

Decoding the temporal dynamics of affective scene processing

Ke Bo^{a,b}, Lihan Cui^a, Siyang Yin^a, Zhenhong Hu^a, Xiangfei Hong^{a,c}, Sungkean Kim^{a,d},
Andreas Keil^e, Mingzhou Ding^{a,*}

^a J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, USA

^b Department of Psychological and Brain Sciences, Dartmouth college, Hanover, NH 03755, USA

^c Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China

^d Department of Human-Computer Interaction, Hanyang University, Ansan, Republic of Korea

^e Department of Psychology, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:

Emotion, affective scenes
IAPS
Multivariate pattern analysis
EEG
fMRI
Representation similarity analysis
Visual cortex

ABSTRACT

Natural images containing affective scenes are used extensively to investigate the neural mechanisms of visual emotion processing. Functional fMRI studies have shown that these images activate a large-scale distributed brain network that encompasses areas in visual, temporal, and frontal cortices. The underlying spatial and temporal dynamics, however, remain to be better characterized. We recorded simultaneous EEG-fMRI data while participants passively viewed affective images from the International Affective Picture System (IAPS). Applying multivariate pattern analysis to decode EEG data, and representational similarity analysis to fuse EEG data with simultaneously recorded fMRI data, we found that: (1) ~80 ms after picture onset, perceptual processing of complex visual scenes began in early visual cortex, proceeding to ventral visual cortex at ~100 ms, (2) between ~200 and ~300 ms (pleasant pictures: ~200 ms; unpleasant pictures: ~260 ms), affect-specific neural representations began to form, supported mainly by areas in occipital and temporal cortices, and (3) affect-specific neural representations were stable, lasting up to ~2 s, and exhibited temporally generalizable activity patterns. These results suggest that affective scene representations in the brain are formed temporally in a valence-dependent manner and may be sustained by recurrent neural interactions among distributed brain areas.

1. Introduction

The visual system detects and evaluates threats and opportunities in complex visual environments to facilitate the organism's survival. In humans, to investigate the underlying neural mechanisms, we record fMRI and/or EEG data from observers viewing depictions of naturalistic scenes varying in affective content. A large body of previous fMRI work has shown that viewing emotionally engaging pictures, compared to neutral ones, heightens blood flow in limbic, frontoparietal, and higher-order visual structures (Lang et al., 1998; Phan et al., 2002; Liu et al., 2012; Bradley et al., 2015). Applying MVPA and functional connectivity techniques to fMRI data, we further reported that affective content can be decoded from voxel patterns across the entire visual hierarchy, including early retinotopic visual cortex, and that the anterior emotion-modulating structures such as the amygdala and the prefrontal cortex are the likely sources of these affective signals via the mechanism of reentry (Bo et al., 2021).

Temporal dynamics of affective scene processing remains to be better elucidated. The event-related potential (ERP), an index of average neural mass activity with millisecond temporal resolution, has been the

main method for characterizing the temporal aspects of affective scene perception (Cuthbert et al., 2000; Keil et al., 2002; Hajcak et al., 2009). Univariate ERPs are sensitive to local neural processes but do not reflect the contributions of multiple neural processes taking place in distributed brain regions underlying affective scene perception. The advent of the multivariate decoding approach has begun to expand the potential of the ERPs (Bae and Luck 2019; Sutterer et al., 2021). By going beyond univariate evaluations of condition differences, these multivariate pattern analyses (MVPA) take into account voltage topographies reflecting distributed neural activities and help uncover the discriminability of experimental conditions not possible with the univariate ERP method. The MVPA method can even be applied to single-trial EEG data. By going beyond mean voltages, the decoding algorithms can examine differences in single-trial EEG activity patterns across all sensors, which further complements the ERP method (Grootswagers et al., 2017; Contini et al., 2017). Conceptually, the presence of decodable information in neural patterns has been taken to index differences in neural representations (Norman et al., 2006). Thus, in the context of EEG/ERP data, the time course of decoder performance may inform on how neural representations linked to a given condition or stimulus form

* Corresponding author.

E-mail address: mding@bme.ufl.edu (M. Ding).

<https://doi.org/10.1016/j.neuroimage.2022.119532>.

Received 15 February 2022; Received in revised form 1 July 2022; Accepted 1 August 2022

Available online 2 August 2022.

1053-8119/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

and evolve over time (Cauchoix et al., 2014; Wolff et al., 2015; Dima et al., 2018).

The first question we considered was how long it takes for the affect-specific neural representations of affective scenes to form. For non-affective images containing objects such as faces, houses or scenes, past work has shown that the neural responses become decodable as early as ~100 ms after stimulus onset (Cichy et al., 2014; Cauchoix et al., 2014). This latency reflects the onset time for the detection and categorization of stereotypical visual features associated with different objects in early visual cortex (Nakamura et al., 1997; Di Russo et al. 2002). For complex scenes varying in affective content, however, although mapped onto rich category-specific visual features in a multivariate fashion (Kragel et al., 2019), there are no stereotypical visual features that unambiguously separate different affective categories (e.g., unpleasant scenes vs neutral scenes). Accordingly, univariate ERP studies have reported robust voltage differences between emotional and neutral content at relatively late times, e.g., ~170–280 ms at the level of the early posterior negativity (Schupp et al., 2006; Foti et al., 2009) and ~300 ms at the level of the late positive potential (LPP) (Cuthbert et al., 2000; Lang and Bradley, 2010; Liu et al., 2012; Sabatinelli et al., 2013). We sought to further examine these issues by applying multimodal neuroimaging and the MVPA methodology. It is expected that perceptual processing of affective scenes would begin ~100 ms following picture onset whereas affect-specific neural representations would emerge between ~150 ms and ~300 ms.

A related question is whether there are systematic timing differences in the formation of neural representations of affective scenes differing in emotional content. Specifically, it has been debated to what extent pleasant versus unpleasant contents emerge over different temporal intervals (e.g., Oya et al., 2002). The negativity bias idea suggests that aversive information receives prioritized processing in the brain and predicts that scenes containing unpleasant elements evoke faster and stronger responses compared to scenes containing pleasant or neutral elements. The ERP results to date have been equivocal (Carretié et al., 2001; Huang and Luo, 2006; Franken et al., 2008). An alternative idea is that the timing of emotional representation formation depends on the specific content of the images (e.g., erotic within the pleasant category vs mutilated bodies within the unpleasant category) rather than on the broader semantic categories such as unpleasant scenes and pleasant scenes (Weinberg and Hajcak, 2010). We sought to test these ideas by applying the MVPA approach to decode subcategories of images using EEG data. It is expected that the timing of representation formation is content-specific.

How do neural representations of affective scenes, once formed, evolve over time? For non-affective images, the neural responses are found to be transient, with the processing locus evolving dynamically from one brain structure to another (Carlson et al., 2013; Cichy et al., 2014; Kaiser et al., 2016). For affective images, in contrast, the enhanced LPP, a major ERP index of affective processing, is persistent, lasting up to several seconds, and supported by distributed brain regions including the visual cortex as well as frontal structures, suggesting sustained neural representations. To test whether neural representations of affective scenes are dynamic or sustained, we applied a MVPA method called the generalization across time (GAT) (King and Dehaene, 2014), in which the MVPA classifier is trained on data at one time point and tested on data from all time points. The resulting temporal generalization matrix, when plotted on the plane spanned by the training time and the testing time, can be used to visualize the temporal stability of neural representations. For a dynamically evolving neural representation, high decoding accuracy will be concentrated along the diagonal in the plane, namely, the classifier trained at one time point can only be used to decode data from the same time point but not data from other time points. For a stable or sustained neural representation, on the other hand, high decoding accuracy extends away from the diagonal line, indicating that the classifier trained at one time point can be used to decode data from other time points. It is expected that the neural representations of affective

scenes are sustained rather than dynamic with the visual cortex playing an important role in the sustained representation.

We recorded simultaneous EEG-fMRI data from participants viewing affective images from the International Affective Picture System (IAPS) (Lang et al., 1997). MVPA was applied to EEG data to assess the formation of affect-specific representations of affective scene in the brain and their stability. EEG and fMRI data were integrated to assess the role of visual cortex in the large-scale recurrent network interactions underlying the sustained representation of affective scenes. Fusing EEG and fMRI data via representation similarity analysis (RSA) (Kriegeskorte et al., 2008), we further tested the timing of perceptual processing of affective scenes in areas along the visual hierarchy and compare that with the formation time of affect-specific representations.

2. Materials and methods

2.1. Participants

Healthy volunteers ($n = 26$) with normal or corrected-to-normal vision signed informed consent and participated in the experiment. Two participants withdraw before recording. Four additional participants were excluded for excessive movements inside the scanner. EEG and fMRI data from these four participants were not considered. Data from the remaining 20 subjects were analyzed and reported here (10 women; mean age: 20.4 ± 3.1).

These data have been published before (Bo et al., 2021) to address a different set of questions. In particular, in Bo et al. (2021), we asked the question of whether affective signals can be found in visual cortex. Analyzing fMRI, an affirmative answer was found when it was shown that pleasant, unpleasant, and neutral pictures evoked highly decodable neural representations in the entire retinotopic visual hierarchy. Using the late positive potential (LPP) and effective functional connectivity as indices of neural reentry we further argued that these affective representations are likely the results of feedback from anterior emotion-modulating structures such as the amygdala and the prefrontal cortex. In the present study we address the temporal dynamics of affective scene processing where the focus was placed on EEG decoding.

2.2. Procedure

2.2.1. The stimuli

The stimuli included 20 pleasant, 20 neutral and 20 unpleasant pictures from the International Affective Picture System (IAPS; Lang et al., 1997): Pleasant: 4311, 4599, 4610, 4624, 4626, 4641, 4658, 4680, 4694, 4695, 2057, 2332, 2345, 8186, 8250, 2655, 4597, 4668, 4693, 8030; Neutral: 2398, 2032, 2036, 2037, 2102, 2191, 2305, 2374, 2377, 2411, 2499, 2635, 2347, 5600, 5700, 5781, 5814, 5900, 8034, 2387; Unpleasant: 1114, 1120, 1205, 1220, 1271, 1300, 1302, 1931, 3030, 3051, 3150, 6230, 6550, 9008, 9181, 9253, 9420, 9571, 3000, 3069. The pleasant pictures included sports scenes, romance, and erotic couples and had average arousal and valence ratings of 5.8 ± 0.9 and 7.0 ± 0.5 , respectively. The unpleasant pictures included threat/attack scenes and bodily mutilations and had average arousal and valence ratings of 6.2 ± 0.8 and 2.8 ± 0.8 , respectively. The neutral pictures were images containing landscapes, adventures, and neutral humans and had average arousal and valence ratings of 4.2 ± 1.0 and 6.3 ± 1.0 , respectively. The arousal ratings for pleasant and unpleasant pictures are not significantly different ($p = 0.2$) but both are significantly higher than that of the neutral pictures ($p < 0.001$). Valence differences between unpleasant vs neutral ($p < 0.001$) and between pleasant vs neutral ($p = 0.005$) are both significant. Based on specific content, the 60 pictures can be further divided into 6 subcategories: disgust/mutilation body, attack/threat scene, erotic couple, happy people, neutral people, and adventure/nature scene. These subcategories provided an opportunity to examine the content-specificity of temporal processing of affective images.

