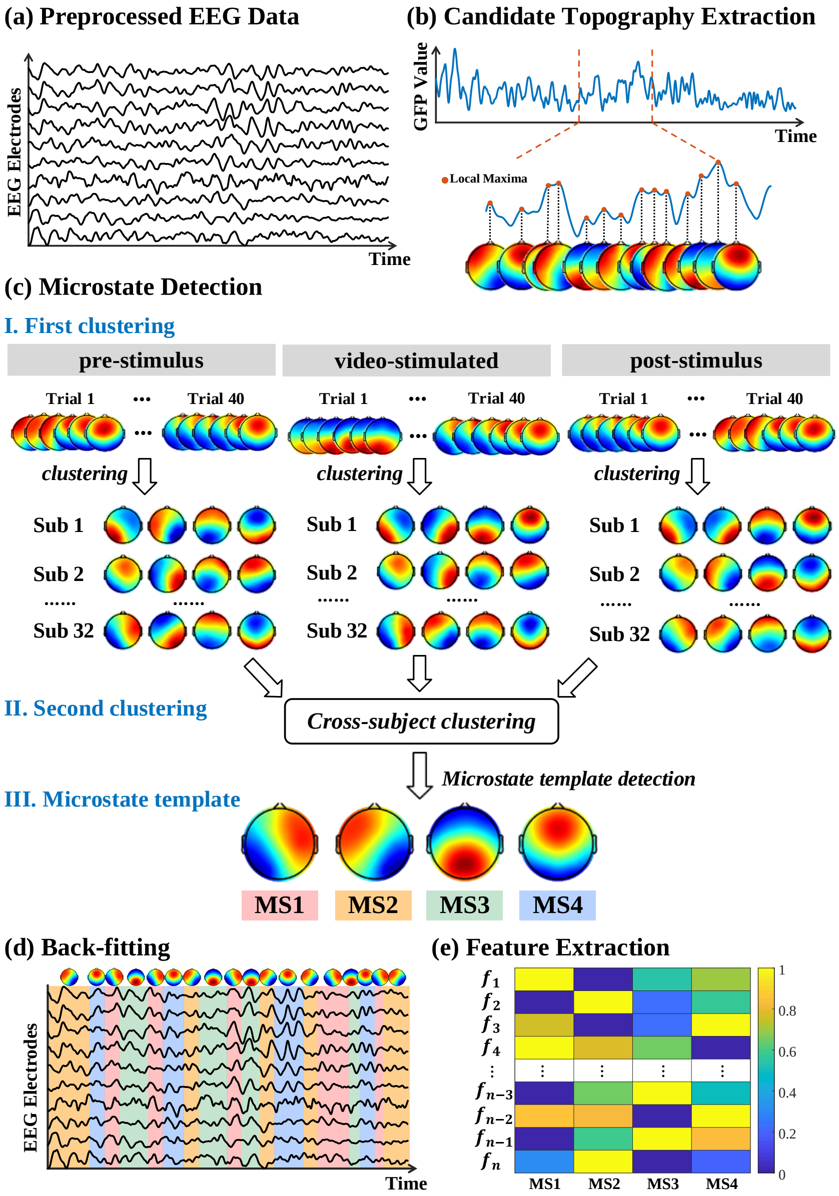


Fig. 2 A standard procedure of EEG microstate analysis. Based on (a) the preprocessed EEG data, (b) candidate topographies with high signal-to-noise ratios are extracted from the local maxima of the GFP curve. (c) A sequential microstate clustering is conducted for representative and reliable EEG microstate detection, where the first clustering is performed at a within-subject level and the second clustering is at a cross-subject level. (d) Based on the detected EEG microstate templates, EEG signals are then re-represented into EEG microstate sequences by assigning each time point to one predominant microstate. (e) A series of microstate features are calculated for quantitative measurement, including duration, occurrence, coverage, and transition probability, named as $\{f_1,\ldots,f_k,\ldots,f_n\}$. Here, f_k refers to the kth extracted microstate feature and n is the total feature number.

2.2 EEG microstate analysis

In this section, we will first present a standard EEG microstate analysis. In the EEG microstate analysis, microstate detection is a fundamental and crucial part, which will directly influence the validity and reliability of the analysis performance. However, the current microstate detection methods mainly focus on the resting-state and may fail to be adaptive enough for the task-state. In this work, we will then introduce a sequential microstate clustering analysis for efficient and representative EEG microstate template detection.

A standardized EEG microstate analysis is implemented as shown in Fig. 2, including candidate topography extraction, microstate detection, back-fitting, and feature extraction (Pascual-Marqui *et al.*, 1995). EEG microstate analysis starts with a bottom-up process (Fig. 2 (b) and (c)), in which the representative EEG microstate templates are first detected from the spontaneous EEG signals. Then, a top-down process termed back-fitting (Fig. 2 (d)) is conducted to re-represent EEG data into a series of dynamic microstate sequences, by calculating the spatial similarity with the identified microstate templates and assigning the original EEG sample points to the microstate with the highest similarity (Lehmann *et al.*, 2005; Zanesco *et al.*, 2020). Here, a global map dissimilarity (GMD) (Murray *et al.*, 2008) is used as a criterion for microstate class assignment, where the spatial correlation between original EEG topographies at each sample point and microstate templates is measured. Then, a temporal smoothing process (Poulsen *et al.*, 2018) is adopted to reject the noisy time segments, and the interrupted microstate segments shorter than 30 ms would be re-assigned to another microstate based on GMD calculations. Finally, the corresponding microstate features are extracted to quantify the dynamic changes of EEG microstate activities (Fig. 2 (e)) and represent the spatial-temporal oscillations of brain activities during video watching.

To improve the representative microstate template detection for emotion-related EEG dynamics analysis, a sequential microstate detection with two-step spatial clustering is introduced (Fig. 2 (c)). The first clustering is implemented at the within-subject level to extract subject-representative microstate topographies. Here, based on 40 trials of EEG data from every single subject, the candidate topographies under three experimental conditions are separately extracted from the local maximum of global field power (GFP) (Lehmann and Skrandies, 1980; Skrandies, 1989) and then output into a modified k-means clustering algorithm (Pascual-Marqui *et al.*, 1995) with a cluster number c ranging from 2 to 8 and an iteration number I of 1000. Within the iteration of spatial clustering, the cluster centroids with high global explained variance (GEV) and low cross-validation (CV) criterion value are identified and extracted as subject-representative microstate topographies. Based on the extracted subject-representative microstate topographies, the second clustering is conducted at a cross-subject level for subject-independent EEG microstate detection under similar parameter settings. The EEG topographies with high spatial similarity are identified as the final EEG microstates for emotion-related neural dynamics analysis. The obtained EEG microstate templates (named as MS1, MS2, MS3, and MS4) show a low variance of residual noise (Pascual-Marqui *et al.*, 1995; Murray *et al.*, 2008) and a good tolerance for individual differences (Michel and Koenig, 2018; D’Croz-Baron *et al.*, 2021) in emotion-evoked EEG dynamics analysis across 32 subjects and 40 trials. After back-fitting using the detected EEG microstates, a series of EEG microstate time sequences are obtained. Four commonly used EEG microstate features are extracted below for further investigation of emotion-related neural mechanisms.

- **Duration:** the average time span that a specific microstate remains dominant, which can reflect the stability of the underlying neural configuration during emotion induction (Khanna *et al.*, 2015).
- **Occurrence:** the times of presentation per second that a specific EEG microstate remains dominant, which indicates the representation tendency of the underlying neural activation (Koenig *et al.*, 2002).
- **Coverage:** the ratio of the period that a specific microstate keeps dominant to the total recording time (Seitzman *et al.*, 2017).
- **Transition probability (TP)**: the transition percentage between any two EEG microstates, which estimates the sequential activation tendency of scalp electric potentials on a millisecond time scale (Koenig *et al.*, 2005; Khanna *et al.*, 2015).

2.3 Study 1: Emotion dynamic analysis

In Study 1, we will investigate the representation differences of emotion-evoked EEG microstate activities from the perspectives of (1) task effect, (2) level effect, and (3) evoking dynamics.

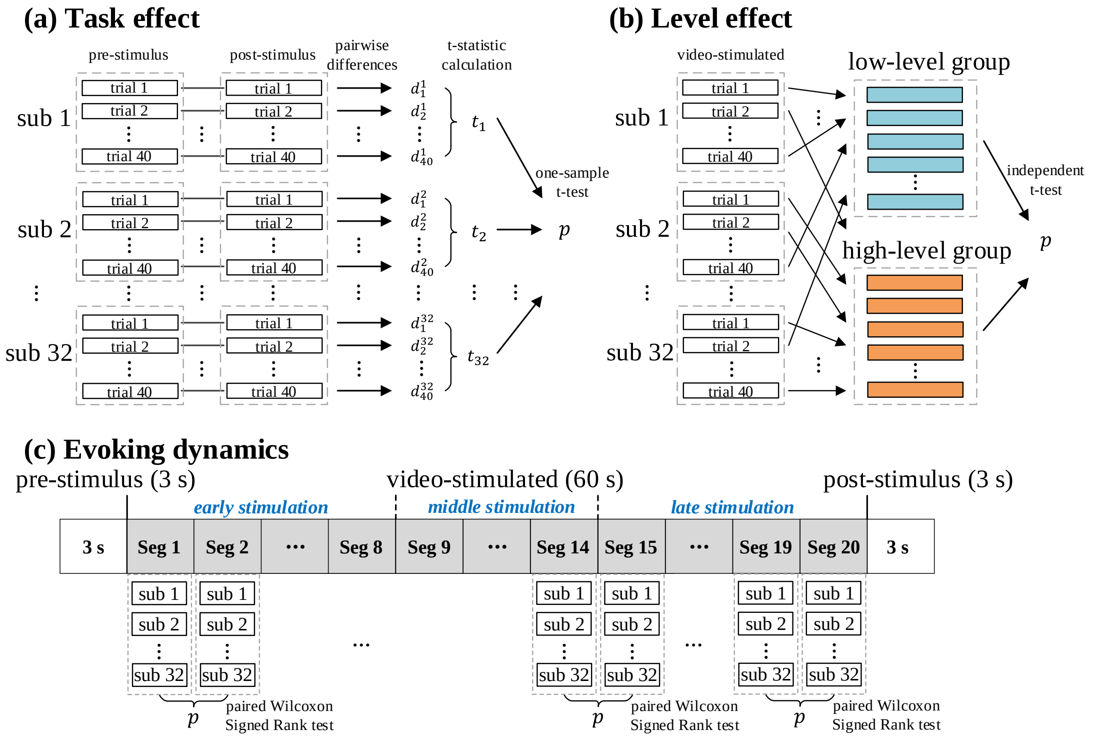


Fig. 3 The statistical analysis processes of emotion dynamics analysis in terms of (a) task effect, (b) level effect, and (c) evoking dynamics.

2.3.1 Task effect

The emotion task effect on EEG microstate activities is measured as the pairwise statistical differences between pre-stimulus and post-stimulus stages (Fig. 3 (a)). First, for each subject, the representation differences of EEG microstate activities before and after emotion-evoking tasks are measured as d_j^i ($i \in [1,32]$ is the subject number and $j \in [1,40]$ is the trial number). The pairwise differences between pre-stimulus and post-stimulus stages across 40 trials are calculated in terms of each microstate feature and then output for t-statistic calculation. A t-value, t_j ($j \in [1,40]$), is calculated as the statistical measurement of subject-specific representation differences in one type of microstate feature. Thus, for one microstate feature, a list of t-values from 32 subjects is obtained. Second, to evaluate the emotion-evoking task effect from a cross-subject perspective, a one-sample t-test is implemented on the obtained list of t-values for one microstate feature and a p-value is measured. Third, to minimize the influence of type I errors, all the obtained p-values are corrected for multiple comparisons using the false discovery rate (FDR) with a significance level of 5% ($p < 0.05$).

