

Decoding the temporal representation of facial expression in face-selective regions

Zhihao Zhang^{a,b}, Tong Chen^{c,d}, Ye Liu^{a,b}, Chongyang Wang^e, Ke Zhao^{a,b,*}, Chang Hong Liu^f, Xiaolan Fu^{a,b,*}

^a State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

^b Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

^c Chongqing Key Laboratory of Non-Linear Circuit and Intelligent Information Processing, Southwest University, Chongqing 400715, China

^d Chongqing Key Laboratory of Artificial Intelligence and Service Robot Control Technology, Chongqing 400715, China

^e Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

^f Department of Psychology, Bournemouth University, Dorset, United Kingdom

ARTICLE INFO

Keywords:

Facial expression
Magnetoencephalography (MEG)
Pairwise classification
Representational similarity analysis
CNN

ABSTRACT

The ability of humans to discern facial expressions in a timely manner typically relies on distributed face-selective regions for rapid neural computations. To study the time course in regions of interest for this process, we used magnetoencephalography (MEG) to measure neural responses participants viewed facial expressions depicting seven types of emotions (happiness, sadness, anger, disgust, fear, surprise, and neutral). Analysis of the time-resolved decoding of neural responses in face-selective sources within the inferior parietal cortex (IP-faces), lateral occipital cortex (LO-faces), fusiform gyrus (FG-faces), and posterior superior temporal sulcus (pSTS-faces) revealed that facial expressions were successfully classified starting from ~100 to 150 ms after stimulus onset. Interestingly, the LO-faces and IP-faces showed greater accuracy than FG-faces and pSTS-faces. To examine the nature of the information processed in these face-selective regions, we entered with facial expression stimuli into a convolutional neural network (CNN) to perform similarity analyses against human neural responses. The results showed that neural responses in the LO-faces and IP-faces, starting ~100 ms after the stimuli, were more strongly correlated with deep representations of emotional categories than with image level information from the input images. Additionally, we observed a relationship between the behavioral performance and the neural responses in the LO-faces and IP-faces, but not in the FG-faces and pSTS-faces. Together, these results provided a comprehensive picture of the time course and nature of information involved in facial expression discrimination across multiple face-selective regions, which advances our understanding of how the human brain processes facial expressions.

1. Introduction

Facial expressions signal important emotional information (Ekman, 1993). They are processed in a rapid and seemingly automatic fashion, as indexed by neural responses, for the regulation and facilitation of social interactions (Tracy and Robins, 2008). Haxby and colleagues (2000) proposed an influential hierarchical face perception model. They identified the posterior superior temporal sulcus (pSTS) along the lateral visual pathway (Pitcher and Ungerleider, 2021) as being responsible for processing changeable aspects of faces, such as expression, eye-gaze, and lip movements. The authors also indicated that the fusiform face

area (FFA), based on inputs from the occipital face area (OFA), within the ventral pathways, encodes invariant aspects, such as face identity. Subsequently, revised models of face perception have suggested that facial expressions are processed in FFA (Bernstein and Yovel, 2015) and in several distributed and interacting regions (Duchaine and Yovel, 2015), including the OFA, FFA, and pSTS. Additional evidence based on higher than chance-level classification accuracy has further confirmed the functional role of the OFA, FFA, and pSTS in facial expression processing (Liang et al., 2017).

Besides the ventral and lateral pathways, the inferior parietal (IP) cortex, which receives input from the occipitoparietal cortex (Kravitz

* Corresponding authors at: State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China.
E-mail addresses: zhaok@psych.ac.cn (K. Zhao), fu xl@psych.ac.cn (X. Fu).

et al., 2011; Pitcher and Ungerleider, 2021), also plays a role in facial expression and face recognition as part of the dorsal pathway. By analyzing the functional connectivity patterns of the right FFA, the left IP cortex was also considered to form the distributed cortical network model for face processing (Li et al., 2009). Baroni's study, utilizing intracranial electrodes to decode faces, discovered that, in addition to the superior temporal sulcus (STS) and FFA, certain electrodes in the bilateral IP were capable of successfully classifying faces (Baroni et al., 2017). Furthermore, two lesion studies have demonstrated that damage to the IP region directly affects facial expression recognition (Adolphs et al., 1996, 2000). Neuroimaging research has also revealed that the activation of the bilateral IP lobule is significantly stronger for facial expression processing than for gender recognition (Johnston et al., 2013; Sarkheil et al., 2013). Additionally, the gray matter volume of bilateral STS and the right IP was associated with facial expression decoding (L. Zhang et al., 2016). These studies suggest that the IP region is not only sensitive to faces but also contributes to processing facial expressions.