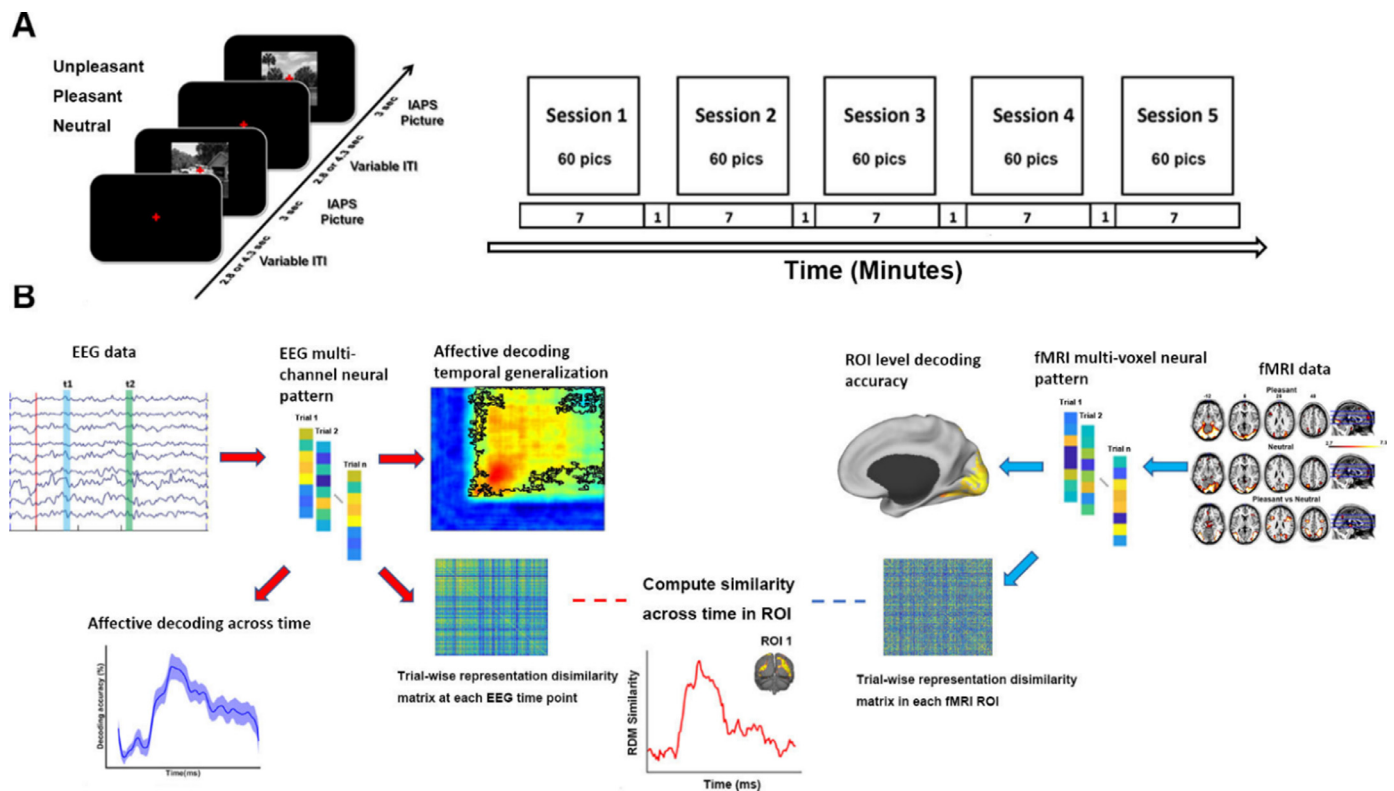


Fig. 1. Experimental paradigm and data analysis pipeline. (A) Affective picture viewing paradigm. Each recording session lasts seven minutes. 60 IAPS pictures including 20 pleasant, 20 unpleasant and 20 neutral pictures were presented in each session in random order. Each picture was presented at the center of screen for 3 s and followed by a fixation period (2.8 or 4.3 s). Participants were required to fixate the red cross at the center of the screen throughout the session while simultaneous EEG-fMRI was recorded. (B) Analysis pipeline illustrating the methods used at different stages of the analysis (see text for more details).

Two considerations went into the selection of the 60 pictures as stimuli in this study. First, these pictures are well characterized, and have been used in a body of research at the UF Center for the Study of Emotion and Attention as well as in previous work from our laboratories. The categories were not solely designated on the basis of normative ratings of valence and arousal, but also taken into account of the pictures' ability to engage emotional responses, as assessed by autonomic, EEG, and BOLD measures (Liu et al., 2012; Deweese et al., 2016; Thigpen et al., 2018; Tebbe et al., 2021). Second, we have used the same picture set previously in a number of studies where EEG LPPs and response times were recorded across several samples of participants (see, e.g., Thigpen et al., 2018), enabling us to benchmark the EEG data from inside the scanner against data recorded in an EEG lab outside the scanner, and to consider the impact of these pictures on modulating overt response time behavior, when interpreting the results of the present study.

2.2.2. The paradigm

The experimental paradigm was illustrated in Fig. 1A. There were five sessions. Each session contains 60 trials corresponding to the presentation of 60 different pictures. The order of picture presentation was randomized across sessions. Each IAPS picture was presented on a MR-compatible monitor for 3 s, followed by a variable (2800 ms or 4300 ms) interstimulus interval. The subjects viewed the pictures via a reflective mirror placed inside the scanner. They were instructed to maintain fixation on the center of the screen. After the experiment, participants rated the hedonic valence and emotional arousal level of 12 representative pictures (4 pictures for each broad category), which are not part of the 60-picture set, based on the paper and pencil version of the self-assessment manikin (Bradley and Lang, 1994; Bo et al., 2021).

2.3. Data acquisition

2.3.1. EEG data acquisition

EEG data were recorded simultaneously with fMRI using a 32 channel MR-compatible EEG system (Brain Products GmbH). Thirty-one sintered Ag/AgCl electrodes were placed on the scalp according to the 10–20 system with the FCz electrode serving as the reference. An additional electrode was placed on subject's upper back to monitor electrocardiogram (ECG); the ECG data was used during data preprocessing to assist in the removal of the cardiobalistic artifacts. EEG signal was recorded with an online 0.1–250 Hz band-pass filter and digitized to 16-bit at a sampling rate of 5 kHz. To ensure the successful removal of the gradient artifacts in subsequent analyses, the EEG recording system was synchronized with the scanner's internal clock throughout recording.

2.3.2. fMRI data acquisition

Functional MRI data were collected on a 3T Philips Achieva scanner (Philips Medical Systems). The recording parameters are as follows: echo time (TE), 30 ms; repetition time (TR), 1.98 s; flip angle, 80°; slice number, 36; field of view, 224 mm; voxel size, 3.5°3.5°3.5 mm; matrix size, 64°64. Slices were acquired in ascending order and oriented parallel to the plane connecting the anterior and posterior commissure. T1-weighted high-resolution structural images were also obtained.

2.4. Data preprocessing

2.4.1. EEG data preprocessing

The EEG data was first preprocessed using Brain Vision Analyzer 2.0 (Brain Products GmbH, Germany) to remove gradient and cardiobalistic artifacts. To remove gradient artifacts, an artifact template was

created by segmenting and averaging the data according to the onset of each volume and subtracted from the raw EEG data (Allen et al., 2000). To remove cardioballistic artifacts, ECG signal was low-pass-filtered, and the R peaks were detected as heart-beat events (Allen et al., 1998). A delayed average artifact template over 21 consecutive heart-beat events was constructed using a sliding-window approach and subtracted from the original signal. After gradient and cardioballistic artifacts were removed, the EEG data were lowpass filtered with the cut-off set at 50 Hz, downsampled to 250 Hz, re-referenced to the average reference, and exported to EEGLAB (Delorme and Makeig, 2004) for further analysis. The second-order blind identification (SOBI) procedure (Belouchrani et al., 1993) was performed to further correct for eye blinking, residual cardioballistic artifacts, and movement-related artifacts. The artifact-corrected data were then lowpass filtered at 30Hz and epoched from -300 ms to 2000 ms with 0ms denoting picture onset. The prestimulus baseline was defined to be -300 ms to 0 ms.

2.4.2. fMRI data preprocessing

The fMRI data were preprocessed using SPM (<http://www.fil.ion.ucl.ac.uk/spm/>). The first five volumes from each session were discarded to eliminate transient activity. Slice timing was corrected using interpolation to account for differences in slice acquisition time. The images were then corrected for head movements by spatially realigning them to the sixth image of each session, normalized and registered to the Montreal Neurological Institute (MNI) template, and resampled to a spatial resolution of 3mm by 3mm by 3mm. The transformed images were smoothed by a Gaussian filter with a full width at half maximum of 8 mm. The low frequency temporal drifts were removed from the functional images by applying a high-pass filter with a cutoff frequency of 1/128 Hz.

2.5. MVPA analysis: EEG data

2.5.1. EEG decoding

MVPA analysis was done using support vector machine (SVM) implemented in Matlab 2014 LIBSVM toolbox (Chang and Lin, 2011). To reduce noise and increase decoding robustness, 5 consecutive EEG data points (no overlap) were averaged, resulting in a smoothed EEG time series with a temporal resolution of 20 ms (50 Hz). Unpleasant vs neutral scenes and pleasant vs neutral scenes were decoded within each subject at each time point to form a decoding accuracy time series. Each trial of the EEG data (100 trials for each emotion category) was treated as a sample for the classifier. The 31 EEG channels provided 31 features for the SVM classifier. A ten-fold cross validation approach was applied. The weight vector or weight map from the classifier was transformed according to Haufe et al. (2014) and its absolute value is visualized as a topographical map to assess the importance of each channel in terms of its contribution to the decoding performance between affective and neutral pictures.

2.5.2. Temporal generalization

The stability of the neural representations evoked by affective scenes was tested using a generalization across time (GAT) method (King and Dehaene, 2014). In this method, the classifier was not only tested on the data from the same time point at which it was trained, it was also tested on data from all other sample points, yielding a two-dimensional temporal generalization matrix. The decoding accuracy at a point on this plane (t_x, t_y) reflects the decoding performance at time t_x of the classifier trained at time t_y .

2.5.3. Statistical significance testing of EEG decoding and temporal generalization

Whether the decoding accuracy was above chance was evaluated by the Wilcoxon sign-rank test. Specifically, the decoding accuracy at each time point was tested against 50% (chance level). The resulting p value was corrected for multiple comparisons by controlling for the

false discovery rate (FDR, $p < 0.05$) across the time course. A further requirement to reduce possible false positives is that the significance cluster contains at least five consecutive such sample points.

The decoding accuracy was expected to be at chance level prior to and immediately after picture onset. The time at which decoding accuracy rose above chance level was taken to be the time when the affect-specific neural representations of affective scenes formed. The statistical significance of the difference between the onset times of above-chance-decoding for different decoding accuracy time series was evaluated by a bootstrap resample procedure. Each resample consisted of randomly picking 20 sample decoding accuracy time series from 20 subjects with replacement and above-chance decoding onset was determined for this resample. The procedure was repeated 1000 times and the onset times from all the resamples formed a distribution. The significant difference between two such distributions was assessed by the two-sample Kolmogorov-Smirnov test.