2.3.2 Level effect

The emotion level effect characterizes the EEG microstate differences at different levels (high or low) of emotion dimensions (valence and arousal). For each emotion dimension, we first divide the emotion-evoked EEG data into low- and high-level emotional groups according to the self-assessment ratings from the 32 subjects. Here, the threshold for level grouping is defined by a self-adaptive method (Yin *et al.*, 2017) that is presented in detail in Appendix II of the Supplementary Materials. Second, the emotion level effect is measured as the representation difference between low- and high-level groups in terms of each microstate feature (Fig. 3 (b)). As the emotion dynamics on different emotion dimensions are independent, the evaluation of the emotion level effect is separately measured on valence and arousal dimensions. Specifically, based on the distribution estimation of a Lilliefors test, an independent t-test (for normally distributed groups) or Wilcoxon Rank Sum test (for non-normally distributed groups) is conducted for statistical analysis on each microstate feature and the inter-group differences across trials and subjects are measured. The obtained p-values reveal the statistical evaluation of emotion level effect on EEG microstate activities.

2.3.3 Evoking dynamics

In emotion-evoking experiments, continuous presentation of multimedia stimulation dynamically influences brain activities during stimulation perception and processing (Zheng and Lu, 2015). An investigation of temporal variations in multimedia-evoked EEG activities helps to understand the dynamic characteristics of emotion processing in the brain. In this study, we measure the temporal dynamics of EEG activities under each trial of emotion-evoking tasks in terms of each microstate feature as follows (Fig. 3 (c)). **(1) Data segmentation.** The video-stimulated EEG data at one trial is divided into a number of short segments with a fixed length of 3 s. Total 20 segments are obtained (video length 60 s / segment length 3 s = 20 segments). There is no overlap between any two adjacent segments. For clarity, the first eight segments are named as the early stimulation stage (1~24 s), the following six segments are marked as the middle stimulation stage (25~42 s), and the last six segments are assigned to the late stimulation stage (43~60 s). **(2) Segment-based feature extraction.** Four types of microstate features (duration, occurrence, coverage, and TP) are extracted from each segment. **(3) Baseline correction.** To obtain a better estimation of emotion dynamics and minimize the emotion-unrelated effect, a baseline correction is employed by normalizing the segment-based microstate features with the corresponding features extracted from the pre-stimulus stage (-3 to 0 s). **(4) Statistical measurement.** For each trial, a paired Wilcoxon Signed Rank test (non-parametric statistical analysis based on the normality estimation of Lilliefors test) is conducted on any two adjacent segments across 32 subjects in terms of each microstate feature, and 19 pairs (total 20 segments for each trial) of segment-based statistical differences are obtained to represent the video-specific temporal characteristics of emotion processing in the brain.

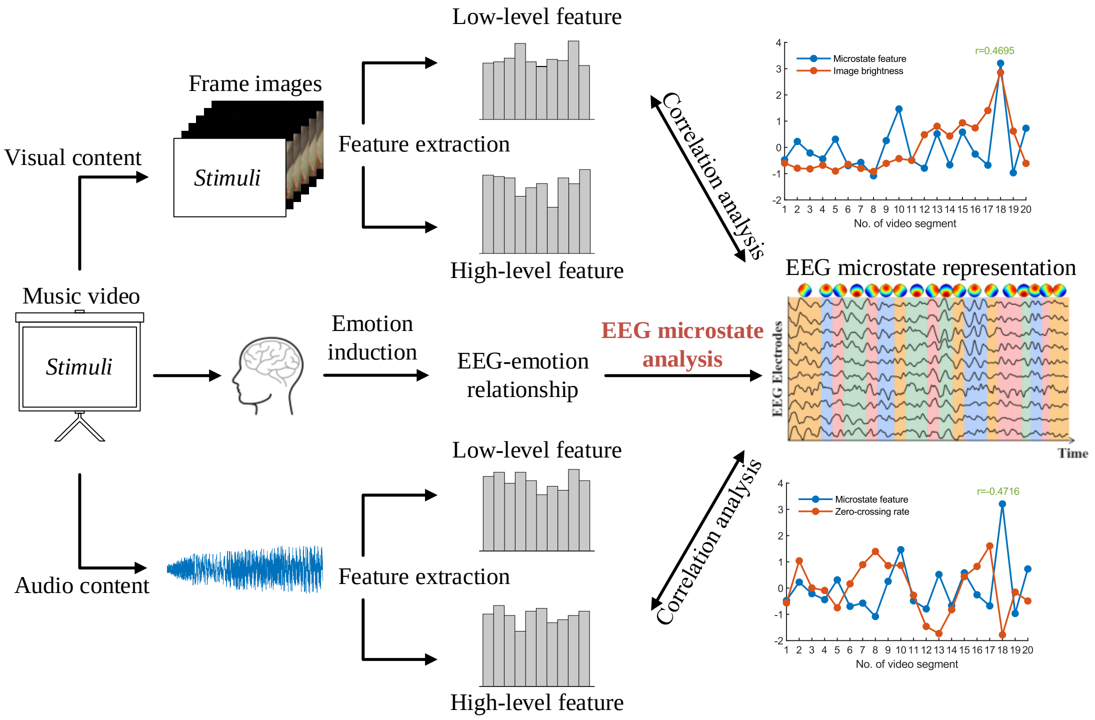
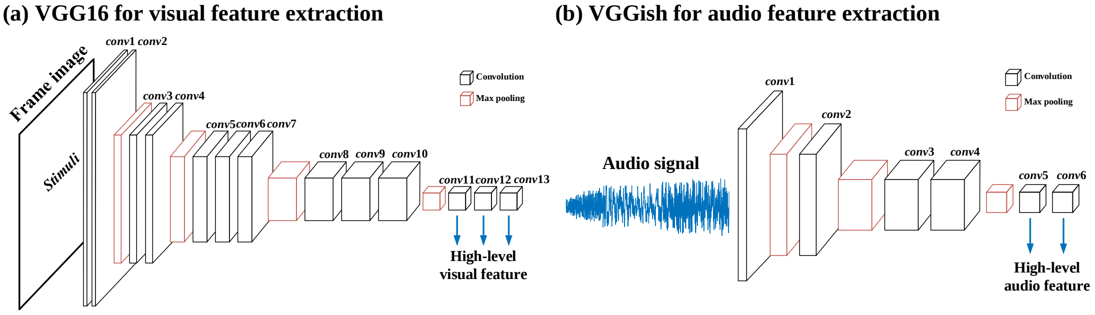


Fig. 4 A general flowchart for correlation detection among evoked emotion states, EEG microstate activities, and multimedia content of visual and audio.

2.4 Study 2: Multimedia stimulation effect analysis

As shown in Fig. 4, the stimulation effects of multimedia on emotional EEG dynamics are analyzed in terms of visual and audio content, respectively. EEG microstates function as an intermediary between multimedia stimulation content and evoked emotions for neural mechanism investigation. The associations among evoked emotions, dynamic brain responses, and multimedia stimulation content are explored, and how the brain perceives stimulation for emotion induction during video watching is studied.


(a) VGG16 for visual feature extraction

(b) VGGish for audio feature extraction

Fig. 5 The pre-trained deep convolutional neural networks (VGG16 and VGGish) for high-level visual and audio feature extraction. For visual content, the features extracted from *conv*11 to *conv*13 in VGG16 are termed high-level visual features. For audio content, the features extracted from *conv*5 and *conv*6 in VGGish are termed high-level audio features.

2.4.1 Visual content

Both low- and high-level visual features are extracted to examine the stimulation effect of visual content on EEG microstate activities during emotion induction. In this study, brightness is measured as a low-level visual feature, since it is a fundamental visual characteristic that is important for emotion induction (Itti *et al.*, 1998). For low-level visual feature extraction, frame-based visual brightness is first extracted and then converted to a segment-based visual feature through an average calculation. The changes of segment-based visual features characterize the temporal variations of visual content in multimedia stimulation, which provides important evoking clues for emotion dynamics analysis.

For high-level visual feature extraction, a pre-trained deep convolutional neural network, VGG16 (Simonyan and Zisserman, 2014), is utilized, in which each convolutional layer functions as an automated feature extractor. With an increase of convolutional layers, deeper visual features with high-level semantic information are learned. To extract high-level visual features for emotion dynamics analysis (Fig. 5 (a)), video frames are sequentially input into VGG16, and a set of feature maps are characterized from the deep convolutional layers (termed as *conv*11, *conv*12, and *conv*13) that corresponding to the semantic information involved in the visual content. Similar to low-level visual feature extraction, frame-based features are obtained and then converted to segment-based visual features through an average calculation. In total, for each segment, there are 1 low-level visual feature and 3 high-level visual features.

2.4.2 Audio content

Similar to the visual feature extraction, both low- and high-level audio features are extracted for audio-based multimedia stimulation effect analysis. Zero-crossing rate (ZCR) (Teixeira *et al.*, 2012) is an inherent acoustic characteristic, which estimates the fundamental frequency components of audio signals by counting the average times that the audio amplitude crosses zero within a given time interval. In this study, ZCR features are extracted at each segment as the low-level audio features, and the temporal variations among segment-based audio features are utilized to present the changes of audio content during emotion induction.

For high-level audio feature extraction, a pre-trained deep convolutional neural network, VGGish (Hershey *et al.*, 2017), is adopted (Fig. 5 (b)). VGGish contains six convolutional layers and has been demonstrated as a powerful audio feature extractor for semantic information searching. Each segment is input into VGGish and the high-level audio features are extracted at the 5th and 6th convolutional layers (termed as *conv*5 and *conv*6) that corresponding to the semantic information involved in the audio content. In total, for each segment, there are 1 low-level audio feature and 2 high-level audio features.

2.4.3 Timing effect

To interpret the dynamic process of multimedia-evoked emotion induction, the timing effect of multimedia stimulation on EEG microstate activities is investigated in a time-shifting manner. The temporal variation relationships between multimedia content and evoking emotions are examined in terms of microstate representation, and the time courses of visual and audio features are shifted in a preceded or succeeded direction. Here, the shifting range is given from −1 s (stimulation preceded) to 1 s (stimulation succeeded), with a step of 100 ms. In total, 21 time-shifting parameters. Note that no time-shifted processing is applied on EEG microstate features. For each time-shifting parameter, the temporal correlation between the shifted multimedia stimulation and EEG microstate responses is measured as follows. **(1) Temporal brain dynamics computation.** For each trial, the subject-specific time-varying EEG microstate activities are characterized by calculating the first-order differences between any two adjacent segments in terms of each microstate feature. **(2)**

Temporal multimedia dynamics computation. For each multimedia stimulation, the contents are first shifted at the given time-shifting parameter. Then, the temporal changes are measured by computing the first-order differences in terms of segment-based visual or audio features. **(3) Subject-specific correlation measurement.** For each subject and each microstate feature, the temporal correlation between the computed temporal brain dynamics and the computed temporal multimedia dynamics (with a specific time-shift parameter) is measured by calculating the Pearson correlation coefficients. For each multimedia stimulus, in total 32 subject-specific correlation coefficients (corresponding to 32 subjects) are obtained for each microstate feature and for each time-shifting parameter. **(4) Video-specific correlation measurement.** For each multimedia stimulus and each microstate feature, the obtained 32 subject-specific correlation coefficients are then verified by t-statistic calculation (cross-subject measurement). The calculated t-values quantify the general changing trend of EEG microstate activities in response to the given multimedia stimulation (after time-shifting) across 32 subjects. **(5) Cross-video temporal correlation evaluation.** To explore the cross-video and cross-subject stimulation effect on the dynamic changes of each EEG microstate feature, the calculated 40 t-values (corresponding to 40 videos) in step (4) are then fed into a two-tailed one-sample t-test. The final obtained t-value characterizes the temporal stimulation effect on EEG microstate activities with a positive or negative correlation, and the corresponding p-value reveals whether the temporal stimulation effect is statistically significant. The above steps (1)-(5) are repeated until each time-shifting parameter and each microstate feature are evaluated, and the overall timing effect of multimedia stimulation on emotional EEG dynamics is obtained.