Despite these advances, the temporal properties of facial expression discriminating in each of these identified regions remain rather unclear. A meta-analysis suggested that the N170 amplitude is reliably modulated by emotional faces relative to neutral ones (Hinojosa et al., 2015). There is also evidence that the component reflects at least some form of facial structure processing instead of emotional concepts (Han et al., 2021). In contrast to univariate methods, multivariate pattern analysis (MVPA) considers the relationship among multiple variables (e.g., channels in M/EEG data and voxels in fMRI data), which can enhance the sensitivity of identifying differences among experimental conditions (Grootswagers et al., 2017; Norman et al., 2006). A few recent studies have employed more sensitive multivariate methods to investigate the temporal dynamics of face and expression processing (Ambrus et al., 2019; Dima et al., 2018; Li et al., 2022; Muukkonen et al., 2020; Smith and Smith, 2019). It has been reported that the neural representations of emotional faces appeared in the right temporalparietal regions and right frontal regions at 130 ms after the stimulus onset (Li et al., 2022). However, the spatial source of EEG decoding depends on input from multiple neighboring sensors, thus localization is limited by spatial resolution. By combining EEG and fMRI, Muukkonen et al. (2020) found that facial expression processing began in the primary visual cortex at approximately 130 ms, extending to the left fusiform face complex, lateral occipital, and temporal-parietal-occipital junction at 190 ms onwards. Because their correlation analysis of EEG and fMRI data included both emotional intensity and category of facial expression, the results may not reflect the spatiotemporal process of facial expression discrimination. Using whole-brain decoding of emotional expression in source space, Dima et al. (2018) discovered that information related to expression discrimination was traced from the visual cortex at 100 ms to higher-level temporal and frontal areas at 200–500 ms. The temporal process of expression decoding was averaged over time using 100 ms time window, which hindered more precise characterization. The contribution of expression decoding in each region was assessed based on the classifier weight instead of the decoding accuracy. However, multivariate classifiers may exhibit large weights at channels that do not pick up the signals of brain activity, as well as small weights at channels containing the signal (Haufe et al., 2014). Thus, a main purpose of our study was to investigate the precise temporal course of regions of interest in facial emotional discrimination.

Apart from the lack of knowledge about the temporal course of facial expression discrimination in the above-identified regions, we also know very little about the nature of information in these regions. In the field of computer vision, convolutional neural networks (CNNs) are among the best-performing models on object recognition, achieving human performance levels in object categorization (He et al., 2016; Krizhevsky et al., 2017). Like human vision, the features in each layer of CNNs exhibit the nature of hierarchical visual representation. Lower layers respond to local corners, edges, and color conjunctions, while higher

layers represent global, abstract, and class-specific information (LeCun et al., 2015; Zeiler and Fergus, 2014). By comparing fMRI data with a trained CNN at different hierarchical stages of human visual representations during object recognition, Cichy et al. (2016a) established that the low-level visual features are primarily extracted in the occipital lobe, whereas more abstract information is represented in the temporal and parietal lobes. Similarly, using a diverse set of models based on image properties and human perceived property models, Tsantani et al. (2021) have investigated the nature of information in the occipital and temporal lobes of face perception. The authors concluded that higher-level perceptual and social face information is encoded in the FFA, while low-level image-based properties are mainly represented in the OFA. In terms of the temporal dimension, previous studies have found that the FFA contains face-specific information around 50–75 ms and encodes individual-level face information invariant over facial expressions between 200 and 500 ms (Ghuman et al., 2014). On the other hand, the occipital and temporal regions primarily represent image-based properties within 100–200 ms and transition toward identity-based representation after 200–300 ms (Vida et al., 2017). However, there is currently no research available on the temporal properties about the nature of information encoded in the brain regions for the discrimination of facial expressions.

Hence, we aimed to address two key questions regarding the temporal representation of facial expressions to provide insights into the precise process of emotional faces information transmitting in the human brain. (1) When do different regions of face-selective regions begin to discriminate facial expressions? (2) What is the nature and time course of the information encoded in these regions, specifically whether it is low-level image-based or high-level category-based, and when does this encoding emerge? To answer these questions, we conducted MEG recordings from adult participants while they viewed facial expression images. The high temporal and spatial resolution of MEG makes it an ideal tool to uncover the spatiotemporal dynamics of brain activities underlying facial expression perception (Gross, 2019). We adopted a passive viewing design, which was commonly used in previous studies on face perception (Dima et al., 2018; Muukkonen et al., 2020). To investigate the time course of facial expression discrimination in face-selective brain regions, we used MVPA to compute time-resolved pairwise decoding accuracy for all possible pairs of facial expression. Additionally, we sought to establish a connection between neural and behavioral responses. This was achieved by comparing the pairwise structure of the neural data with behavioral judgments of the stimuli used in our MEG decoding task. We employed deep features from a lower layer and a higher layer of a fine-tuned CNN to construct image-based and category-based representations of the stimuli. We then evaluated the information encoded in the neural data by comparing its pairwise dissimilarity structure within each region to these representations. This process helped determine which representation—image-based or category-based—more closely resembled the neural data. A visual overview of this study workflow is shown in Fig. 1.