To test the statistical significance of temporal generalization, we conducted Wilcoxon sign-rank test at each pixel in the temporal generalization map the decoding accuracy against 50% (chance level). The corresponding p value is corrected for multiple comparisons according to FDR $p < 0.05$. Cluster size is a further control (>10 points).

2.6. MVPA analysis: fMRI data

The picture-evoked BOLD activation was estimated on a trial-by-trial basis using the beta series method (Mumford et al., 2012). In this method, the trial of interest was represented by a regressor, and all the other trials were represented by another regressor. Six motion regressors were included to account for any movement-related artifacts during scanning. Repeating the process for all the trials we obtained the BOLD response to each picture presentation in all brain voxels. The single-trial voxel patterns evoked by pleasant, unpleasant, and neutral pictures were decoded between pleasant and neutral as well as between unpleasant and neutral using a ten-fold validation procedure within the retinotopic visual cortex defined according to a recently published probabilistic visual retinotopic atlas (Wang et al., 2015). Here the retinotopic visual cortex consisted of V1v, V1d, V2v, V2d, V3v, V3d, V3a, V3b, hV4, hMT, VO1, VO2, PHC1, PHC2, LO1, LO2, and IPS. For some analyses, the voxels in all these regions were combined to form a single ROI called visual cortex, whereas for other analyses, these regions were divided into early, ventral, and dorsal visual cortex (see below).

2.7. Fusing EEG and fMRI data via RSA

Decoding between affective scenes vs neutral scenes, as described above, yields information on the formation and dynamics of affect-specific neural representations. For comparison purposes, we also obtained the onset time of perceptual or sensory processing of affective images in visual cortex, which is expected to precede the formation of affect-specific representations, by fusing EEG and fMRI data via representation similarity analysis (RSA) (Kriegeskorte et al., 2008). RSA is a multivariate method that assesses the representational similarity (e.g., using cross correlation) evoked by a set of stimuli and expresses the result as a representational dissimilarity matrix (1- cross correlation matrix) (RDM). Correlating the fMRI-based RDMs from different ROIs and the EEG-based RDMs from different time points, one can obtain the spatiotemporal profile of information processing in the brain.

In the current study, for each trial, 31 channels of EEG data at a given time point provided a 31-dimensional feature vector, which was correlated with the 31-dimensional feature vector from another trial at the same time point. For all 300 trials (60 trials per session \times 5 sessions) a 300×300 representational dissimilarity matrix (RDM) was constructed at each time point. For fMRI data, following the previous work (Bo et al., 2021), we divided the visual cortex into three ROIs: early (V1v, V1d, V2v, V2d, V3v, V3d), ventral (VO1, VO2, PHC1, PHC2), and dorsal (IPS0-5) visual cortex. For each ROI, the fMRI feature vector was

extracted from each trial and correlated with the fMRI feature vector from another trial, yielding a 300×300 RDM for the ROI. To fuse EEG and fMRI, a correlation between the EEG-based RDM at each time point and the fMRI-based RDM from a ROI was computed, and the result was the representational similarity time course for the ROI. This procedure was carried out at single subject level first and then averaged across subjects.

We note that in our study, since EEG and fMRI were simultaneously recorded, there is trial-to-trial correspondence between EEG and fMRI, which makes single trial RSA analysis possible. Single trial level RDMs, by containing more variability, may enhance the sensitivity of the RSA fusion analysis. In most previous RSA studies fusing MEG/EEG and fMRI (e.g., Cichy et al., 2014; Muukkonen et al., 2020), the single trial-based RSA analysis is not possible, because MEG/EEG and fMRI were recorded separately and there was no trial-to-trial correspondence between the two types of recordings. In those situations, the only available option was to average trials from the same exemplar or experimental condition and construct RDM matrices whose dimension equals the number of exemplars or experimental conditions.

To assess the onset time of significant similarity between EEG RDM and fMRI RDM, we first computed the mean and standard deviation of the similarity measure during the baseline period (-300 ms to 0 ms). Along the representational similarity time course, similarity measures that are five standard deviations above the baseline mean were considered statistically significant ($p < 0.003$). To further control for multiple comparisons, clusters containing fewer than five consecutive such time points were discarded. For a given ROI, the first time point that meets the above significance criteria was considered the onset time for perceptual or sensory processing for that ROI. To statistically compare the onset times from different ROIs, we conducted a bootstrap resample procedure. Each resample consisted of randomly picking 20 sample RDM similarity time series from the 20 subjects with replacement and the onset time was determined for the resample. The procedure was repeated 1000 times and the onset times from all the resamples formed a distribution. The significant difference between distributions was then assessed by the two-sample Kolmogorov-Smirnov test.

3. Results

3.1. Affect-specific neural representations: formation onset time

We decoded multivariate EEG patterns evoked by pleasant, unpleasant, and neutral affective scenes and obtained the decoding accuracy time courses for pleasant-vs-neutral and unpleasant-vs-neutral. As shown in Fig. 2A, for pleasant vs neutral, above-chance level decoding began ~ 200 ms after stimulus onset, whereas for unpleasant vs neutral, the onset time of above-chance decoding was ~ 260 ms. Using a bootstrap procedure, the distributions of the onset times were obtained and shown in Fig. 2B, where the difference between the two distributions was evident, with pleasant-specific representations forming significantly earlier than that of unpleasant-specific representations (ks value = 0.87, effect size = 1.49, two-sample Kolmogorov-Smirnov test). To examine the contribution of different electrodes to the decoding performance, Fig. 2C shows the classifier weight maps at the indicated times. These weight maps suggested that neural activities that contributed to classifier performance was mainly located in occipital-temporal channels, in agreement with prior studies using fMRI where enhanced and/or decodable BOLD activities evoked by affective scenes was observed in visual cortex and temporal structures (Sabatinelli et al., 2006; Sabatinelli et al., 2013; Bo et al., 2021).

Given that above-chance decoding started ~ 200 post picture onset, it is unlikely that the decoding results were driven by low-level visual features, which would have entailed earlier above-chance decoding time (e.g., ~ 100 ms). To firm up this notion, we further tested if there are systematic low level visual feature differences across emotion categories. Low level visual features were extracted by GIST using a method from a

previous publication (Khosla et al., 2012). We hypothesized that if GIST features depend on category labels, we should be able to decode between different categories based on these features. A SVM classification analysis was applied to image-based GIST features, and the decoding accuracy is at chance level: pleasant vs neutral is 49% ($p = 0.9$, random permutation test) and unpleasant vs neutral is 52.5% ($p = 0.8$, random permutation test). These results suggest that the decoding results in Fig. 2 are not likely to be driven by low-level visual features.

Dividing the scenes into 6 subcategories: erotic couple, happy people, mutilation body/disgust, attack, nature scene/adventure, and neutral people, we further decoded multivariate EEG patterns evoked by these subcategories of images. Against neutral people, the onset times of above-chance decoding for erotic couple, attack, and mutilation body/disgust were ~ 180 ms, ~ 280 ms, and ~ 300 ms, respectively, with happy people not significantly decoded from neutral people. The onset times were significantly different between erotic couple and attack with erotic couple being earlier (ks value = 0.81, effect size = 2.1), and between erotic couple and mutilation body/disgust with erotic couple being earlier (ks value = 0.92, effect size = 2.3). The onset times between attack and mutilated body/disgust were only weakly different with attack being earlier (ks value = 0.35, effect size = 0.34). Against natural scenes, the onset times of above-chance level decoding for erotic couple, attack, and mutilation body/disgust were ~ 240 ms, ~ 300 ms, and ~ 300 ms, respectively, with happy people not significantly decoded from natural scenes. The onset times were significantly different between erotic couple and attack with erotic couple being earlier (ks value = 0.7, effect size = 1.3) and between erotic and mutilation body/disgust with erotic couple being earlier (ks value = 0.87, effect size = 1.33); the onset timings were not significantly different between attack and mutilation body/disgust (ks value = 0.25, effect size = 0.25). Combining these data, for subcategories of affective scenes, the formation time of affect-specific neural representations appear to follow the temporal sequence: erotic couple \rightarrow attack \rightarrow mutilation body/disgust.

Affective pictures are characterized along two dimensions: valence and arousal. We tested to what extent these factors influenced the decoding results. Erotic (arousal: 6.30, valence: 6.87) and Disgust/Mutilation (arousal: 6.00, valence: 2.18) pictures have similar arousal ($p = 0.76$) but significantly different valence ($p < 0.001$). As shown in Fig. 3A, the decoding accuracy between these two subcategories rose above chance level ~ 200 ms after picture onset, suggesting that the patterns evoked by affective scenes to a large extent reflect valence. In contrast, natural scenes/adventure (arousal: 5.4, valence: 7.0) and neutral people (arousal: 3.5, valence: 5.5) have significantly different arousal ratings ($p = 0.05$), but the two subcategories cannot be decoded, as shown in Fig. 3B, suggesting that arousal is not a very strong factor driving decodability.

3.2. Affect-specific neural representations: temporal stability

How do affect-specific neural representations, once formed, evolve over time? A serial processing model, in which neural processing progresses from one brain region to the next, would predict that the representations will evolve dynamically, resulting in a temporal generalization matrix as schematically shown in Fig. 4A Left. In contrast, a recurrent processing model, in which the representations are undergirded by the recurrent interactions among different brain regions, would predict sustained neural representations, resulting in a temporal generalization matrix as schematically shown in Fig. 4A Right. We applied a temporal generalization method called the generalization across time (GAT) to test these possibilities. A classifier was trained on data recorded at time t_y and tested on data at time t_x . The decoding accuracy is then displayed as a color-coded two-dimensional function (called the temporal generalization matrix) on the plane spanned by t_x and t_y . As can be seen in Fig. 4B, a stable neural representation emerged ~ 200 ms after picture onset and remained stable as late as 2000 ms post stimulus onset, with the peak decoding accuracy occurring within the time in-