3 Results

In this section, we will present the results of emotion-related EEG microstate dynamics under the investigation of emotion dynamics analysis (Study 1) and multimedia stimulation effect analysis (Study 2).

3.1 EEG microstates

Based on the DEAP database, four EEG microstates are detected in a data-driven manner as presented in Section 2.1. The corresponding GEV and CV values in microstate detection are 82.23% and 63.33%, respectively (Fig. 6 (a)). The detected microstate templates are presented in Fig. 6 (b), which share similar topographical configurations to the canonical microstates reported in the literature (Pascual-Marqui *et al.*, 1995; Koenig *et al.*, 2002; Britz *et al.*, 2010; Michel and Koenig, 2018). For consistency, we label the detected microstates as MS1, MS2, MS3, and MS4 according to topographical orientation.

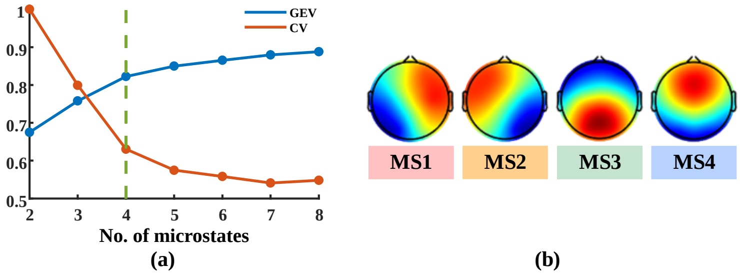


Fig. 6 The detected EEG microstate templates for emotion-related EEG dynamics analysis. (a) Detection performance. Under consideration of CV and GEV values in the detection process, an optimal cluster number of 4 is detected and the corresponding GEV and CV values are 82.23% and 63.33%. For visualization, the CV value is normalized to the range of [0, 1]. (b) The final detected EEG microstates, named MS1, MS2, MS3, and MS4.

3.2 Results of Study 1

Emotion-related EEG dynamics are analyzed to examine the activation differences of EEG microstates under different emotion states in the perspective of task effect, level effect, and evoking dynamics.

3.2.1 Task effect

For emotion task effect analysis, we examine the activation difference of EEG microstates in the pre-stimulus (before emotion-evoking task) and post-stimulus (after emotion-evoking task) stages. As shown in Fig. 7, a significant increasing trend (after FDR) is observed in MS2 coverage ($p < 0.05$) and MS4 coverage ($p < 0.05$), while a significant decreasing trend is observed in MS3 coverage ($p < 0.01$), duration ($p < 0.05$), and occurrence ($p < 0.05$). The transitions from MS3 to MS2 and from MS4 to MS2 significantly increase after emotion task manipulation, while the transition from MS4 to MS3 significantly decreases. These results reflect that the emotion-evoking task leads to a change in EEG microstate activities with distinct patterns. It is found that a positive task effect is observed in MS2 and MS4, meanwhile, a negative task effect is observed in MS3.

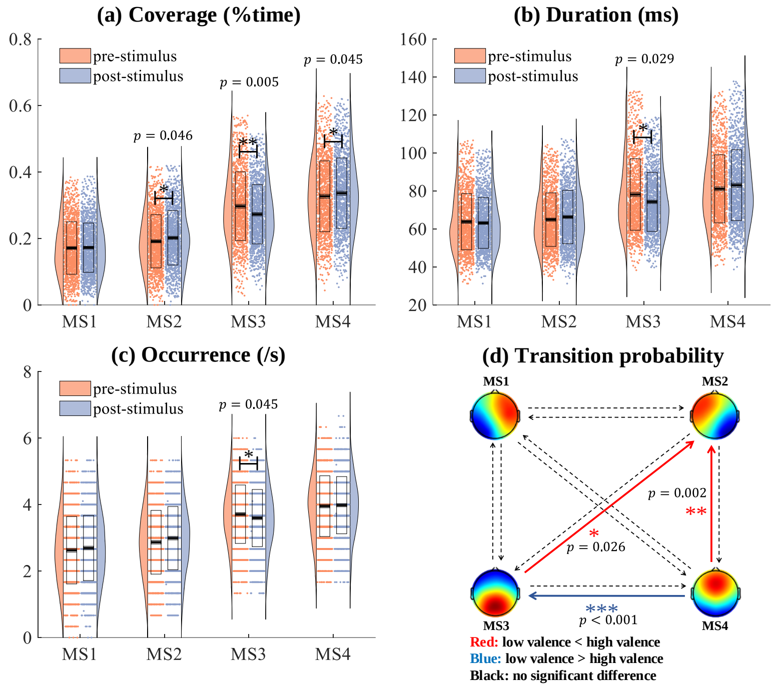


Fig. 7 Emotion-evoking task effect evaluation on EEG microstate activities by measuring the paired difference between pre-stimulus and post-stimulus stages. The cross-subject statistical results of paired differences in terms of each microstate feature: (a) coverage, (b) duration, and (c) occurrence. Here, the dots represent the original feature distribution and the outlines of violin plot represent the kernel probability density estimation. Box plots illustrate the inter-quartile ranges of the features, along with median lines in black. (d) The cross-subject statistical results of paired differences in terms of microstate transition. Red arrows indicate an increasing transition from pre-stimulus to post-stimulus, while blue arrows represent a decreasing transition after the emotion-evoking task occurs. All p-values are corrected with FDR, setting the statistical significance at 5%. (* p<0.05, ** p<0.01, *** p<0.001)

3.2.2 Level effect

The influence of emotion levels (low/high) on brain responses is examined on two independent emotion dimensions (valence and arousal). For the valence dimension (Fig. 8), a lower MS4 occurrence is observed in the high valence group as compared to the low valence group, with the corresponding p-value of 0.038 ($p < 0.05$). For the other EEG microstates (MS1, MS2, and MS3), no significant statistical difference is observed between low and high valence groups. By comparing the microstate transition probability between low and high valence groups, we observe a greater transition from MS1 to MS2 ($p < 0.05$) and a lower transition from MS1 to MS4 ($p < 0.05$) in the high valence group.

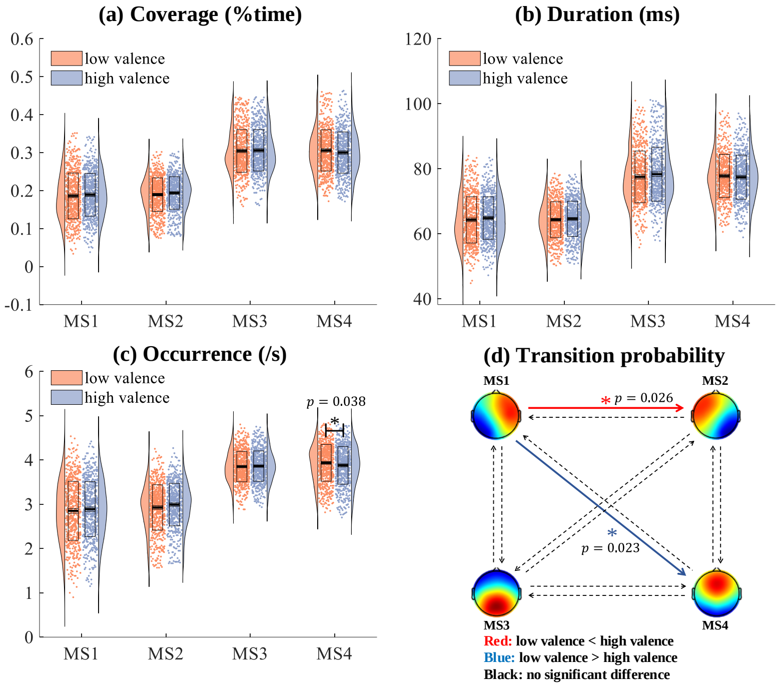


Fig. 8 Valence-based level effect on EEG microstate activities. An independent t-test/Wilcoxon Rank Sum test is utilized to examine the statistical differences in EEG microstate activities between low and high valence groups. In the statistical results of (a) coverage, (b) duration, and (c) occurrence, the dots represent the original feature distribution and the outlines of violin plot represent the kernel probability density estimation. Box plots illustrate the inter-quartile ranges of the features, along with median lines in black. In the statistical results of transition probability (d), red arrows indicate an increasing transition from low to high valence group, while blue arrows represent a decreasing transition in the high valence group as compared to the low valence group. (* p<0.05)

For the arousal dimension, the inter-group statistical differences between low and high arousal groups are shown in Fig. 9. It shows greater coverage ($p < 0.05$) and occurrence ($p < 0.05$) of MS3 are found in the high arousal group as compared to the low arousal group. For MS1, MS2, and MS4, no significant difference is observed between low and high arousal groups. By comparing the differences in microstate transition probability between low and high arousal groups, a higher transition probability from MS2 to MS3 ($p < 0.05$) is observed in the high arousal group.

The above results show that the level differences in emotion could be reflected by the patterns of microstate activities, especially MS3 and MS4. Distinct activation patterns of EEG microstates are observed on valence and arousal. MS4 is sensitive to the changes in valence levels, where high valence leads to a lower MS4 occurrence. In contrast, MS3 activity is related to arousal levels, where high arousal leads to higher MS3 coverage and occurrence.