2. Methods

2.1. Participants

Twenty healthy volunteers were recruited to participate in the study and received monetary compensation for their involvement. One participant (sub11) was excluded due to their behavioral classification performance falling below 3 standard deviations from the mean. The other two participants (sub14 and sub15) were excluded due to excessive head motion, defined as exceeding 5 mm in at least 10 MEG blocks. Data from the remaining 17 participants (10 females, mean age 26, SD = 5.08) were used in all analyses throughout the study. All participants were right-handed, had normal or corrected-to-normal vision, and had no diagnosed difficulties in recognizing faces.

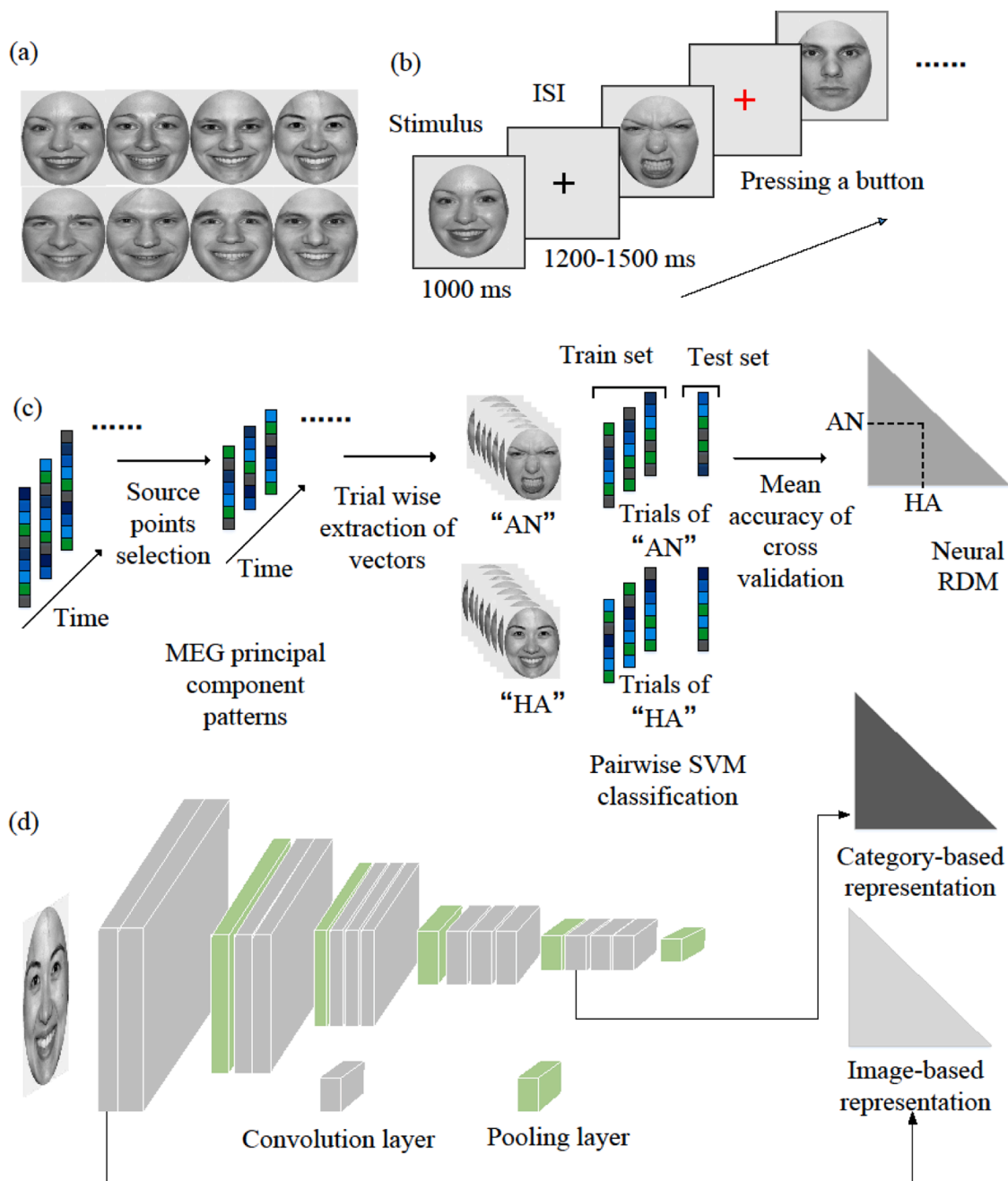


Fig. 1. Illustrations of stimuli, experimental paradigm, multivariate decoding analysis, and image- and category-based representation of the fine-tuned CNN are as follows: (a) Examples of face stimuli. All facial expression images were selected from the NimStim Face Stimulus Set. The examples shown here depict faces with a happy expression. (b) Participants viewed 56 facial expression images while performing a color change detection task. Each image was presented for 1000 ms, followed by a variable inter-stimulus interval (ISI; 1200–1500 ms). (c) Multivariate pattern analyses were performed in a time-resolved manner on trials from MEG source space separately for each face-selective region and participant. For each time point, the principal components (PCs) pattern of the response extracted after trials was averaged, and pairwise support vector machines (SVM) classification was performed with leave-one-out cross-validation (LOOCV). The resulting decoding accuracy values were assembled in a 7×7 representational similarity matrix (RDM) at each time point. (d) The image- and category-based representations were computed based on the fine-tuned CNN.

2.2. Stimuli

The stimulus consisted of 56 facial expression images (Fig. 1a), which were selected from the NimStim Face Stimuli Set (Tottenham et al., 2009). These images included seven types of facial expressions (happiness, sadness, anger, disgust, fear, surprise, and neutral) from eight individuals (four females and four males). Specifically, each facial expression category had eight images from eight different individuals. First, all facial expression images with 68 face key points were

automatically identified using the Dlib face detector (King, 2009). Using the key points of the facial contour, eyes, and mouth, the facial expression images were transformed to minimize differences in alignment. Second, an oval mask of constant size was applied to each facial expression image to remove hair cues and other visible indicators. Third, each image was converted into grayscale and adjusted to match the mean luminance and contrast.

2.3. Experimental procedure