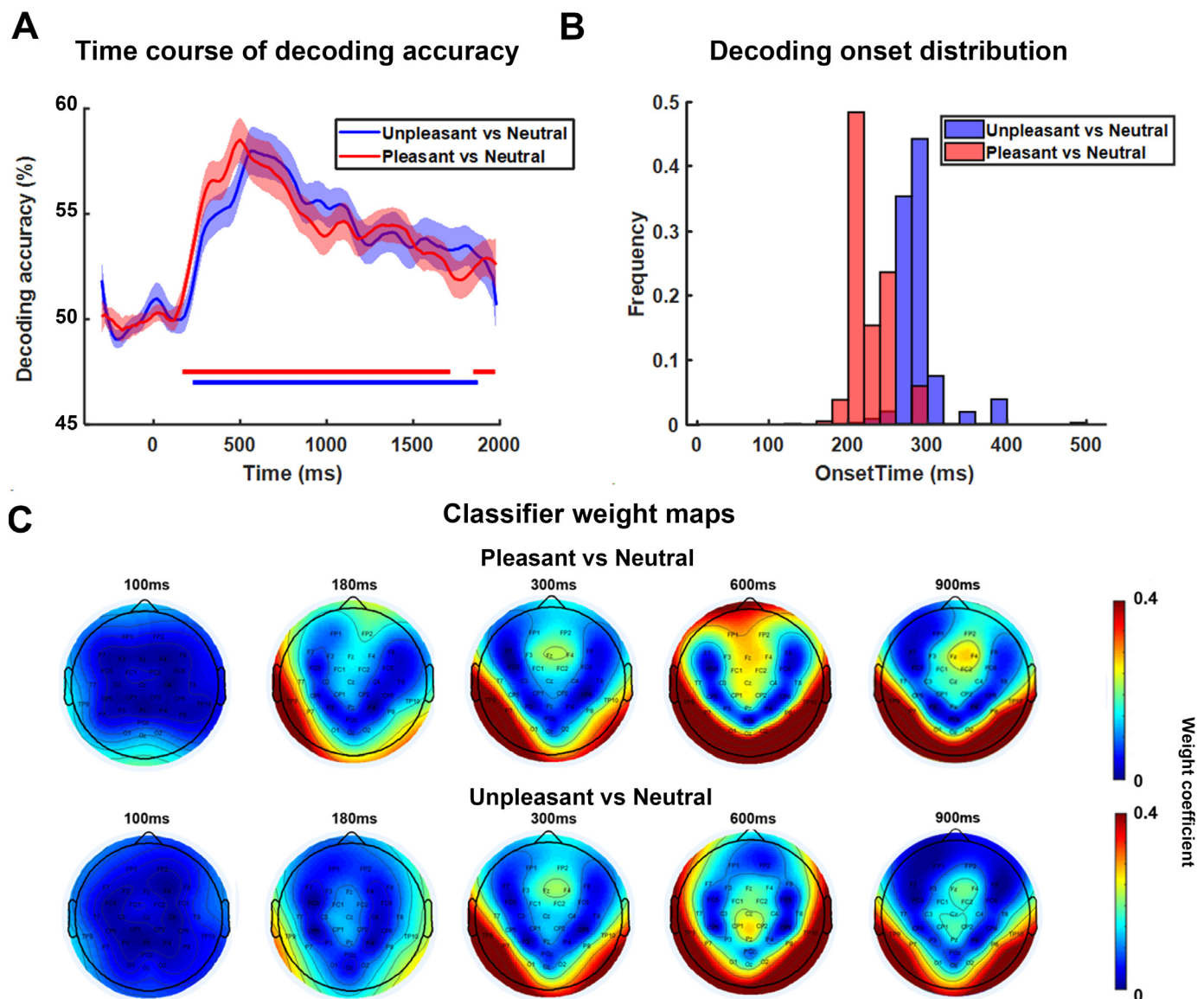


Fig. 2. Decoding EEG data between affective and neutral scenes across time. (A) Decoding accuracy time courses. (B) Bootstrap distributions of above-chance decoding onset times. Subjects are randomly selected with replacement and onset time was computed for each bootstrap resample (a total of 1000 resamples were considered). (C) Weight maps showing the contribution of different channels to decoding performance at different times.

interval 300–800 ms. Although the decoding accuracy decreased after the peak time, it remained significantly above chance, as shown by the large area within the black contour. These results demonstrate that the affect-specific neural representations of affective scenes, whether pleasant or unpleasant, are stable and sustained over extended periods of time, suggesting that affective scene processing could be supported by recurrent interactions in the engaged neural circuits. Repeating the same temporal generalization analysis for emotional subcategories, as shown in Fig. 5, we observed similar stable neural representations for each emotion subcategory.

3.3. Visual cortical contributions to sustained affective representations

Weight maps in Fig. 2 suggest that occipital and temporal structures are the main neural substrate underlying affect-specific neural representations, which is in line with previous studies showing patterns of visual cortex activity encoding rich, category-specific emotion representations (Kragel et al., 2019; Bo et al., 2021). Whether these structures participate in the recurrent interactions that give rise to sustained neu-

ral representations of affective scenes was the question we considered next. Previous work, based on temporal generalization, has shown that cognitive operations such as attention, working memory, and decision-making are characterized by sustained neural representations, in which sensory cortex is an essential node in the recurrent network (Büchel and Friston, 1997; Gazzaley et al., 2004; Wimmer et al., 2015). We tested whether the same holds true in affective scene processing. It is reasonable to expect that if this is indeed the case, then the more stable and sustained the neural interactions (measured by the EEG temporal generalization), the more distinct the neural representations in visual cortex (measured by the fMRI decoding accuracy in visual cortex). Fig. 6A shows above-chance fMRI decoding accuracy for pleasant vs neutral ($p < 0.001$) and unpleasant vs neutral ($p < 0.001$) in visual cortex. We quantified the strength of the temporal generalization matrix by averaging the decoding accuracy inside the black contour (see Fig. 4B) and correlated this strength with the fMRI decoding accuracy in visual cortex. As shown in Fig. 6B, for unpleasant vs neutral decoding, there was a significant correlation between fMRI decoding accuracy in visual cortex and the strength of temporal generalization ($R = 0.66$, $p = 0.0008$),

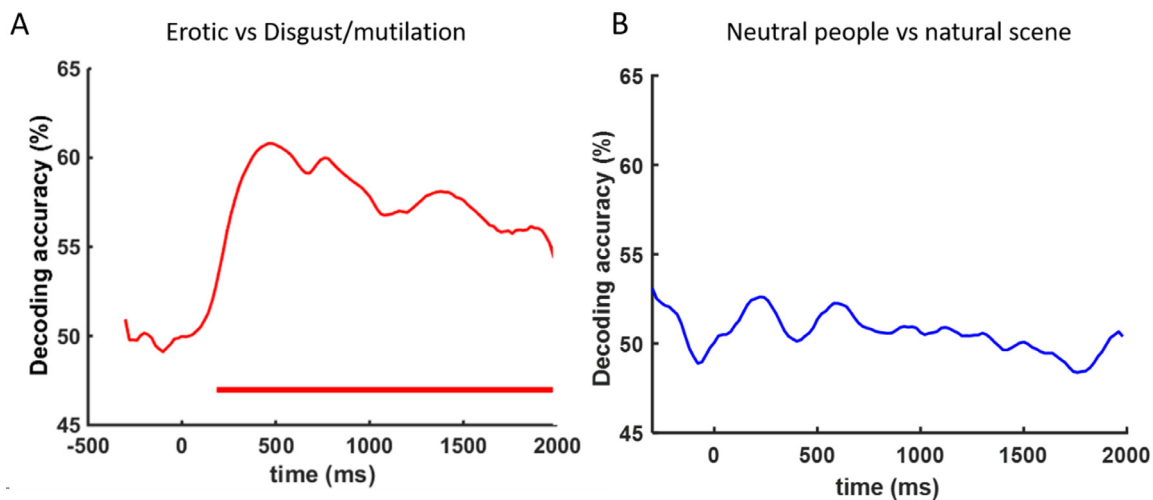


Fig. 3. Further decoding analysis testing the influence of valence vs arousal. **(A)** EEG decoding between Erotic (normative valence: 6.87, arousal: 6.30) vs Disgust/Mutilation pictures (normative valence: 2.18, arousal: 6.00). Red horizontal bar indicates period of above chance decoding (FDR $p < 0.05$). **(B)** EEG decoding between Neutral people (normative valence: 5.5, arousal: 3.5) vs Natural scenes/adventure (normative valence: 7.0, arousal: 5.4). Above chance level decoding is not found.

whereas for pleasant vs neutral decoding, the correlation is not as strong but is still marginally significant ($R = 0.32$, $p = 0.07$). Dividing subjects into high and low decoding accuracy group based on their fMRI decoding accuracies in the visual cortex, the corresponding temporal generalization for each group is shown in Fig. 6C, where it is again intuitively clear that temporal generalization is stronger in subjects with higher decoding accuracy in the visual cortex. Statistically, the strength of temporal generalization for unpleasant vs neutral was significantly larger in the high decoding accuracy group ($p = 0.01$) than the low accuracy group; the same was also observed for pleasant vs neutral but the statistical effect is again weaker ($p = 0.065$). We note that the method used here to quantify the strength of temporal generalization may be influenced by the level of decoding accuracy. In the Supplementary Materials we explored a different method of quantifying the strength of temporal generalization and obtained similar results (Fig. S5).

3.4. Onset time of perceptual processing of affective scenes

Past work has found that perceptual processing of simple visual objects begins ~ 100 ms after image onset in visual cortex (Cichy et al., 2016). This time is earlier than the onset time of affect-specific neural representations (~ 200 ms). Since the present study used complex visual scenes rather than simple visual objects as stimuli, it would be helpful to obtain information on the onset time of perceptual processing of these complex images, providing a reference for comparison. We fused simultaneous EEG-fMRI data using representational similarity analysis (RSA) (Cichy et al., 2016; Cichy and Teng, 2017) and computed the time at which visual processing of IAPS images began in visual cortex. Visual cortex was subdivided into early, ventral, and dorsal parts (see Methods). Their anatomical locations are shown in Fig. 7A. We found that shared variance between EEG recorded on the scalp and fMRI recorded from early visual cortex (EVC), ventral visual cortex (VVC), and dorsal visual cortex (DVC) began to exceed statistical significance level at ~ 80 ms, ~ 100 ms, and ~ 360 ms post picture onset, respectively, and remained significant until ~ 1800 ms; see Fig. 7B. These onset times are significantly different from one another according to the KS test applied to bootstrap generated onset time distributions: EVC < VVC (ks value = 0.21, effect size = 0.37), VVC < DVC (ks value = 0.75, effect size = 1.38), and EVC < DVC (ks test = 0.79, effect size = 1.79); see Fig. 7C.

An additional analysis was conducted to test the influence of low-level visual features on the RSA results (Groen et al., 2018; Grootswagers et al., 2020). Specifically, we computed partial correlation between EEG RDM and fMRI RDM while controlling for the effect of low-level feature RDM. Low level features were extracted by GIST using a method from a previous publication (Khosla et al., 2012). 300×300 GIST RDM was constructed in a similar way as EEG and fMRI RDMs. If GIST is an important factor driving the similarity between EEG RDM and fMRI RDM, it will have a significant contribution to EEG RDM-fMRI RDM correlation, and controlling for this contribution would reduce EEG RDM-fMRI RDM correlation. As can be seen, the results in Fig. 7D,E, where the partial correlation results are shown, are almost the same as Fig. 7B,C, suggesting that low-level features are not an important factor driving the RSA result.

Furthermore, we sought to examine if affect features are a factor driving the RSA result. A 300×300 emotion-category RDM was constructed. Specifically, if two trials belong to the same emotion category, the corresponding element in RDM is coded as '0,' otherwise it is coded as '1'. Fig. 7F showed that this categorical RDM becomes correlated with EEG RDM ~ 240 ms post picture onset, which agrees with the onset time of affect-specific representations from EEG decoding, suggesting that the EEG patterns beyond ~ 240 ms manifested the emotional content of affective scenes.