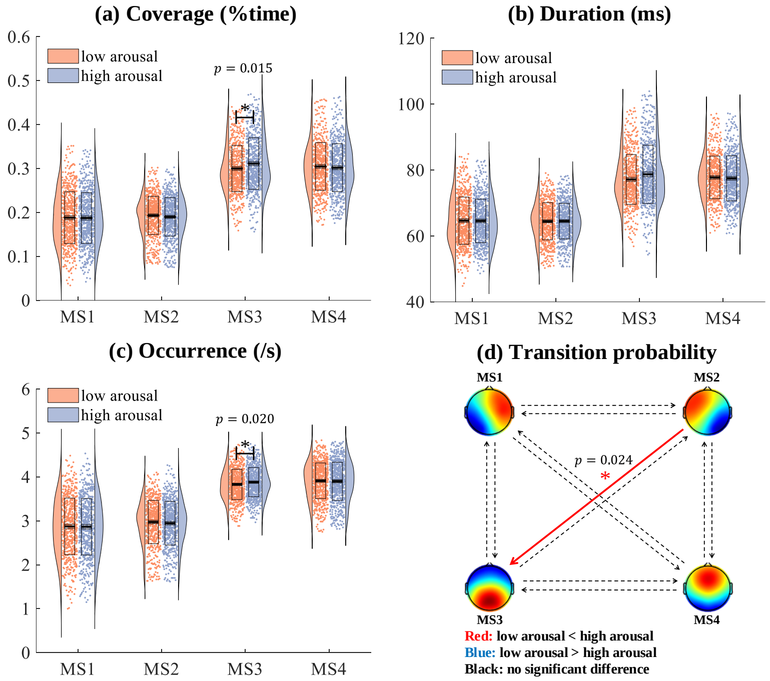
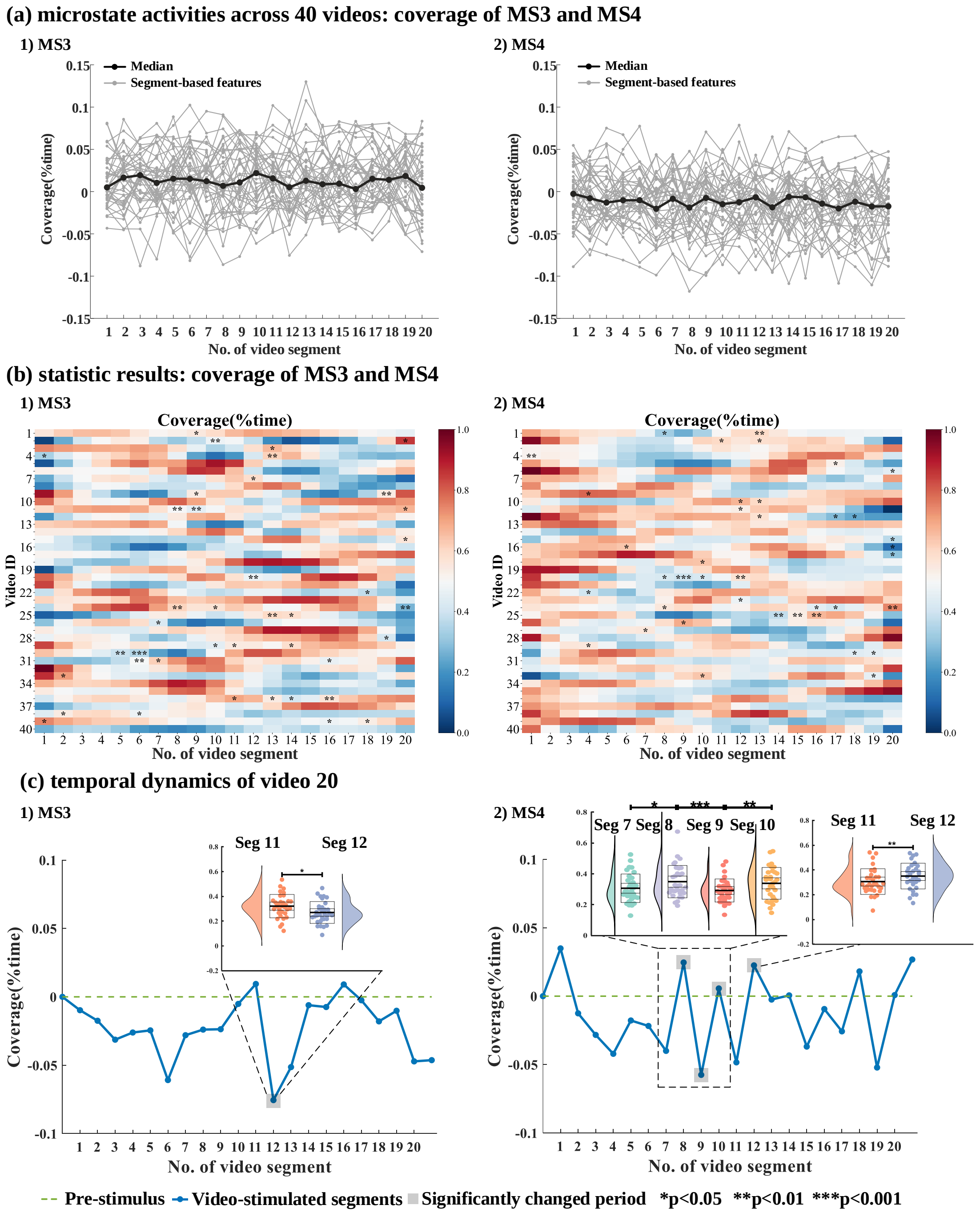


Fig. 9 Arousal-based level effect on EEG microstate activities. An independent t-test/Wilcoxon Rank Sum test is utilized to examine the statistical differences in EEG microstate activities between low and high arousal groups. In the statistical results of (a) coverage, (b) duration, and (c) occurrence, the dots represent the original feature distribution and the outlines of violin plot represent the kernel probability density estimation. Box plots illustrate the inter-quartile ranges of the features, along with median lines in black. In the statistical results of transition probability (d), red arrows indicate an increasing transition in the high arousal group comparing to the low arousal group, while blue arrows represent a decreasing transition from low to high arousal group. (* p<0.05)



(a) microstate activities across 40 videos: coverage of MS3 and MS4

1) MS3

2) MS4

(b) statistic results: coverage of MS3 and MS4

1) MS3

2) MS4

(c) temporal dynamics of video 20

1) MS3

2) MS4

Fig. 10 Video-evoked emotion dynamics in terms of segment-based EEG microstate activities. (a) Visualization of dynamic microstate activities in terms of MS3 and MS4 coverage across 40 videos. A gray line refers to one video's temporal dynamic pattern and the black line is the median of all the temporal dynamic patterns across 40 videos. (b) Statistic results of the temporal variations of EEG microstate activities during emotion induction. In the heatmap, the colors refer to the microstate feature values, where the feature values are calculated as an average of segment-based MS3 or MS4 coverage features across 32 subjects for each video and normalized into the range of [0, 1] for visualization. The segments marked by "*", "**", or "***" refer to the turning points at which the microstate activities of the current time segment are significantly changed comparing to the previous time segment. (c) A detailed example of the temporal changes in terms of MS3 and

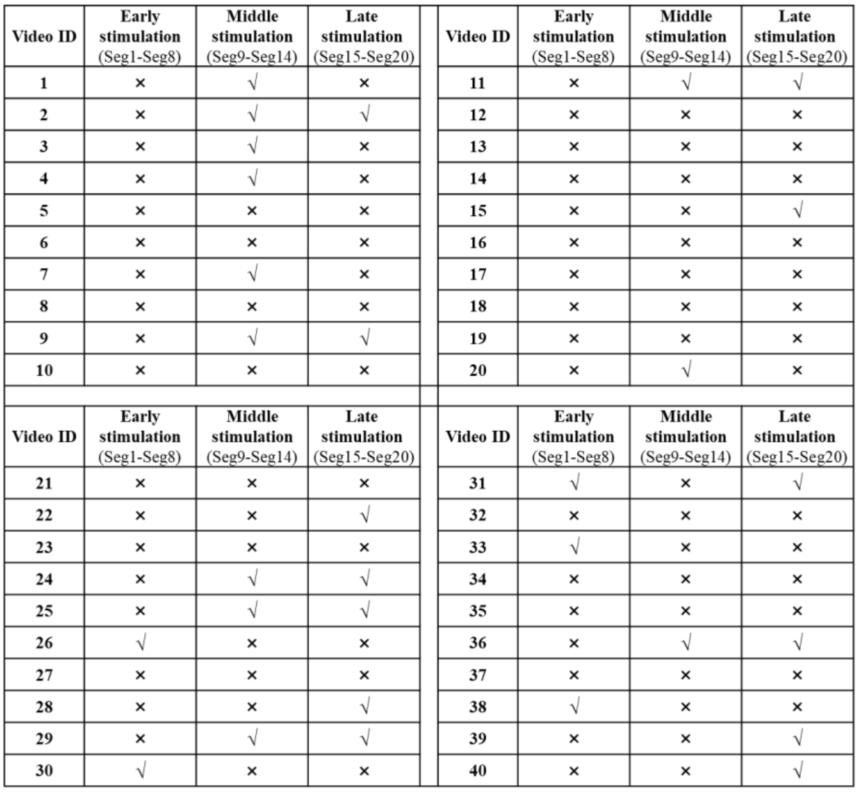
MS4 coverage under an emotion evoking using video 20. (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

3.2.3 Evoking dynamics

The evoking emotion dynamics are studied in terms of EEG microstate activities, by measuring the changing differences between any two adjacent segments. The microstate differences from moment to moment are evaluated as the temporal variation of emotion-evoked EEG dynamics. In line with the previous observations that MS3 and MS4 play an important role in high-level cognitive function and conscious processing (Khanna *et al.*, 2015), our results in the study of task effect and level effect also demonstrate that MS3 and MS4 are more related to emotion perception. Next, the exploration of emotion-related evoking dynamics will mainly focus on these two microstates (MS3 and MS4). For each trial, the evoking dynamics in terms of MS3 and MS4 features are analyzed across 32 subjects and a cross-subject multimedia-specific activation pattern is obtained for each video. Here, we take the coverage feature as an example and report the results in Fig. 10. More detailed results about the other microstate features are presented in Appendix III of the Supplementary Materials. As presented in Fig. 10 (a), varied activation patterns of EEG microstate responses in terms of MS3 and MS4 coverage are observed at 40 trials (videos). For each video, we carefully examine the temporal variations of EEG microstate activities. As shown in Fig. 10 (b), the time segments with significant statistical differences to the previous segments could be considered as "turning points" in the multimedia-evoked brain responses. According to our observations, different videos (different multimedia content) lead to different evoking patterns on the temporal characteristics of EEG microstate activities, where the turning points happen at different time moments and the temporal dynamic patterns of EEG microstate activities perform differently. A detailed observation about the evoking dynamics of video 20 is given as an example (Fig. 10 (c)). The turning points mainly occur from Seg 7 to Seg 12 (in total 20 segments; video length: 60 s; segment length: 3 s). For MS3 coverage, one turning point is observed from Seg 11 to Seg 12 with a significant decreasing trend. For MS4 coverage, significantly increasing trends are observed from Seg 7 to Seg 8, from Seg 9 to Seg 10, and from Seg 11 to Seg 12, and a significant decreasing trend is found from Seg 8 to Seg 9. The differences of the temporal dynamic evoking effects on emotional EEG microstate activities could be possibly explained by assuming that the content differences in multimedia stimulation would lead to different evoking reactions in the brain during emotion induction.

Besides, the temporal distributions of the identified turning points for each video are reported in Table 1 (for MS3 coverage) and Table 2 (for MS4 coverage). It is found that the turning points are observed in 55% of videos (22/40). For MS3 coverage, it is found 22.7% of turning points are observed at the early stimulation stage, 54.5% at the middle stimulation stage, and 59.1% at the late stimulation stage. For MS4 coverage, 22.7% of turning points are found at the early stimulation stage, 50.0% at the middle stage, and 45.5% at the late stimulation stage. These results generally show that the turning points are mainly distributed at the middle and late stimulation stages, which suggests the temporal patterns of stimulation perception during emotion induction. Overall, the results obtained in Study 1 reveal that the dynamic changes of the evoked emotions can be well characterized by EEG microstate representations, especially by MS3 and MS4.

Table 1 The temporal distribution of turning points of 40 videos in terms of MS3 coverage.



Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)	Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)
1	×	√	×	11	×	√	√
2	×	√	√	12	×	×	×
3	×	√	×	13	×	×	×
4	×	√	×	14	×	×	×
5	×	×	×	15	×	×	√
6	×	×	×	16	×	×	×
7	×	√	×	17	×	×	×
8	×	×	×	18	×	×	×
9	×	√	√	19	×	×	×
10	×	×	×	20	×	√	×
Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)	Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)
21	×	×	×	31	√	×	√
22	×	×	√	32	×	×	×
23	×	×	×	33	√	×	×
24	×	√	√	34	×	×	×
25	×	√	√	35	×	×	×
26	√	×	×	36	×	√	√
27	×	×	×	37	×	×	×
28	×	×	√	38	√	×	×
29	×	√	√	39	×	×	√
30	√	×	×	40	×	×	√

Table 2 The temporal distribution of turning points of 40 videos in terms of MS4 coverage.

Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)	Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)
1	×	√	×	11	×	√	×
2	×	√	×	12	×	√	√
3	×	×	×	13	×	×	×
4	×	×	×	14	×	×	×
5	×	×	√	15	×	×	√
6	×	×	√	16	√	×	√
7	×	×	×	17	×	×	√
8	×	×	×	18	×	√	×
9	√	×	×	19	×	×	×
10	×	√	×	20	×	√	×
Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)	Video ID	Early stimulation (Seg1-Seg8)	Middle stimulation (Seg9-Seg14)	Late stimulation (Seg15-Seg20)
21	×	×	×	31	×	×	×
22	√	×	×	32	×	×	×
23	×	√	×	33	×	√	√
24	×	√	√	34	×	×	×
25	×	×	√	35	×	×	×
26	×	√	×	36	×	×	×
27	√	×	×	37	×	×	×
28	×	×	×	38	×	×	×
29	√	×	×	39	×	×	×
30	×	×	√	40	×	×	×

3.3 Results of Study 2

To characterize the stimulation effect of multimedia content on emotion induction, we separately analyze the temporal correlation between EEG microstate activities and multimedia stimulation in terms of visual content, audio content, and timing effect.

3.3.1 Visual content

The correlations between EEG microstate activities and visual content in terms of low-level and high-level visual features are reported in Fig. 11 (a) and (b). The results show that the changes in visual content have a close relationship with the dynamic activities of EEG microstates, which is mainly reflected in the MS4 activity. For low-level visual features, a positive correlation is found between the brightness and MS4 occurrence, where a higher MS4 occurrence is observed when the presented videos with high brightness. For high-level visual features, a positive correlation is observed between the visual features extracted from *conv*11 to *conv*13 and MS4 coverage and duration. Besides, a negative association is also observed between the high-level visual features extracted from *conv*11 to *conv*12 and MS3 coverage. Compared to the stimulation effect of low-level visual features, a more complex stimulation effect is found for high-level visual features, suggesting that high-level features are more related to the changing activities of MS3 and MS4 and play a more important role in emotion induction.

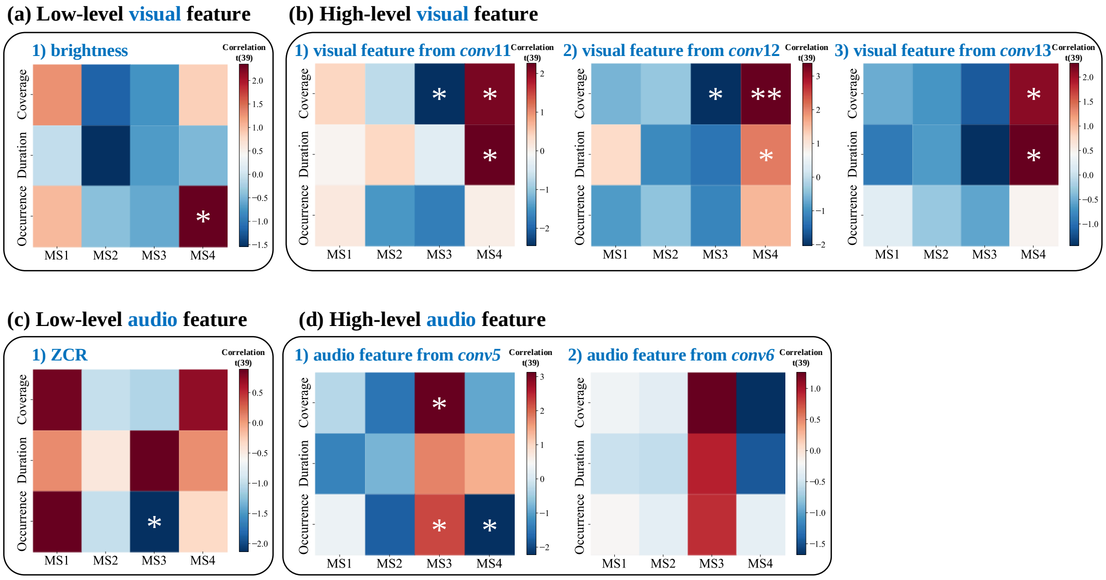


Fig. 11 Stimulation effect of visual and audio content on segment-based microstate features. The heatmap is a visual display of the calculated t values from correlation relationship measurement. A positive t value is marked as red, which indicates a positive correlation between stimulation content and EEG microstate activities. A negative t value is marked as blue, which refers to a negative correlation between stimulation content and microstate activities. (* p<0.05, ** p<0.01)

3.3.2 Audio content

Similar to the visual content effect analysis in Section 3.3.1, the correlations between audio content and EEG microstate activities are reported in Fig. 11 (c) and (d). The results show that audio content mainly influences MS3 activity. For low-level audio features, a negative correlation (an increase of ZCR feature leads to a decrease of MS3 occurrence) is observed. For high-level audio features, a positive correlation (the audio features extracted from *conv*5 activate MS3 responses for a larger coverage and occurrence) is found. At the same time, a negative correlation is observed between the features extracted from *conv*5 and MS4 occurrence. In the investigation of the audio content effect in terms of high-level audio features, a complex correlation between EEG microstate activities of MS3 and MS4 and the high-level audio features extracted from *conv5* is observed. However, for the high-level audio features extracted from *conv*6, no significant correlation with EEG microstate activities is found. One possible reason could be that the utilized VGGish was trained for audio classification based on the Youtube-8M database (Abu-El-Haija *et al.*, 2016), and the audio features characterized at the last convolutional layer (*conv*6) could reflect more about classification information instead of the affective information.

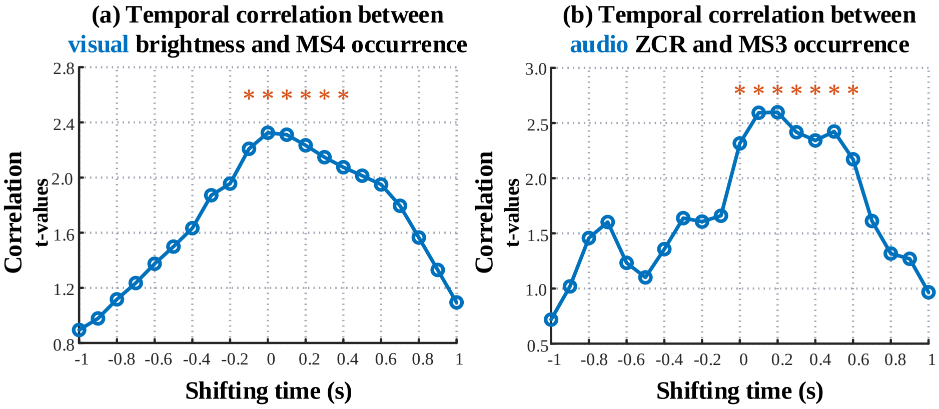


Fig. 12 Temporal correlation between shifted multimedia content and EEG microstate activities. (a) Temporal correlation between the time-shifting visual content (brightness) and MS4 occurrence. (b) Temporal correlation between the time-shifting audio content (ZCR) and MS3 occurrence. (* p<0.05; ZCR: zero-crossing rate)

3.3.3 Timing effect

We shift the multimedia stimulation content in a preceded or succeeded direction with a time range of -1s to 1s and measure the corresponding temporal correlation with the EEG microstate activities at every time-shifting parameter. For visual content perception, the results (Fig. 12 (a)) reveal that the valid time effect is in the range of −100 ms to 400 ms. The highest correlation is reached at 0 ms (stimulation onset), and then the correlation declines from 0 ms to 400 ms. Besides, a preceded effect of visual content on EEG microstate dynamics is also observed before the stimulation onset (-100 ms). One possible reason could be that there may exist an expectation effect before visual content presentation, as the adjacent stimulation content in continuous videos is closely content related. For audio content perception, it is found that the timing effect mainly occurs from 0 to 600 ms (Fig. 12 (b)). The highest correlation happens at 200 ms, which shows a post-stimulus effect of audio content on brain responses during emotion induction. These results show that, for visual and audio content, the timing effects on simultaneous brain responses are different, which results in a distinct activation pattern of EEG microstate activities during multimedia-evoked emotion induction. Stimulation perception of visual content is closely related to MS4 activities with an onset effect, and that of audio content is more related to MS3 activities with a post-stimulus effect.

4 Discussion

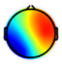
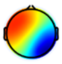
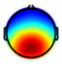
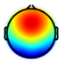
The main goal of this study is to explore emotions from a dynamic perspective, in which the associations among evoked emotion states, spontaneous EEG activities, and the content of multimedia stimulation are measured. The observations of EEG microstate dynamics in video-evoked emotion study and the potential neural mechanisms underlying emotion perception are further discussed in this section.

4.1 Neurophysiological significance of EEG microstates

In this work, four EEG microstate templates are detected from a series of reference-free potential topographies using a sequential clustering method. It is found that the identified EEG microstates share similar topographical configurations to the canonical EEG microstates in the previous studies (Britz *et al.*, 2010; Musso *et al.*, 2010; Khanna *et al.*, 2015; Michel and Koenig, 2018). Our results consistently show that the intrinsic EEG dynamics of various emotion states under video-watching tasks can be efficiently characterized by the spatial-temporal variations of EEG microstate activities. Based on the literature, MS1, MS2, MS3, and MS4 are believed to be correlated with four important functional brain networks (Table 3), including the auditory, visual, default mode, and dorsal attention networks. For example, Britz *et al.* (Britz *et al.*, 2010) validated the spatial correlation between EEG microstates and the four functional brain networks based on the simultaneous EEG-fMRI data. In their study, it was detected that MS1 was spatially correlated with the negative BOLD activation over bilateral superior and middle temporal gyri, MS2 was mapped with the negative BOLD activation in bilateral extrastriate visual areas, MS3 was related with the positive BOLD signals in the anterior cingulate cortex and bilateral inferior frontal gyri that are important brain regions for emotional salience information integration, and MS4 was correlated with the negative BOLD activation in the right-lateralized dorsal and ventral areas of frontal and parietal cortex (these brain regions functionally activate for attention shifting and reorienting task during external stimulation processing). Through estimating the language processing of unresponsive patients, Gui *et al.* (Gui *et al.*, 2020) found a higher activation of MS1 and MS2 but a lower activation of MS3 and MS4 in patients when compared to the healthy group. Their results also suggested that MS1 and

MS2 mainly respond to low-level sensory brain processing, whereas MS3 and MS4 respond to high-level cognitive functional networks. Hence, taking the functional significance of EEG microstates in neurophysiology into consideration, the underlying neural mechanism of emotions could be further revealed by inspecting the functional patterns of EEG microstate activities.

Table 3 A summary of neurophysiological significances for EEG microstates

Microstates	Our observations	Functional significance
MS1 	-	1. Associated with auditory network; 2. Spatially correlated with negative BOLD activation in bilateral superior and middle temporal gyri.
MS2 	1. Negative association with visual stimulation 2. Lower coverage after emotion-evoking tasks	1. Associated with the visual network; 2. Spatially associated with the negative BOLD activation in bilateral extrastriate visual areas.
MS3 	1. Positive relation with arousal states (higher coverage and occurrence in high-level arousal) 2. Significantly respond to audio stimulation	1. Associated with the default mode network; 2. Related with the positive BOLD signals in the anterior cingulate cortex, bilateral inferior frontal gyri; 3. Integrate information of subjective perception and emotion salience (task-negative).
MS4 	1. Negative relation with valence states (lower occurrence in high-level valence) 2. Significantly respond to visual stimulation	1. Associated with the dorsal attention network; 2. Correlated with the negative BOLD activation in ventral areas of frontal and parietal cortex and right-lateralized dorsal; 3. Related with the activity of attention switching and reorientation (task-positive).

4.2 The influence of emotion states on EEG microstate activities

In consideration of the influence of emotion-evoking tasks on EEG microstate dynamics, a significant representation difference is observed by comparing the EEG microstate activities at the pre-stimulus and post-stimulus stages. It is found that emotional task manipulations lead to an increase in MS2 and MS4 coverage and a decrease in MS3 coverage, duration, and occurrence. The fast-changing visual content in an emotional video activates the visual network, which could be reflected as an increase in MS2. For the decrease of MS3 activities, one possible reason could be that the corresponding default mode network is task-negative so that emotion perception and processing would inhibit the activation of the default mode network. The increase of MS4 activities is in line with the functional activity of the dorsal attention network (task-positive) that would be activated in multimedia-directed emotion induction for attention reorientation and focus switching. However, no significant difference is found in MS1 (related to the auditory network). According to Britz *et al.*'s work (Britz *et al.*, 2010), it was found that MS1 simultaneously corresponded to a negative BOLD activation mainly in bilateral superior and middle gyri (auditory-related functional brain regions) and the primary visual cortex of the bilateral extrastriate cortex (visual-related functional brain regions) (Mantini *et al.*, 2007). In Milz *et al.*'s work for cognitive task-based EEG microstate analysis, a larger MS1 duration was found during the visual-stimulated task comparing to the resting-state and audio-stimulated task (Milz *et al.*, 2016). In our case, as video-based multimedia stimulation is adopted for emotion induction, the simultaneous presentation of audiovisual content would possibly inhibit the activities of MS1.