4. Discussion

We investigated the temporal dynamics of affective scene processing and reported four main observations. First, EEG patterns evoked by both pleasant and unpleasant scenes were distinct from those evoked by neutral scenes, with above-chance decoding occurring ~ 200 ms post image onset. The formation of pleasant-specific neural representations led that of unpleasant-specific neural representations by about 60 ms (~ 200 ms vs ~ 260 ms); the peak decoding accuracies were about the same (59% vs 58%). Second, dividing affective scenes into six subcategories, the onset of above-chance decoding between affective and neutral scenes followed the sequence: erotic couple (~ 210 ms) \rightarrow attack (~ 290 ms) \rightarrow mutilation body/disgust (~ 300 ms), suggesting that the speed at which neural representations form depends on specific picture content. Third, for both pleasant and unpleasant scenes, the neural representations were sustained rather than transient, and the stability of the representations was associated with the fMRI decoding accuracy in the visual cortex, sug-

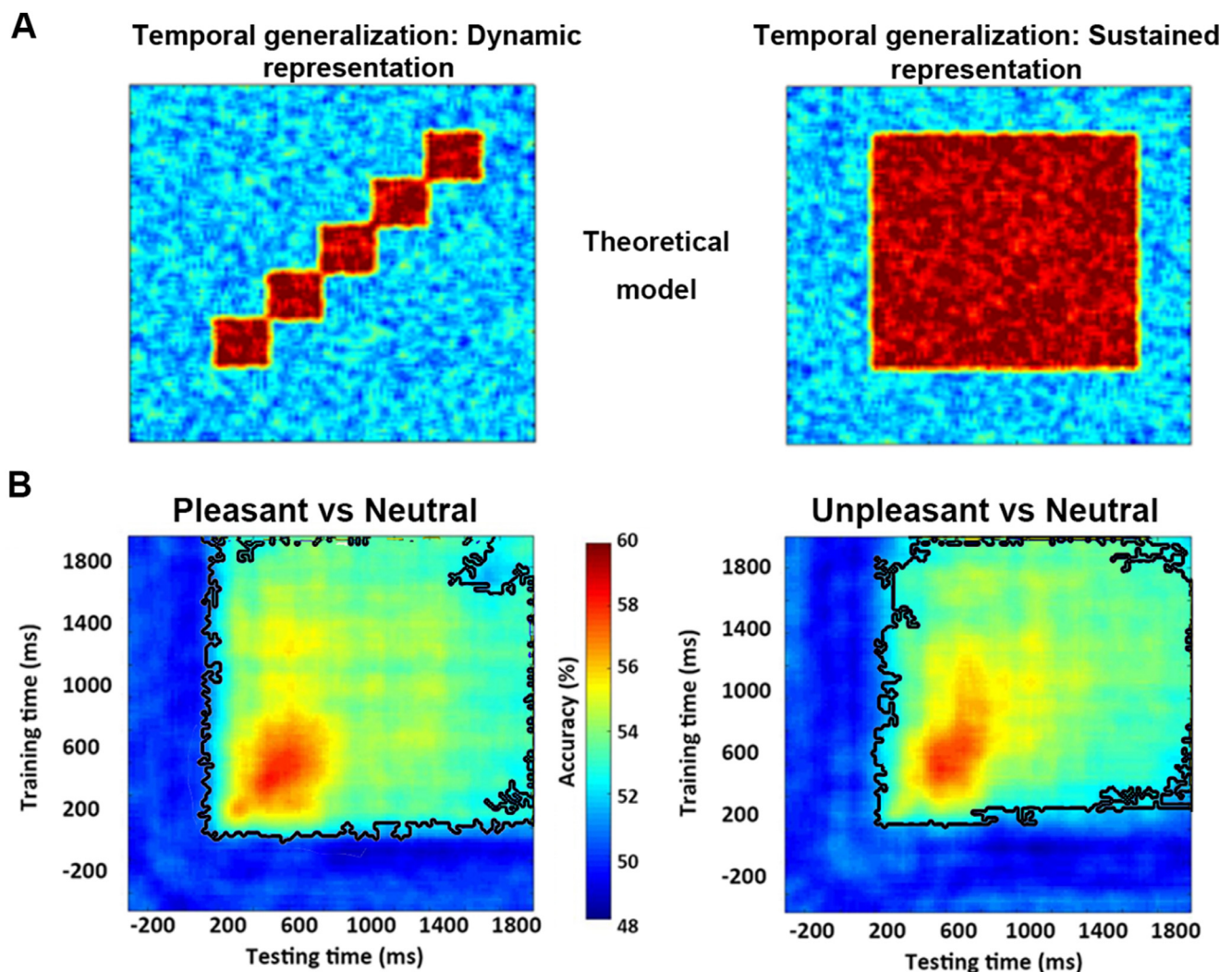


Fig. 4. Temporal generalization analysis. Classifier trained at each time point was tested on all other time points in the time series. The decoding accuracy at a point on this plane reflects the performance at time t_x of the classifier trained at time t_y . (A) Schematic temporal generalizations of dynamic or transient (Left) vs sustained or stable (Right) neural representations. (B) Temporal generalization for decoding between pleasant vs neutral (Left) and between unpleasant vs neutral (Right). Wilcoxon sign-rank test applied at each pixel in the temporal generalization map to test the significance of decoding accuracy against 50% (chance level). The corresponding p value is corrected for multiple comparisons according to FDR $p < 0.05$. Cluster size is further controlled (>10 points). Back contours enclose pixels with above chance decoding accuracy.

gesting, albeit indirectly, a role of visual cortex in the recurrent neural network that supports the affective representations. Fourth, applying RSA to fuse EEG and fMRI, perceptual processing of complex visual scenes was found to start in early visual cortex ~ 80 ms post image onset, preceding to ventral visual cortex at ~ 100 ms.

4.1. Formation of affect-specific neural representations

The question of how long it takes for affect-specific neural representations to form has been considered in the past. An intracranial electroencephalography study reported enhancement of gamma oscillations for emotional pictures compared to neutral pictures in occipital-temporal lobe in the time period of 200–1000 ms (Boucher et al., 2015). In our data, the ~ 200 ms onset of above-chance decoding and ~ 500 ms occurrence of peak decoding accuracy, with the main contribution to decoding performance coming from occipital and temporal electrodes, are consistent with the previous report. Compared to nonaffective images such as faces, houses and scenes, where decodable differences in neural representations in visual cortex started to emerge ~ 100 ms post stimulus onset with peak decoding accuracy occurring at ~ 150 ms (Cichy et al., 2016; Cauchoix et al., 2014), the formation times of these affect-

specific representations appear to be quite late. From a theoretical point of view, this delay may be explained by the reentry hypothesis which holds that anterior emotion regions such as the amygdala and the pre-frontal cortex, upon receiving sensory input, send feedback signals to visual cortex to enhance sensory processing and facilitate motivated attention (Lang and Bradley, 2010). In a recent fMRI study (Bo et al., 2021), we found that scenes expressing different affect can be decoded from multivoxel patterns in the retinotopic visual cortex and the decoding accuracy is correlated with the effective connectivity from anterior regions to visual cortex, in agreement with the hypothesis. What has not been established is how long it takes for the reentry signals to reach visual cortex. To provide a reference time for addressing this question, we fused EEG and fMRI data via RSA and found that sensory processing of complex visual scenes such as those contained IAPS pictures began ~ 100 ms post picture onset. This gave us an estimate of the reentry time which is on the order of ~ 100 ms or shorter. We caution that these estimates are somewhat speculative as our inferences are made rather indirectly.

Univariate ERP analysis, presented in the Supplementary Materials, was also carried to provide additional insights. Four groups of electrodes centered on Oz, Cz, Pz, and Fz were chosen as ROIs. ERPs evoked by affective pictures and neutral pictures were contrasted at each ROI. At

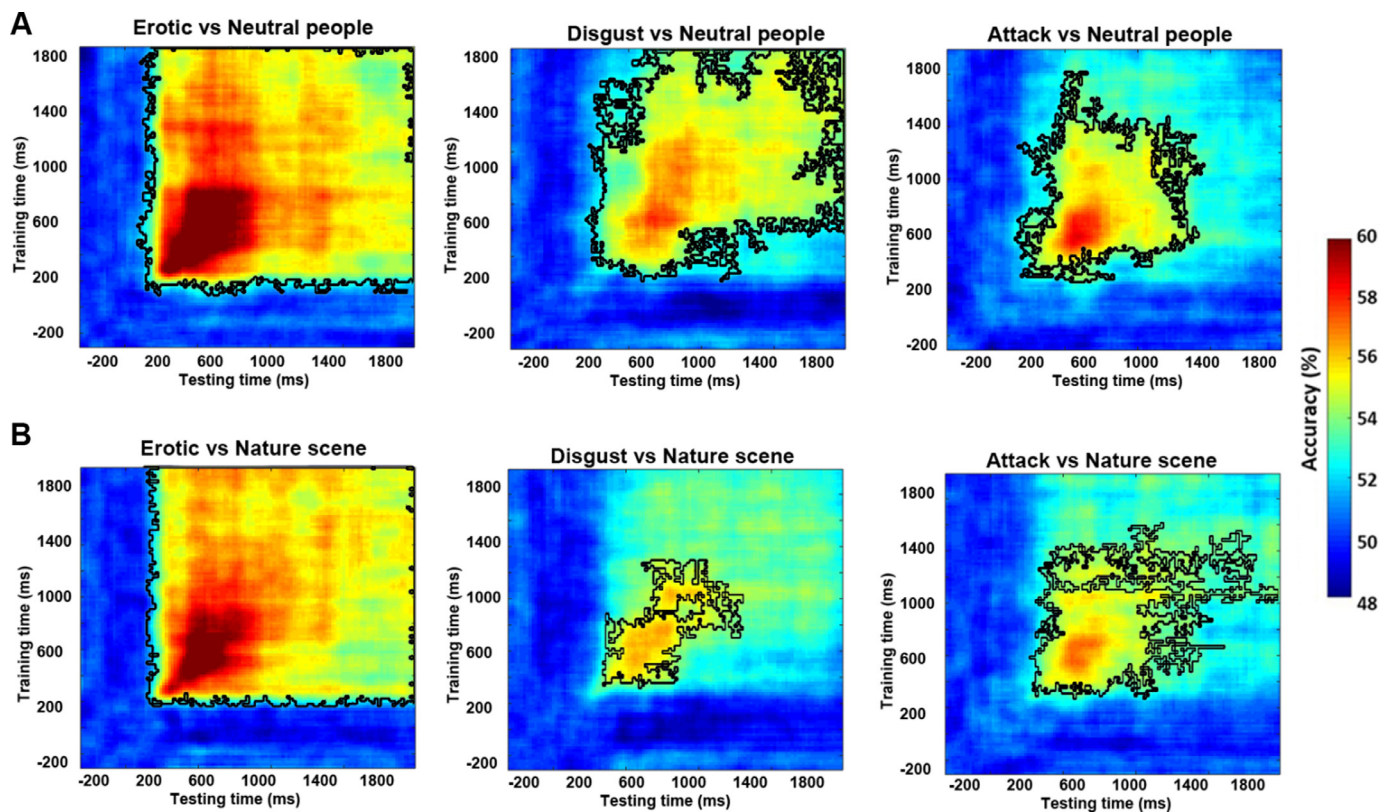


Fig. 5. Temporal generalization analysis for subcategories of affective scenes. (A) Decoding emotion subcategories against neutral people. (B) Decoding emotion subcategories against natural scenes. See Figure 4 for explanation of notations.

Cz, the difference ERP waves between pleasant vs neutral showed clear activation starting at ~ 172 ms, whereas for unpleasant vs neutral, the activation started at ~ 200 ms, both in general agreement with the timing information obtained from MVPA analysis.

The foregoing indicates that pleasant scenes evoked earlier affect-specific representations than unpleasant scenes. This positivity bias appears to be at variance with the negativity bias idea, which holds that negative events elicit more rapid and stronger responses compared to pleasant events (Rozin and Royzman, 2001; Vaish et al., 2008). While the idea has received support in behavioral data, e.g., subjects tend to locate unpleasant faces among pleasant distractors in shorter time than the reverse (Öhman et al., 2001), the neurophysiological support is mixed. Some studies using affective picture viewing paradigms reported shorter ERP latency and larger ERP amplitude for unpleasant pictures compared to pleasant ones in central P2 and late positive potential (LPP) (Carretié et al., 2001; Huang and Luo, 2006), but other ERP studies found that positive scene processing can be as strong and as fast as negative scene processing when examining early posterior negativity (EPN) in occipital channels (Schupp et al., 2006; Franken et al., 2008; Weinberg and Hajcak, 2010). One possible explanation for the discrepancy might be the choice of stimuli. The inclusion of exciting and sports images, which have high valence but average arousal, as stimuli in the pleasant category weakens the pleasant ERP effects when compared against threatening scenes included in the unpleasant category which have both low valence and high arousal (Weinberg and Hajcak, 2010). In the present work, by including images such as erotica and affiliative happy scenes in the pleasant category, which have comparable arousal ratings as images included in the unpleasant category, we were able to mitigate the possible issues associated with stimulus selection. Other explanations needed to be sought.

Subdividing the images into 6 subcategories: erotic couples, happy people, mutilation body/disgust, attack scene, neutral scene, and neu-

tral people, and decoding the emotion subcategories against the neutral subcategories, we found the following temporal sequence of formation of neural representations: erotic couple (pleasant) \rightarrow attack (unpleasant) \rightarrow mutilation body/disgust (unpleasant), with happy people failing to be decoded from neutral images. This finding can be seen as providing neural support to previous electrodermal findings showing that erotic scenes evoked largest responses within IAPS pictures, which was followed by mutilation and threat scenes (Sarlo et al., 2005), suggesting the temporal dynamic of emotion processing depends on specific scene content. It also supports a behavioral study that found a fast discrimination of erotic pictures compared to other categories, assessed using choice and simple response time experiments, using the same pictures as used here (Thigpen et al., 2018). In a neural study of nude body processing (Alho et al. 2015), the authors reported an early 100–200 ms nude-body sensitive response in primary visual cortex, which was maintained in a later period (200–300 ms). Their consistent occipitotemporal activation is comparable with our weight map analysis which implicates the occipitotemporal cortex as the main neural substrate sustaining the affective representations.

The faster discrimination between erotic scenes vs neutral people compared to erotic scenes vs natural scenes is worth discussing. One possibility is that the neutral people category has lower arousal ratings (3.458) compared to natural scenes (5.42) and arousal influences decodability. In addition, comparing discrimination performance and ERPs for pictures with no people versus pictures with people, Ihssen and Keil (2013) found no evidence that affective subcategories with people were better discriminated against subcategories with objects than subcategories with people. Instead, a face/portrait category was most rapidly discriminated when using a go/no-go format for responding. Despite the similarities, the exact mechanisms underlying our decoding findings, remain to be better understood.

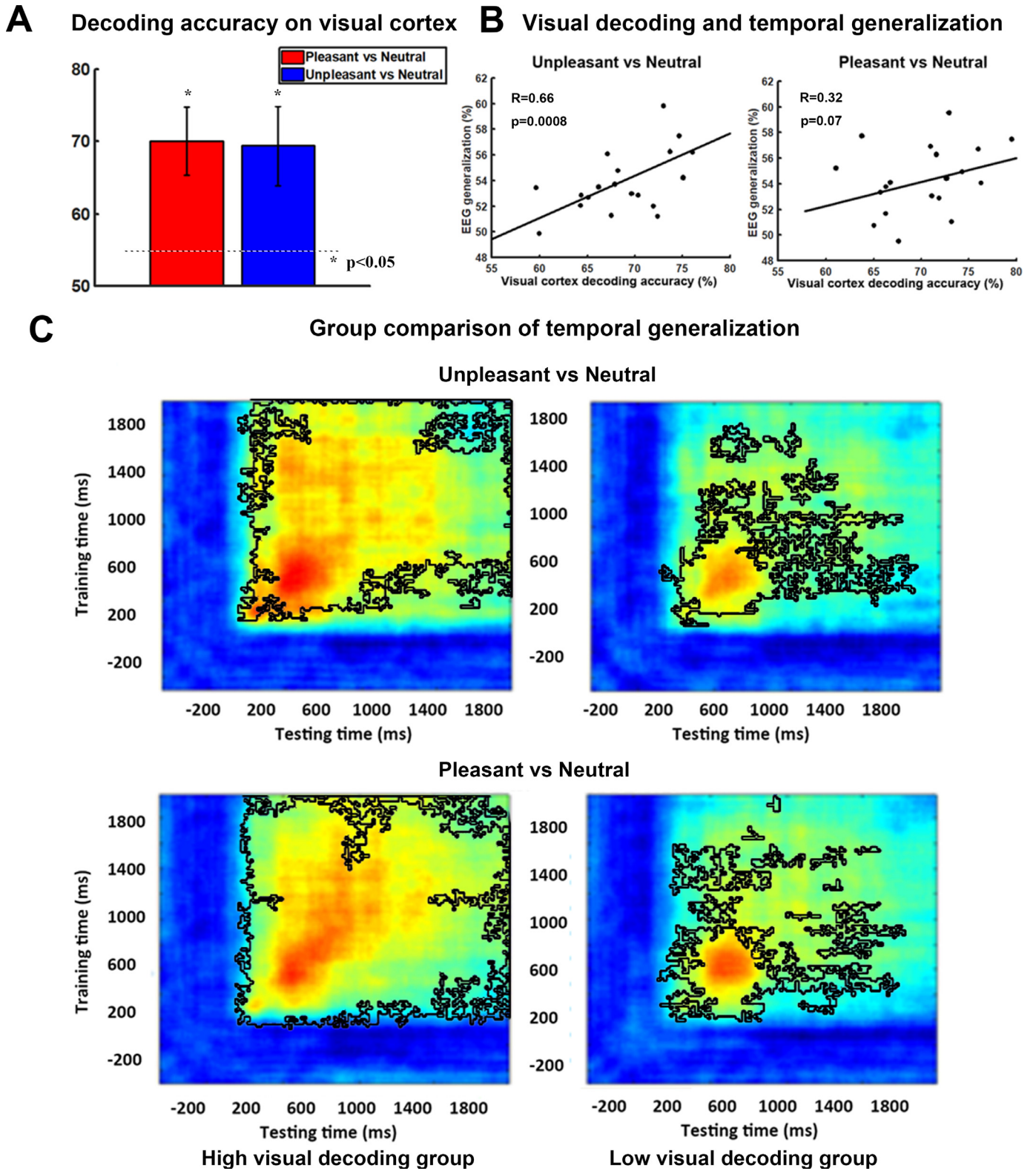


Fig. 6. Visual cortical contribution to stable representations of affect. **(A)** fMRI decoding accuracy in visual cortex. $p < 0.05$ threshold indicated by the dashed line. **(B)** Correlation between strength of EEG temporal generalization and fMRI decoding accuracy in visual cortex. **(C)** Subjects are divided into two groups according to their fMRI decoding accuracy in visual cortex. Temporal generalization for unpleasant vs neutral (Upper) and pleasant vs neutral (Lower) was shown for each group (high accuracy group on the Left vs low accuracy group on the Right). Black contours outline the statistically significant pixels ($p < 0.05$, FDR).

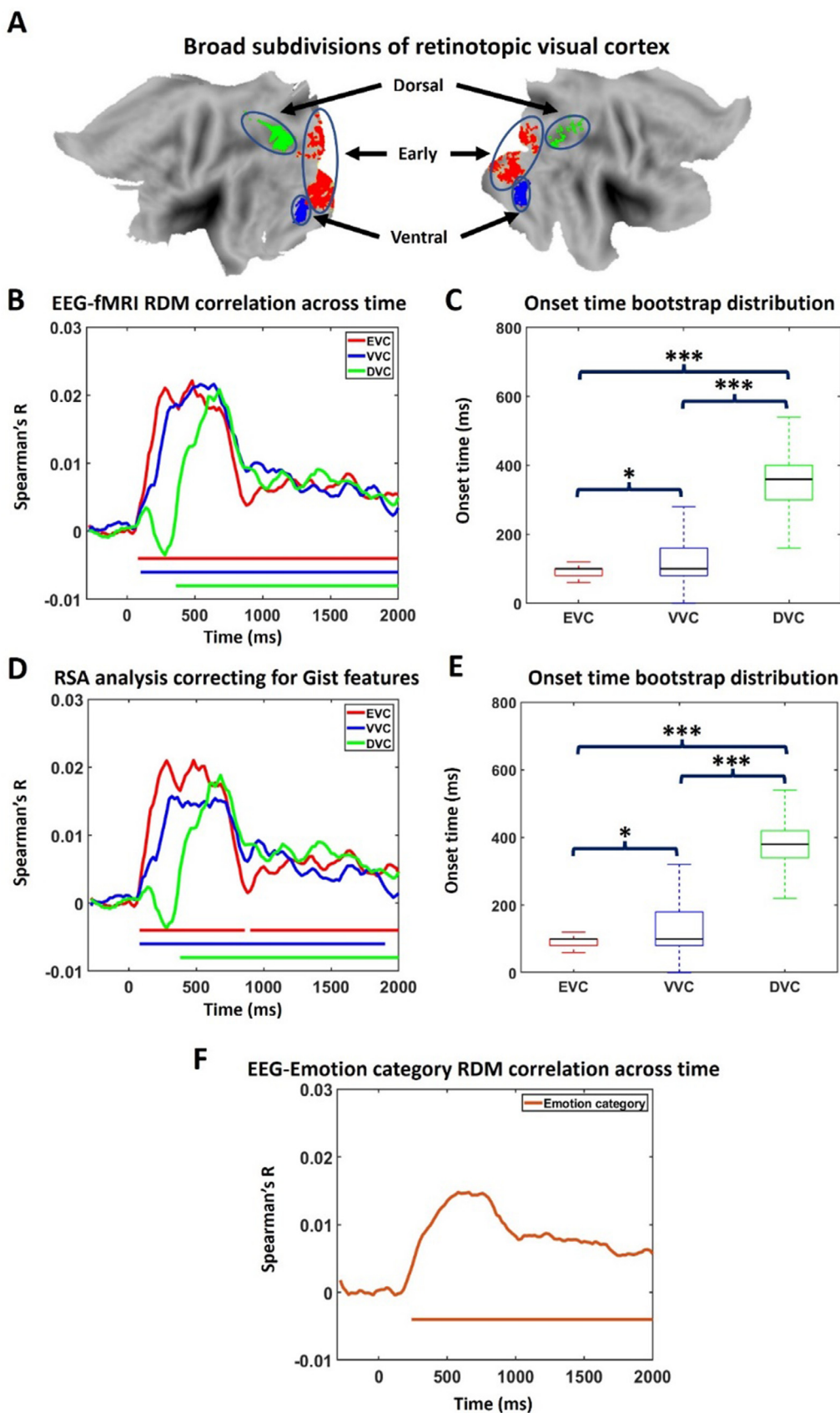


Fig. 7. Representational similarity analysis (RSA). (A) Regions of interest (ROIs): early visual cortex (EVC), ventral visual cortex (VVC), and dorsal visual cortex (DVC). (B) Similarity between EEG RDM and fMRI RDM across time for the three ROIs. Similarity larger than five baseline standard deviations for more than 5 consecutive time points are marked as statistically significant. (C) Onset time of significant similarity for each ROI in B. * Small effect size. ***Large effect size. (D) Partial correlation between EEG RDM and fMRI RDM with GIST RDM being set as control variable. (E) Onset time of significant similarity for each ROI in D. (F) Time course of similarity between EEG RDM and emotion category RDM.