For the EEG microstate dynamics analysis, our results are highly consistent with the findings in the previous works. For example, Seitzman *et al.* (Seitzman *et al.*, 2017) found that task manipulation would inhibit the activities of MS3 and promote the activities of MS4. For serial subtraction tasks, the coverage, duration, and occurrence of MS3 significantly decreased and these features of MS4 significantly increased after task manipulation. Similar results were also observed by Kim *et al.* (Kim *et al.*, 2021) that a significant decrease of MS3 and a significant increase of MS4 were found in good performance groups while performing mental arithmetic tasks. These findings generally suggest that the influences of cognitive task manipulation (refers to emotion induction in this work) on microstate activities are in line with task effects on the default mode network (task-negative, associated with MS3) and dorsal attention network (task-positive, associated with MS4).

4.3 The effect of emotion level differences on EEG microstate activities

In the investigation of microstate representation differences in valence and arousal dimension, we observe that emotion level differences mainly influence the activities of MS3 and MS4. Valence level is positively correlated with MS4 occurrence, while arousal level is negatively correlated with MS3 coverage and occurrence. These findings are congruent with the observations of emotion level effects on the activities of functional brain networks in the previous studies. Nummenmaa *et al.* (Nummenmaa *et al.*, 2012) found that low valence increased the activities of the default mode network for emotion perception and induction, while high arousal activated the dorsal attention network for attention switching via external multimedia stimulation. Besides, Colibazzi *et al.* (Colibazzi *et al.*, 2010) obtained similar observations in a self-generated emotion induction task. For low valence emotions, higher BOLD signals were detected in the right dorsolateral prefrontal cortex and rostral dorsal anterior cingulate cortex that are spatially belonged to the dorsal attention network. For high arousal emotions, higher BOLD responses were detected in the midline and medial temporal lobes that are belonged to the default mode network. All the above results suggest that the dorsal attention network and default mode network are essential for emotion processing, where the activities of the dorsal attention network (MS4) are negatively related to the valence level, and the default mode network (MS3) is positively correlated with arousal level.

Furthermore, the representation differences of EEG microstate activities in valence and arousal dimensions can be interpreted by the functional significance of the dorsal attention network and default mode network during emotion induction. As well investigated in the literature, the dorsal attention network plays a dominant role in stimulus-directed sensory and functions for extracting salient emotional information for valence induction (Szczepanski *et al.*, 2013). The default mode network is mainly involved in self-reflection and internal perception, both of which are important for arousal processing (McKiernan *et al.*, 2003). These findings are also reflected in our observations of the emotion level effect on EEG microstate activities, resulting in a positive association between valence level and MS4 activities and a negative association between arousal level and MS3 activities.

4.4 Multimedia stimulation effect on microstate activities

In this work, the temporal association between multimedia content and EEG microstate activities is measured to describe the dynamic process of emotion perception in the brain. In previous affective computing studies, it was found that visual brightness as a fundamental and commonly used visual feature has been validated as an essential visual attribute for emotional valence induction in video affective content analysis (Wang and Ji, 2015). For the audio ZCR features, it has been validated as a key acoustic feature for emotion arousal enhancement and has been widely applied for multimedia content-based emotion recognition (Zhang *et al.*, 2010). In our study, these multimedia features are also found to be highly associated with emotion-evoked EEG microstate activities that a higher brightness leads to a higher occurrence of MS4, and a higher ZCR leads to a lower MS3 occurrence.

Through investigating the temporal correlation between multimedia stimulation and dynamic EEG activities during emotion induction, our findings demonstrate that emotion perception is a temporally dynamic process coordinated with the stimulation of multimedia content (Effron *et al.*, 2006). For visual content, the temporal association between visual content and dynamic EEG activities is observed in the time range of −100–400 ms. These results are consistent with the previous findings. For example, Oya *et al.* (Oya *et al.*, 2002) observed a strong gamma response around 150–450 ms on the amygdala under the visual task using emotional pictures. In Potts *et al.*'s visual-stimulated ERP task (Potts, 2004), it was found the P2a (about 220–316 ms) component significantly increased in response to the visual content. In this work, a preceded correlation (–100 ms) is found between visual content and the triggered emotions, which could be explained by the fact that emotion induction is a temporally dynamic process and the previous stimulation continuously affects the following emotion states. Through exploring the potential associations between continuous multimedia content and dynamic EEG microstate activities, it helps us to understand the temporal characteristics of video-evoked emotion processing in the brain.

4.5 Limitations and future work

There remains a lack of clarified identification of emotion-specific functional brain regions and their specialized roles in emotion processing and regulation. Few studies have precisely identified the originated location of surface EEG signals for emotion-related neuronal activity measurement. EEG microstate analysis with the aim of voxel-based source localization would greatly enhance the spatial resolution for accurate reflection of global neuronal activities, and will significantly improve the performance of functional brain activity estimation, monitoring, and regulation. Moreover, fMRI data with high spatial resolution can capture the hemodynamic changes in deep brain structures. Combining EEG microstates and fMRI data will provide a more comprehensive investigation of emotion-related neurophysiological mechanisms in high temporal and spatial resolution.

5 Conclusion

In summary, our work mainly focuses on the dynamic neural mechanism of emotions in terms of EEG microstate analysis and investigates the complex relationship among evoking emotion states, dynamic EEG activities, and multimedia stimulation. The results show EEG microstates, especially for MS3 and MS4, have a close association with high-level functional brain networks and are capable of representing the dynamic characteristics of emotional EEG activities inducted by continuous multimedia stimulation. Distinctive EEG microstate activities are observed under different emotion states from the perspectives of task effect, level effect, and timing effect. Also, it is found that EEG microstate patterns can reflect the emotional stimulation effect from the video content which is used for emotion induction. This work deepens our understanding of emotion dynamics and provides an attractive approach for time-varied emotion-related neural mechanism exploration.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61906122 and Grant 91859122, in part by the Tencent “Rhinoceros Birds” - Scientific Research Foundation for Young Teachers of Shenzhen University, and in part by the High Level University Construction under Grant 000002110133.

References

Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., Vijayanarasimhan, S., 2016. Youtube-8m: A large-scale video classification benchmark. arXiv preprint arXiv:1609.08675.

Alarcão, S.M., Fonseca, M.J., 2019. Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing* 10, 374-393.

Britz, J., Van De Ville, D., Michel, C.M., 2010. BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *NeuroImage* 52, 1162-1170.

Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., Zhu, H., Kangarlu, A., Duan, Y., Russell, J.A., 2010. Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion* 10, 377-389.

D’Croz-Baron, D.F., Bréchet, L., Baker, M., Karp, T., 2021. Auditory and Visual Tasks Influence the Temporal Dynamics of EEG Microstates During Post-encoding Rest. *Brain Topography* 34, 19-28.

Effron, D.A., Niedenthal, P.M., Gil, S., Droit-Volet, S., 2006. Embodied temporal perception of emotion. *Emotion* 6, 1-9.

Gianotti, L.R.R., Faber, P.L., Schuler, M., Pascual-Marqui, R.D., Kochi, K., Lehmann, D., 2008. First Valence, Then Arousal: The Temporal Dynamics of Brain Electric Activity Evoked by Emotional Stimuli. *Brain Topography* 20, 143-156.

Gui, P., Jiang, Y., Zang, D., Qi, Z., Tan, J., Tanigawa, H., Jiang, J., Wen, Y., Xu, L., Zhao, J., Mao, Y., Poo, M.-m., Ding, N., Dehaene, S., Wu, X., Wang, L., 2020. Assessing the depth of language processing in patients with disorders of consciousness. *Nature Neuroscience* 23, 761-770.

Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J., Wilson, K., 2017. CNN architectures for large-scale audio classification. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 131-135.

Horikawa, T., Kamitani, Y., 2017. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications* 8, 15037.

Hu, L., Zhang, Z., 2019. *EEG signal processing and feature extraction*. Springer.

Hu, X., Chen, J., Wang, F., Zhang, D., 2019. Ten challenges for EEG-based affective computing. *Brain Science Advances* 5, 1-20.

Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 1254-1259.

Jenke, R., Peer, A., Buss, M., 2014. Feature Extraction and Selection for Emotion Recognition from EEG. *IEEE Transactions on Affective Computing* 5, 327-339.

Katsigiannis, S., Ramzan, N., 2018. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics* 22, 98-107.

Khanna, A., Pascual-Leone, A., Michel, C.M., Farzan, F., 2015. Microstates in resting-state EEG: Current status and future directions. *Neuroscience & Biobehavioral Reviews* 49, 105-113.

Kim, K., Duc, N.T., Choi, M., Lee, B., 2021. EEG microstate features according to performance on a mental arithmetic task. *Scientific Reports* 11, 343.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2012. DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 18-31.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P.V., Braeker, E., Kleinlogel, H., Isenhart, R., John, E.R., 2002. Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *NeuroImage* 16, 41-48.

Koenig, T., Studer, D., Hubl, D., Melie, L., Strik, W.K., 2005. Brain connectivity at different time-scales measured with EEG. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1015-1024.

Lehmann, D., Faber, P.L., Galderisi, S., Herrmann, W.M., Kinoshita, T., Koukkou, M., Mucci, A., Pascual-Marqui, R.D., Saito, N., Wackermann, J., Winterer, G., Koenig, T., 2005. EEG microstate duration and syntax in acute, medication-naïve, first-episode schizophrenia: a multi-center study. Psychiatry Research: *Neuroimaging* 138, 141-156.

Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology* 48, 609-621.

Liu, X., Li, T., Tang, C., Xu, T., Chen, P., Bezerianos, A., Wang, H., 2019. Emotion Recognition and Dynamic Functional Connectivity Analysis Based on EEG. *IEEE Access* 7, 143293-143302.

Liu, Y., Yu, M., Zhao, G., Song, J., Ge, Y., Shi, Y., 2018. Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals. *IEEE Transactions on Affective Computing* 9, 550-562.

Mantini, D., Perrucci, M.G., Del Gratta, C., Romani, G.L., Corbetta, M., 2007. Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences of the United States of America* 104, 13170-13175.

McKiernan, K.A., Kaufman, J.N., Kucera-Thompson, J., Binder, J.R., 2003. A Parametric Manipulation of Factors Affecting Task-induced Deactivation in Functional Neuroimaging. *Journal of Cognitive*

Neuroscience 15, 394-408.

Michel, C.M., Koenig, T., 2018. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage* 180, 577-593.

Milz, P., Faber, P.L., Lehmann, D., Koenig, T., Kochi, K., Pascual-Marqui, R.D., 2016. The functional significance of EEG microstates—Associations with modalities of thinking. *NeuroImage* 125, 643-656.

Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP Analyses: A Step-by-Step Tutorial Review. *Brain Topography* 20, 249-264.

Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., Winterer, G., 2010. Spontaneous brain activity and EEG microstates. A novel EEG/fMRI analysis approach to explore resting-state networks. *NeuroImage* 52, 1149-1161.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I.P., Hari, R., Sams, M., 2012. Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences* 109, 9599-9604.

Oya, H., Kawasaki, H., Howard, M.A., Adolphs, R., 2002. Electrophysiological Responses in the Human Amygdala Discriminate Emotion Categories of Complex Visual Stimuli. *The Journal of Neuroscience* 22, 9502-9512.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1995. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Transactions on Biomedical Engineering* 42, 658-665.

Potts, G.F., 2004. An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition* 56, 5-13.

Poulsen, A.T., Pedroni, A., Langer, N., Hansen, L.K., 2018. Microstate EEGlab toolbox: An introductory guide. *bioRxiv*, 289850.

Scherer, K.R., 2009. The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion* 23, 1307-1351.

Seitzman, B.A., Abell, M., Bartley, S.C., Erickson, M.A., Bolbecker, A.R., Hetrick, W.P., 2017. Cognitive manipulation of brain electric microstates. *NeuroImage* 146, 533-543.

Shen, X., Hu, X., Liu, S., Song, S., Zhang, D., 2020. Exploring EEG microstates for affective computing: decoding valence and arousal experiences during video watching. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 841-846.

Shen, Y.-W., Lin, Y.-P., 2019. Challenge for Affective Brain-Computer Interfaces: Non-stationary Spatio-spectral EEG Oscillations of Emotional Responses. *Frontiers in Human Neuroscience* 13, 366.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., Yang, X., 2018. A Review of Emotion Recognition Using Physiological Signals. *Sensors* 18.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*:1409.1556.

Skrandies, W., 1989. Data reduction of multichannel fields: Global field power and Principal Component Analysis. *Brain Topography* 2, 73-80.

Sporns, O., 2011. The human connectome: a complex network. *Annals of the New York Academy of Sciences* 1224, 109-125.

Szczepanski, S.M., Pinsk, M.A., Douglas, M.M., Kastner, S., Saalmann, Y.B., 2013. Functional and structural architecture of the human dorsal frontoparietal attention network. *Proceedings of the National Academy of Sciences* 110, 15806-15811.

Teixeira, R.M.A., Yamasaki, T., Aizawa, K., 2012. Determination of emotional content of video clips by low-level audiovisual features. *Multimedia Tools and Applications* 61, 21-49.

Wang, S., Ji, Q., 2015. Video Affective Content Analysis: A Survey of State-of-the-Art Methods. *IEEE Transactions on Affective Computing* 6, 410-430.

Yin, Z., Wang, Y., Liu, L., Zhang, W., Zhang, J., 2017. Cross-Subject EEG Feature Selection for Emotion Recognition Using Transfer Recursive Feature Elimination. *Frontiers in Neurorobotics* 11, 19.

Yuan, H., Zotev, V., Phillips, R., Drevets, W.C., Bodurka, J., 2012. Spatiotemporal dynamics of the brain at rest — Exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *NeuroImage* 60, 2062-2072.

Zanesco, A.P., King, B.G., Skwara, A.C., Saron, C.D., 2020. Within and between-person correlates of the temporal dynamics of resting EEG microstates. *NeuroImage* 211, 116631.

Zhang, S., Huang, Q., Jiang, S., Gao, W., Tian, Q., 2010. Affective Visualization and Retrieval for Music Video. *IEEE Transactions on Multimedia* 12, 510-522.

Zheng, W., Lu, B., 2015. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development* 7, 162-175.

Zheng, W., Zhu, J., Lu, B., 2019. Identifying Stable Patterns over Time for Emotion Recognition from EEG. *IEEE Transactions on Affective Computing* 10, 417-429.

Supplementary Materials

Appendix I: DEAP Database and EEG Processing

In this work, the public DEAP database with 32 participants' EEG recordings is used to evaluate the effectiveness of EEG microstate analysis in the application of affective computing. This publicly well-known DEAP database is a multimodality physiological database for affective decoding (Koelstra *et al.*, 2012). Forty music videos with a fixed length of 60 seconds were presented in a random sequence to every participant for emotion-evoking. Simultaneously, 32-electrode EEG signals were recorded at a sampling rate of 512 Hz, where the electrodes were placed based on the international 10-20 system. Every single trial included four parts: 5 s screen fixation before emotion triggering, 60 s music video playing for emotion-evoking, 3 s screen fixation after video playing, and self-assessment as subjective feedback on evoked emotions. Emotional self-assessments in terms of valence (related with pleasantness) and arousal (related with excitation) were conducted by utilizing a self-assessment manikin (SAM) (Morris, 1995) with a continuous scale from 1 to 9. The experimental pipeline of the DEAP database is shown in Fig. S1.

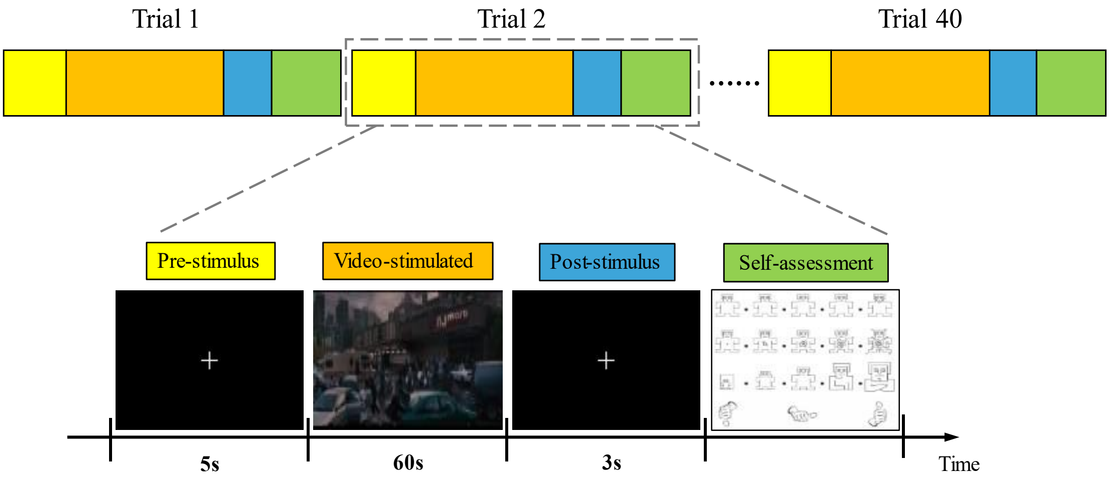


Fig. S1 The experimental pipeline of the DEAP database.

Based on the collected raw EEG signals, a standard preprocessing pipeline is conducted for unwanted noise removal. Firstly, each trial of EEG recordings is filtered with a bandpass filter of 1 - 45 Hz for unrelated artifact filtering and a notch filter of 50 Hz for power line noise removal. Secondly, noisy electrodes are interpolated using an average strategy with the neighboring 3 electrodes. Thirdly, the filtered EEG data are adjusted to spatial zero-mean distribution by conducting a common average re-reference (Brunet *et al.*, 2011). The common average referencing is implemented not only to remove zero-mean random noise for data quality improvement but also to adjust the preprocessed EEG data to a reference-free electric potential distribution for further spatial similarity detection (Pascual-Marqui *et al.*, 1995; Murray *et al.*, 2008). Finally, independent component analysis (ICA) is implemented for ocular artifact removals, such as eye movements and blinks. Meanwhile, other artifacts caused by body movement, cardiac or muscular activity are rejected as much as possible by manual screening. After the preprocessing, the noises in the EEG raw data are greatly removed and a clean EEG data are reconstructed by using the remained independent components. Then, EEG data segmentation is conducted to separate each single-trial data into three conditions: **pre-stimulus (3 s), video-stimulated (60 s), and post-stimulus(3 s)**. Here, to minimize the effect from the previous trial, we only segment the last 3 s of the original pre-stimulus recordings for further analysis.

Appendix II: Self-Adaptive Threshold Reassignment

In the DEAP database, emotional ratings of valence and arousal were collected in a continuous range from 1 to 9, where 1 and 9 indicated the lowest and highest degree of valence or arousal, respectively. Traditionally, a fixed threshold of 5 is directly used to generate binary emotional groups, e.g., low (<5) valence/arousal group and high (>5) valence/arousal group. However, the ranges of the returned self-assessment ratings could be very different from different subjects, due to individual-specific understandings in emotion terms, emotion levels, or personal preference (Kemp *et al.*, 2004; Lee *et al.*, 2005). A two-class grouping using a fixed threshold (e.g., 5) could bring out bias and fail to correctly detect neural differences for each subject. Inspired by Yin *et al.*'s work (Yin *et al.*, 2017), we introduce a self-adaptive threshold reassignment to group video-stimulated trials into low valence (or arousal) and high valence (or arousal) for each subject. More specifically, the self-adaptive threshold reassignment method mainly follows 3 steps:

1) ***Individual-level clustering***. A classical k-means clustering is first conducted to partition 40 ratings of one subject in valence-arousal space into 2 clusters.

2) ***Threshold calculation***. A self-adaptive threshold is then computed as the midpoint of these two cluster centroids.

3) ***Binary emotional grouping***. Based on the computed threshold, 40-trial EEG data collected in the video-stimulated stage are separately divided into low/high valence and low/high arousal groups at the individual level.

An example of subject 4 is shown in Fig. S2. Table S1 displays the individual thresholds of valence and arousal for 32 subjects, which are slightly varied across subjects.

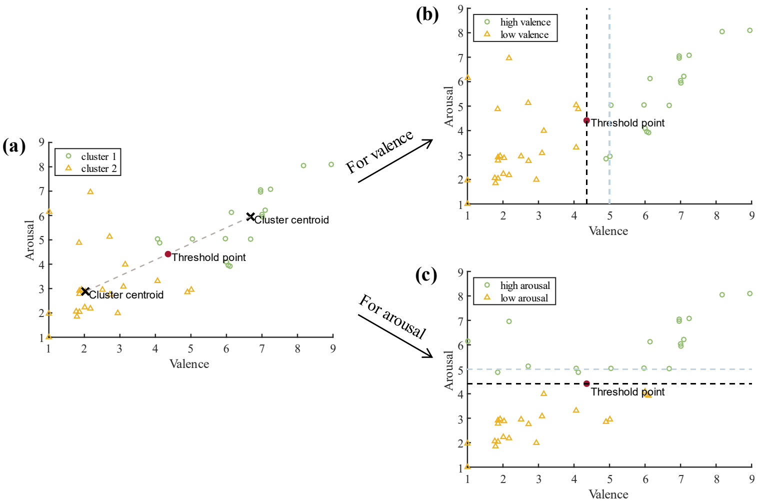


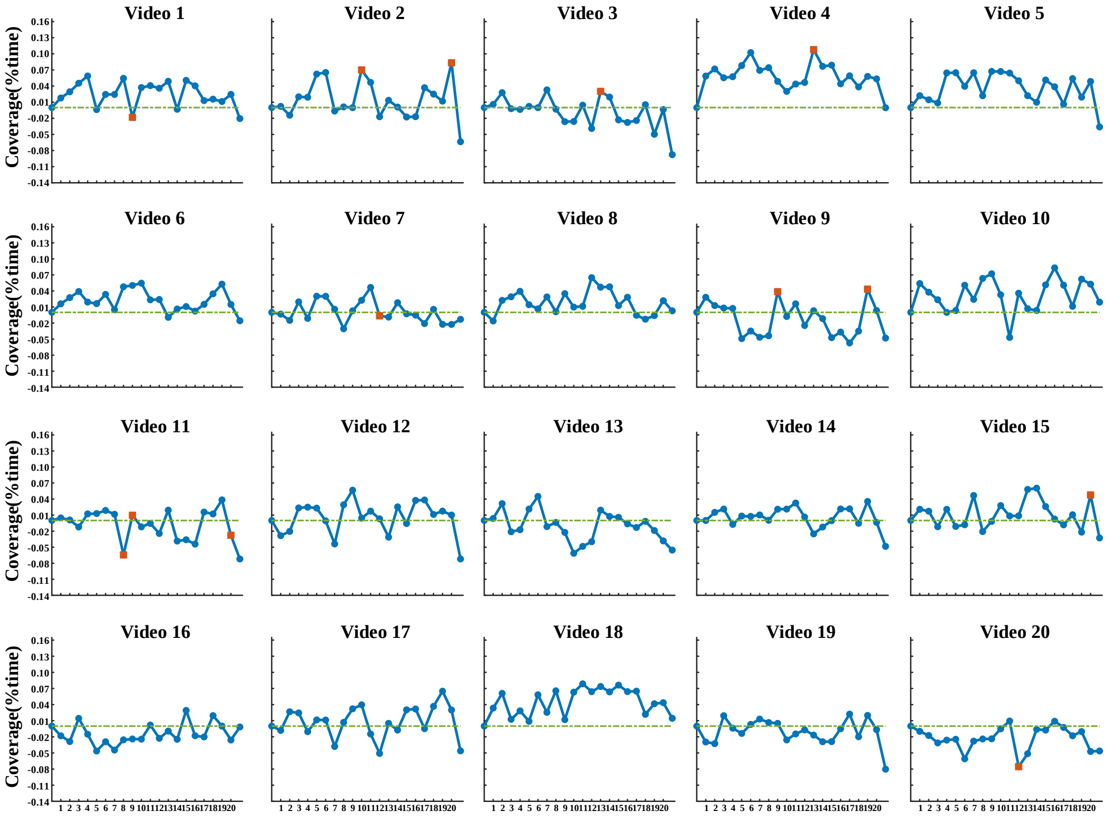
Fig. S2 An example of self-adaptive threshold detection and binary emotional grouping for subjet 4. (a) Self-adaptive threshold is described as the midpoint of two cluster centroids after a classical k-means clustering. 40-trial EEG recordings of subject 4 are divided into (b) low and high valence groups according to the adaptive threshold in valence dimension, and (c) low and high arousal groups according to the adaptive threshold in arousal dimension.