4.2. Temporal evolution of neural representations of affective scenes

Once the affect-specific neural representations form, how do these representations evolve over time? If emotion processing is sequential, namely, if it progresses from one brain region to the next as time passes, we would expect dynamically evolving neural patterns. On the other hand, if the emotional state is stable over time undergirded by recurrent processing in distributed brain networks, we would expect a sustained neural pattern. A technique for testing these possibilities is the temporal generalization method (King and Dehaene, 2014). In this method, a classifier trained on data at one time is applied to decode data from all other times, resulting in a 2D plot of decoding accuracy called the temporal generalization matrix. Past studies decoding between non-emotional images such as neutral faces vs objects have found a transient temporal generalization pattern (Carlson et al., 2013; Cichy et al., 2014; Kaiser et al., 2016), supporting a sequential processing model for object recognition (Carlson et al., 2013). The temporal generalization results from our data revealed that the neural representations of affective scenes are stable over a wide time window (~200 ms to 2000 ms). Such stable representations may be maintained by sustained motivational attention, triggered by affective content (Schupp et al., 2004; Hajcak et al., 2009), which could in turn be supported by recurrent interactions between sensory cortex and anterior emotion structures (Keil et al., 2009; Sabatinelli et al., 2009; Lang and Bradley 2010). In addition, the time window in which sustained representations were found is broadly consistent with previous ERP studies where elevated LPP lasted multiple seconds, extending even beyond the offset of the stimuli (Foti and Hajcak, 2008; Hajcak et al., 2009).

4.3. Role of visual cortex in sustained neural representations of affective scenes

The visual cortex, in addition to its role in processing perceptual information, is also expected to play an active role in sustaining affective representations, because the purpose of sustained motivational attention is to enhance vigilance towards threats or opportunities in the visual environment (Lang and Bradley, 2010). The sensory cortex's role in sustained neural computations has been shown in other cognitive paradigms, including decision-making (Mostert et al., 2015), where stable neural representations are shown to be supported by the reciprocal interactions between prefrontal decision structures and sensory cortex. In face perception and imagery, neural representations are also found to be stable and sustained by communications between high and low order visual cortices (Dijkstra et al., 2018). In our data, two lines of evidence appear to support a sustained role of visual cortex in emotion representation. First, over an extended time period, the weight maps obtained from EEG classifiers were comprised of channels located mainly in occipital-temporal areas. Second, if the emotion-specific neural representations in the visual cortex stem from the recurrent processing within distributed networks, then the stronger and longer these interactions, the stronger and more distinct the affective representations in visual cortex. This is supported by the finding that the strength of temporal generalization is correlated with the fMRI decoding accuracy in visual cortex.

4.4. Temporal dynamic of sensory processing in visual pathway

The temporal dynamics of sensory processing of complex visual scenes can be revealed by fusing EEG-fMRI using RSA. The results showed that visual processing of IAPS images started ~80 ms post picture onset in early visual cortex (EVC) and proceeded to ventral visual cortex (VVC) at ~100 ms. It is instructive to compare this timing information with a previous ERP study where it is found that during the recognition of natural scenes, the low-level features are best explained by the ERP component occurring ~90 ms post picture onset while high-level features are best represented by the ERP component occurring ~170 ms after picture onset (Greene and Hansen, 2020). Compared with

the ~100 ms start time of perceptual processing in visual cortex, the ~200 ms formation onset of affect-specific neural representations likely includes the time it took for the reentry signals to travel from emotion processing structures such as the amygdala or the prefrontal cortex to the visual cortex (see below), which then give rise to the affect-specific representations seen in the occipital-temporal channels. The dorsal visual cortex (DVC), a brain region important for action and movement preparation (Wandell and Winawer, 2011), is activated at ~360 ms, which is relatively late and may reflect the processing of action predispositions resulting from affective perceptions. This sequence of temporal activity is consistent with that established previously using the fast-fMRI method where early visual cortex activation preceded ventral visual cortex activation which preceded dorsal visual cortex activation (Sabatinelli et al., 2014).

It is worth noting the RSA similarity time courses in all three visual ROIs stayed highly activated for a relatively long time period, which may be taken as further evidence, along with the temporal generalization analysis, to support sustained neural representations of affective scenes. From a methodological point of view, the RSA differs from the decoding analysis in that decoding analysis captures affect-specific distinction between neural representations, whereas the RSA fusing of EEG-fMRI is sensitive to evoked pattern similarity shared by EEG and fMRI imaging modalities, with early effects likely driven by sensory perceptual processing and late effects by both sensory and affective processing.

4.5. Beyond the visual cortex

The visual cortex is not the only brain region activated by affective scenes. In the Supplementary Materials, we performed a whole-brain decoding analysis of fMRI data (Fig. S1), and found above-chance decoding in many areas in prefrontal, limbic, as well as occipital-temporal cortices. Interestingly, the strongest decoding was found in the occipital-temporal areas, lending support to our focus on the visual cortex. Shedding light on the timing of these activations, a previous EEG source localization study reported that affect-related activation began to appear in visual cortex, prefrontal cortex and limbic systems ~200 ms after stimulus onset (Costa et al., 2014), complementing our fMRI analysis and the fMRI analysis by others (Saarimäki et al., 2016). Fusing EEG and fMRI with RSA, we further tested the temporal dynamics in several emotion-modulating structures, including amygdala, dACC, anterior insula, and fusiform cortex. As shown in Fig. S3, visual input reached the amygdala ~100 ms post picture onset, which is comparable with the activation time of early visual cortex. A similar activation time has been reported in a previous intracranial electrophysiological study (Méndez-Bértolo et al., 2016). Early activation was also found in dACC. Despite these early arrivals of visual input, it takes longer for affect-specific signals to arise, however. Recording from single neurons, Wang et al. showed that it takes ~250 ms for the emotional judgement signal of faces to emerge in the amygdala (Wang et al., 2014). It is intriguing to note that the ~100 ms difference between the arrival of visual input and the emergence of affect-specific activity is similar to our suggested reentry time of ~100 ms.

4.6. Limitations

This study is not without limitations. First, the suggestion that sustained affective representations are supported by recurrent neural interactions is speculative and based on indirect evidence, as we have already acknowledged above. Second, we used cross correlation to construct RDMs. A previous study has shown that a decoding-based analysis leads to more reliable RDMs (Guggenmos et al., 2018). Unfortunately, this method is not applicable to our data, because we do not have enough repetitions for each picture (five times) to permit a reliable decoding accuracy for every pair of pictures. Third, the inclusion of adventure scenes, which contain humans (small in size relative to the overall image), while providing a more relevant, interesting group of

scenes to help avoid that any decoding effects be solely due to the homogenous, low-interest neutral people in the neutral category of images, could complicate the animate-inanimate comparison.

4.7. Summary

We recorded simultaneous EEG-fMRI data from participants viewing affective pictures. Applying multivariate analyses including SVM and RSA, we found that perceptual processing of affective pictures began ~100 ms in visual cortex, whereas affect-specific representations began to form ~200 ms post image onset. The neural representations of affective scenes are sustained rather than dynamic and the visual cortex might be an important node in the recurrent network that supports these sustained representations.

Ethics statement

The experimental protocol was approved by the Institutional Review Board of the University of Florida. Written informed consent was obtained from all the participants.

Credit authorship contribution statement

Ke Bo: Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Methodology, Investigation, Conceptualization. **Lihan Cui:** Writing – review & editing, Formal analysis, Visualization, Methodology. **Siyang Yin:** Investigation. **Zhenhong Hu:** Conceptualization, Investigation. **Xiangfei Hong:** Conceptualization, Methodology. **Sungkean Kim:** Writing – review & editing, Visualization. **Andreas Keil:** Writing – review & editing, Supervision, Methodology. **Mingzhou Ding:** Writing – review & editing, Conceptualization, Supervision, Project administration, Funding acquisition, Methodology.

Acknowledgments

This work was supported by NIH grants R01 MH112558 and R01 MH125615. The authors declare no competing interests.

Data/code availability statement

Data has been uploaded to NIH Data Archive (https://nda.nih.gov/edit_collection.html?id=2645) and can be accessed by submitting requests to NIH Data Archive. The software and code (EEGLAB, SPM, Matlab 2014) used in the study are open-source and publicly available.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119532.

Reference

- Allen, P.J., Josephs, O., Turner, R., 2000. A method for removing imaging artifact from continuous EEG recorded during functional MRI. *Neuroimage* 12 (2), 230–239.
- Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., Lemieux, L., 1998. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *Neuroimage* 8 (3), 229–239.
- Alho, J., Salminen, N., Sams, M., Hietanen, J.K., Nummenmaa, L., 2015. Facilitated early cortical processing of nude human bodies. *Biol. Psychol.* 109, 103–110.
- Bae, G.Y., Luck, S.J., 2019. Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *Neuroimage* 184, 242–255.
- Belouchrani, A., Abed-Meraim, K., Cardoso, J.F., Moulines, E., 1993. May. Second-order blind separation of temporally correlated sources. In: *Proceedings of the International Conference on Digital Signal Processing*. Citeseer, pp. 346–351.
- ... & Bo, K., Yin, S., Liu, Y., Hu, Z., Meyyappan, S., Kim, S., Ding, M., 2021. Decoding neural representations of affective scenes in retinotopic visual cortex. *Cereb. Cortex* 31 (6), 3047–3063.
- Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25 (1), 49–59.