Table S1 The obtained self-adaptive thresholds for subjective ratings in valence and arousal dimensions.

Subject ID	Valence	Arousal	Subject ID	Valence	Arousal
s1	5.125	6.580	s17	5.110	5.705
s2	6.965	4.758	s18	5.375	5.515
s3	5.845	3.668	s19	5.080	5.660
s4	4.355	4.415	s20	6.010	5.540
s5	5.010	4.580	s21	5.543	6.505
s6	6.305	4.800	s22	4.520	6.005
s7	5.025	5.290	s23	6.340	3.030
s8	6.298	5.820	s24	5.510	6.015
s9	5.520	5.550	s25	5.535	6.540
s10	5.590	5.185	s26	4.628	3.495
s11	5.340	4.248	s27	6.518	5.043
s12	4.825	6.930	s28	4.560	4.530
s13	5.045	7.025	s29	4.775	4.465
s14	4.933	6.013	s30	6.030	5.085
s15	6.035	4.585	s31	4.838	5.760
s16	4.475	4.778	s32	5.545	5.550
Sample size of low valence		659	Sample size of low arousal		637
Sample size of high valence		621	Sample size of high arousal		643
Mean threshold for valance ratings				5.394	
Mean threshold for arousal ratings				5.271	

Appendix III: Emotion-Evoking Temporal Dynamics in Microstate Sequences

The dynamic changes in MS3 and MS4 coverage of 40 videos across 32 subjects are shown in Fig. S3 and Fig. S4, respectively.



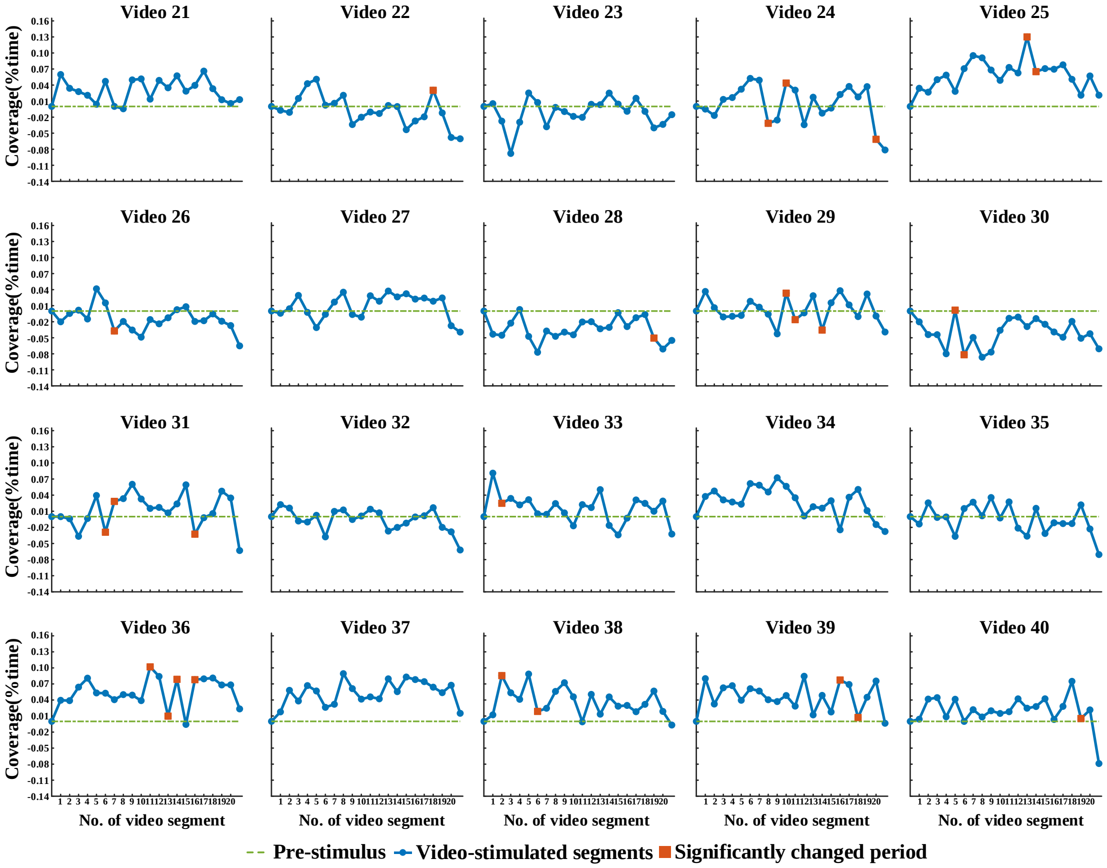


Fig. S3 The temporal dynamics in MS3 coverage during emotion-evoking under 40 videos. The blue line represents the average MS3 coverage across 32 subjects along with video-triggered emotion-evoking, which shows a general time-varied trend of microstate dynamics evoked by one video. The red square marks the turning points that MS3 coverage extracted from this segment is statistically different from the previous segment across 32 subjects.

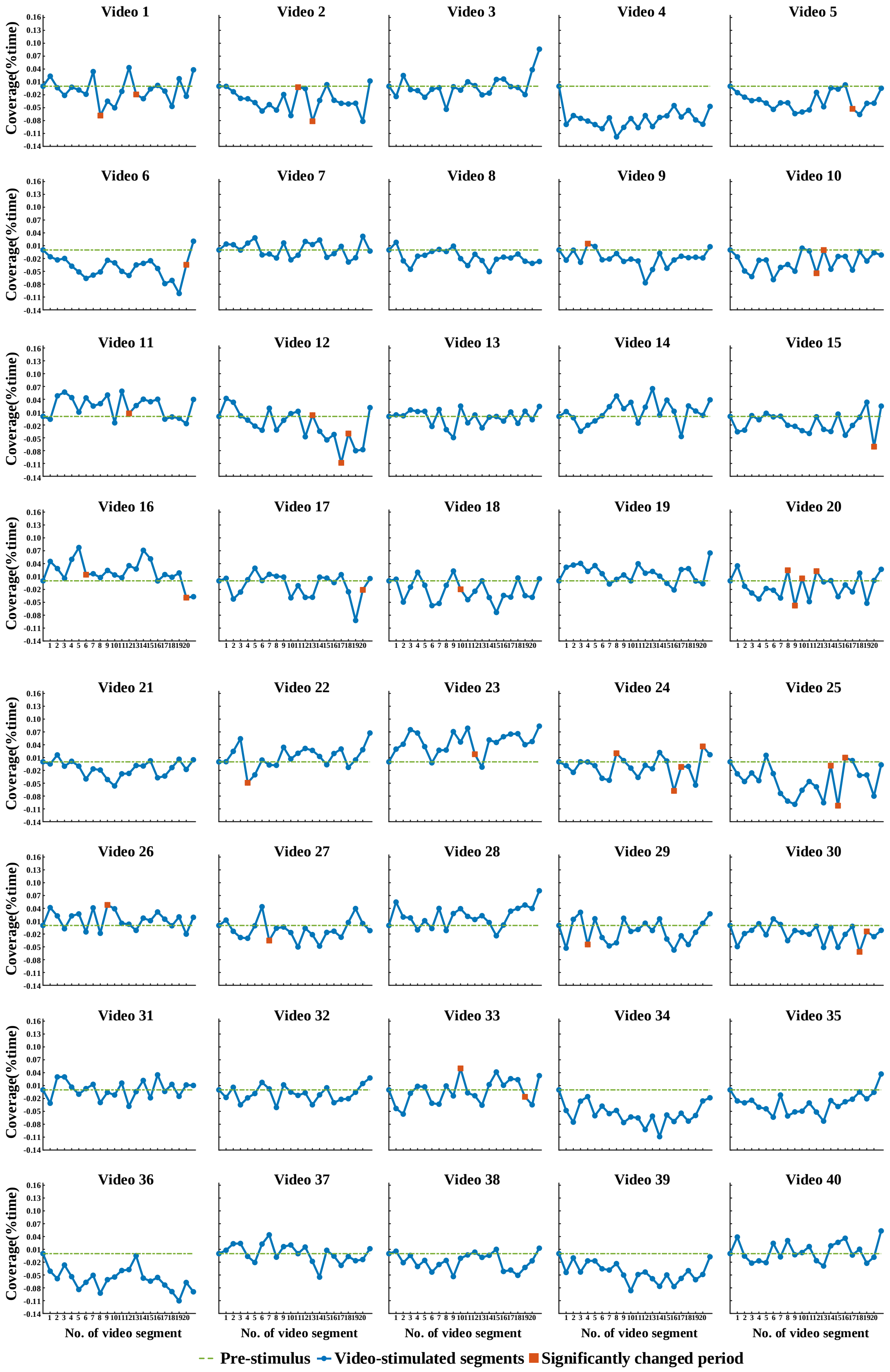


Fig. S4 The temporal dynamics in MS4 coverage during emotion-evoking under 40 videos. The blue line represents the average MS4 coverage across 32 subjects along with video-triggered emotion-evoking, which shows a general time-varied trend of microstate dynamics evoked by one video. The red square marks the turning points that MS4 coverage extracted from this segment is statistically different from the previous segment across 32 subjects.

Reference

Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal Analysis of Multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience* 2011, 813870.

Kemp, A.H., Silberstein, R.B., Armstrong, S.M., Nathan, P.J., 2004. Gender differences in the cortical electrophysiological processing of visual emotional stimuli. *NeuroImage* 21, 632-646.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2012. DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 18-31.

Lee, T.M.C., Liu, H.L., Chan, C.C.H., Fang, S.Y., Gao, J.H., 2005. Neural activities associated with emotion recognition observed in men and women. *Molecular Psychiatry* 10, 450-455.

Morris, J.D., 1995. SAM: the Self-Assessment Manikin. An efficient cross-cultural measurement of emotional response. *Journal of Advertising Research*, p. 63+.

Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP Analyses: A Step-by-Step Tutorial Review. *Brain Topography* 20, 249-264.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1995. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Transactions on Biomedical Engineering* 42, 658-665.

Yin, Z., Wang, Y., Liu, L., Zhang, W., Zhang, J., 2017. Cross-Subject EEG Feature Selection for Emotion Recognition Using Transfer Recursive Feature Elimination. *Frontiers in Neurorobotics* 11, 19.