- Bradley, M.M., Costa, V.D., Ferrari, V., Codispoti, M., Fitzsimmons, J.R., Lang, P.J., 2015. Imaging distributed and massed repetitions of natural scenes: Spontaneous retrieval and maintenance. *Hum. Brain Mapp.* 36 (4), 1381–1392.
- ... & Boucher, O., D'Hondt, F., Tremblay, J., Lepore, F., Lassonde, M., Vannasing, P., Nguyen, D.K., 2015. Spatiotemporal dynamics of affective picture processing revealed by intracranial high-gamma modulations. *Hum. Brain Mapp.* 36 (1), 16–28.
- Büchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb. Cortex* 7 (8), 768–778 (New York, NY: 1991).
- Carretié, L., Mercado, F., Tapia, M., Hinojosa, J.A., 2001. Emotion, attention, and the 'negativity bias', studied through event-related potentials. *Int. J. Psychophysiol.* 41 (1), 75–85.
- Carlson, Tovar, Alink, A.D.A., Kriegeskorte, N., 2013. Representational dynamics of object vision: the first 1000 ms. *J. Vis.* 13 (10).
- Cauchois, M., Barragan-Jason, G., Serre, T., Barbeau, E.J., 2014. The neural dynamics of face detection in the wild revealed by MVPA. *J. Neurosci.* 34 (3), 846–854.
- Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* 2 (3), 1–27.
- Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. *Nat. Neurosci.* 17 (3), 455.
- Cichy, R.M., Pantazis, D., Oliva, A., 2016. Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex* 28 (8), 3563–3579. doi:10.1093/cercor/bhw135.
- Cichy, R.M., Teng, S., 2017. Resolving the neural dynamics of visual and auditory scene processing in the human brain: a methodological approach. *Philos. Trans. R. Soc. B Biol. Sci.* 372 (1714), 20160108.
- Contini, E.W., Wardle, S.G., Carlson, T.A., 2017. Decoding the time-course of object recognition in the human brain: from visual features to categorical decisions. *Neuropsychologia* 105, 165–176.
- Costa, T., Cauda, F., Crini, M., Tatu, M.K., Celegghin, A., de Gelder, B., Tamietto, M., 2014. Temporal and spatial neural dynamics in the perception of basic emotions from complex scenes. *Soc. Cogn. Affect. Neurosci.* 9 (11), 1690–1703.
- Cuthbert, B.N., Schupp, H.T., Bradley, M.M., Birbaumer, N., Lang, P.J., 2000. Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol. Psychol.* 52 (2), 95–111.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21.
- Deweese, M.M., Müller, M., Keil, A., 2016. Extent and time-course of competition in visual cortex between emotionally arousing distractors and a concurrent task. *Eur. J. Neurosci.* 43, 961–970.
- Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., Hillyard, S.A., 2002. Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp.* 15 (2), 95–111.
- Dijkstra, N., Mostert, P., de Lange, F.P., Bosch, S., van Gerven, M.A., 2018. Differential temporal dynamics during visual imagery and perception. *Elife* 7, e33904.
- Dima, D.C., Perry, G., Messaritaki, E., Zhang, J., Singh, K.D., 2018. Spatiotemporal dynamics in human visual cortex rapidly encode the emotional content of faces. *Hum. Brain Mapp.* 39 (10), 3993–4006.
- Franken, I.H., Muris, P., Nijs, I., van Strien, J.W., 2008. Processing of pleasant information can be as fast and strong as unpleasant information: implications for the negativity bias. *Neth. J. Psychol.* 64 (4), 168–176.
- Foti, D., Hajcak, G., 2008. Deconstructing reappraisal: descriptions preceding arousing pictures modulate the subsequent neural response. *J. Cogn. Neurosci.* 20 (6), 977–988.
- Foti, D., Hajcak, G., Dien, J., 2009. Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. *Psychophysiology* 46 (3), 521–530.
- Gazzaley, A., Rissman, J., D'Esposito, M., 2004. Functional connectivity during working memory maintenance. *Cogn. Affect. Behav. Neurosci.* 4 (4), 580–599.
- Greene, M.R., Hansen, B.C., 2020. Disentangling the independent contributions of visual and conceptual features to the spatiotemporal dynamics of scene categorization. *J. Neurosci.* 40 (27), 5283–5299.
- Groen, I.I., Greene, M.R., Fei-Fei, L., Beck, D.M., Baker, C.I., 2018. Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife*, 7 doi:10.7554/eLife.32962.
- Grootswagers, T., Kennedy, B.L., Most, S.B., Carlson, T.A., 2020. Neural signatures of dynamic emotion constructs in the human brain. *Neuropsychologia* 145, 106535. doi:10.1016/j.neuropsychologia.2017.10.016.
- Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* 29 (4), 677–697.
- Guggenmos, M., Sterzer, P., Cichy, R.M., 2018. Multivariate pattern analysis for MEG: a comparison of dissimilarity measures. *Neuroimage* 173, 434–447.
- Haufe, S., Meinecke, F., Gørgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110.
- Hajcak, G., Dunning, J.P., Foti, D., 2009. Motivated and controlled attention to emotion: time-course of the late positive potential. *Clin. Neurophysiol.* 120 (3), 505–510.
- Huang, Y.X., Luo, Y.J., 2006. Temporal course of emotional negativity bias: an ERP study. *Neurosci. Lett.* 398 (1–2), 91–96.
- Ihssen, N., Keil, A., 2013. Accelerative and decelerative effects of hedonic valence and emotional arousal during visual scene processing. *Q. J. Exp. Psychol.* 66 (7), 1276–1301.
- Kaiser, D., Azzalini, D.C., Peelen, M.V., 2016. Shape-independent object category responses revealed by MEG and fMRI decoding. *J. Neurophysiol.* 115 (4), 2246–2250.
- Keil, A., Bradley, M.M., Hauk, O., Rockstroh, B., Elbert, T., Lang, P.J., 2002. Large-scale neural correlates of affective picture processing. *Psychophysiology* 39 (5), 641–649.

- Keil, A., Sabatinelli, D., Ding, M., Lang, P.J., Ihssen, N., Heim, S., 2009. Re-entrant projections modulate visual cortex in affective perception: evidence from Granger causality analysis. *Hum. Brain Mapp.* 30 (2), 532–540.
- King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* 18 (4), 203–210.
- Khosla, A., Xiao, J., Torralba, A., Oliva, A., 2012. Memorability of image regions. In: *Proceedings of the Advances in Neural Information Processing Systems*, p. 25.
- Kragel, P.A., Reddan, M.C., LaBar, K.S., Wager, T.D., 2019. Emotion schemas are embedded in the human visual system. *Sci. Adv.* 5 (7), eaaw4358.
- Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1997. International affective picture system (IAPS): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention 1 (39–58), 3.
- Lang, P.J., Bradley, M.M., Fitzsimmons, J.R., Cuthbert, B.N., Scott, J.D., Moulder, B., Nangia, V., 1998. Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology* 35 (2), 199–210.
- Lang, P.J., Bradley, M.M., 2010. Emotion and the motivational brain. *Biol. Psychol.* 84 (3), 437–450.
- Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., Ding, M., 2012. Neural substrate of the late positive potential in emotional processing. *J. Neurosci.* 32 (42), 14563–14572.
- ... & Mendez-Bertolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martinez-Alvarez, R., Mah, Y.H., Strange, B.A., 2016. A fast pathway for fear in human amygdala. *Nat. Neurosci.* 19 (8), 1041–1049.
- Mostert, P., Kok, P., De Lange, F.P., 2015. Dissociating sensory from decision processes in human perceptual decision making. *Sci. Rep.* 5, 18253.
- Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59 (3), 2636–2643.
- Muukkonen, I., Ölander, K., Numminen, J., Salmela, V.R., 2020. Spatio-temporal dynamics of face perception. *Neuroimage* 209, 116531.
- Nakamura, A., Kakigi, R., Hoshiyama, M., Koyama, S., Kitamura, Y., Shimojo, M., 1997. Visual evoked cortical magnetic fields to pattern reversal stimulation. *Cogn. Brain Res.* 6 (1), 9–22.
- Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10 (9), 424–430.
- Oya, H., Kawasaki, H., Howard, M.A., Adolphs, R., 2002. Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. *J. Neurosci.* 22 (21), 9502–9512.
- Öhman, A., Lundqvist, D., Esteves, F., 2001. The face in the crowd revisited: a threat advantage with schematic stimuli. *J. Personal. Soc. Psychol.* 80, 381–396.
- Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I., 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16 (2), 331–348.
- Rozin, P., Royzman, E.B., 2001. Negativity bias, negativity dominance, and contagion. *Personal. Soc. Psychol. Rev.* 5 (4), 296–320.
- ... & Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Nummenmaa, L., 2016. Discrete neural signatures of basic emotions. *Cereb. Cortex* 26 (6), 2563–2573.
- Sabatinelli, D., Lang, P.J., Keil, A., Bradley, M.M., 2006. Emotional perception: correlation of functional MRI and event-related potentials. *Cereb. Cortex* 17 (5), 1085–1091.
- Sabatinelli, D., Lang, P.J., Bradley, M.M., Costa, V.D., Keil, A., 2009. The timing of emotional discrimination in human amygdala and ventral visual cortex. *J. Neurosci.* 29 (47), 14864–14868.
- Sabatinelli, D., Keil, A., Frank, D.W., Lang, P.J., 2013. Emotional perception: correspondence of early and late event-related potentials with cortical and subcortical functional MRI. *Biol. Psychol.* 92 (3), 513–519.
- Sabatinelli, D., Frank, D.W., Wanger, T.J., Dhamala, M., Adhikari, B.M., Li, X., 2014. The timing and directional connectivity of human frontoparietal and ventral visual attention networks in emotional scene perception. *Neuroscience* 277, 229–238.
- Sutterer, D.W., Coia, A.J., Sun, V., Shevell, S.K., Awh, E., 2021. Decoding chromaticity and luminance from patterns of EEG activity. *Psychophysiology* 58 (4), e13779.
- Sarlo, M., Palomba, D., Buodo, G., Minghetti, R., Stegagno, L., 2005. Blood pressure changes highlight gender differences in emotional reactivity to arousing pictures. *Biol. Psychol.* 70 (3), 188–196.
- Schupp, H., Cuthbert, B., Bradley, M., Hillman, C., Hamm, A., Lang, P., 2004. Brain processes in emotional perception: Motivated attention. *Cogn. Emot.* 18 (5), 593–611.
- Schupp, H.T., Flaisch, T., Stockburger, J., Junghöfer, M., 2006. Emotion and attention: event-related brain potential studies. *Prog. Brain Res.* 156, 31–51.
- Tebbe, A.-L., Friedl, W.M., Alpers, G.W., Keil, A., 2021. Effects of affective content and motivational context on neural gain functions during naturalistic scene perception. *Eur. J. Neurosci.* 53 (10), 3323–3340.
- Thigpen, N.N., Keil, A., Freund, A.M., 2018. Responding to emotional scenes: effects of response outcome and picture repetition on reaction times and the late positive potential. *Cogn. Emot.* 32, 24–36.
- Vaish, A., Grossmann, T., Woodward, A., 2008. Not all emotions are created equal: the negativity bias in social-emotional development. *Psychol. Bull.* 134 (3), 383.
- Wang, L., Mruczek, R.E., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. *Cereb. Cortex* 25 (10), 3911–3931.
- Wang, S., Tudusciuc, O., Mamelak, A.N., Ross, I.B., Adolphs, R., Rutishauser, U., 2014. Neurons in the human amygdala selective for perceived emotion. *Proc. Natl. Acad. Sci.* 111 (30), E3110–E3119.
- Wandell, B.A., Winawer, J., 2011. Imaging retinotopic maps in the human brain. *Vis. Res.* 51 (7), 718–737.
- Weinberg, A., Hajcak, G., 2010. Beyond good and evil: the time-course of neural activity elicited by specific picture content. *Emotion* 10 (6), 767.
- Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., De La Rocha, J., 2015. Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nat. Commun.* 6 (1), 1–13.
- Wolff, M.J., Ding, J., Myers, N.E., Stokes, M.G., 2015. Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* 9, 123.

Further reading

- Hajcak, G., MacNamara, A., Olvet, D.M., 2010. Event-related potentials, emotion, and emotion regulation: an integrative review. *Dev. Neuropsychol.* 35 (2), 129–155.
- Junghöfer, M., Weike, A.I., Stockburger, J., Hamm, A.O., 2004. The facilitated processing of threatening faces: an ERP analysis. *Emotion* 4 (2), 189.
- King, J.R., Gramfort, A., Schurger, A., Naccache, L., Dehaene, S., 2014. Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS One* 9 (1), e85791.
- Stokes, M.G., Wolff, M.J., Spaak, E., 2015. Decoding rich spatial information with high temporal resolution. *Trends Cogn. Sci.* 19 (11), 636–638.