During the MEG experiment, participants were seated upright while viewing the face stimuli. Electromagnetic coils were attached to the nasion and preauricular points on the scalp of each participant to determine head location. As shown in Fig. 1b, each trial began with a face stimulus presented on a gray background for 1000 ms. This was followed by the presentation of a central fixation cross with a variable ISI that lasted between 1200 and 1500 ms. The color of the fixation cross sometimes changed from black to red. Participants were instructed to press a button using their right index finger when they saw this color change. The purpose of this task was to maintain the participants' attention. The trials used to maintain the participants' attention during the experiment were removed from subsequent analyses. Participants viewed 20 blocks of trials in which each of the 56 images was presented once, randomly interleaved with 11 color-detection trials, for a total of 67 trials per block. In the subsequent decoding analysis, a total of 1120 trials were included, with 160 trials for each emotional category. To ensure that the participants could better understand the experiment, one practice block was presented before the formal phase.

After the MEG experiment, the participant were asked to complete two behavioral tasks. Firstly, the participant performed a facial emotion classification task. Each trial began with a face image chosen in order from the 56 stimuli used in the MEG experiment. The stimuli were displayed in the center of the screen, and the participants were asked to choose one of the seven labels ("anger", "disgust", "fear", "happiness", "sadness", "surprise", and "neutral") that matched the face image. These labels were presented at the bottom of the screen. Once a response was made, the next trial started. In addition to measuring participants' ability to identify facial expressions, we also used this task to compare the behavioral performance with the classification performance of a deep neural network. Once the classification task was complete, participants viewed a pair of face images and rated the similarity of the facial emotion on a 9-point scale, with a value of 1 indicating different facial emotions and a value of 9 indicating the same facial emotion. All stimulus pairs used in the task consisted of face images of two different expressions (e.g., neutral vs. happy). Participants completed a total of 336 emotional similarity assessments with the 21 facial expression pairs. Each facial emotion pair was evaluated 16 times because four of the eight facial expression images from the two emotional categories were randomly selected. In each trial, a pair of images was displayed on the screen side by side until the participant rated the pair and pressed a key for the next trial.

All experiment paradigms were implemented using Python and PsychoPy (Peirce., 2007). The MEG experiment and behavioral tasks lasted approximately 60 min and 30 min, respectively.

2.4. MEG acquisition and preprocessing

MEG data were collected on a 275-channel CTF system (MEG International Services LP, Coquitlam, British Columbia, Canada) with a sampling rate of 1200 Hz in a magnetically shielded and sound attenuated room at the Institute of Biophysics, Chinese Academy of Sciences. Recordings were available from 272 channels, with the exclusion of bad channels MLF55, MRT16, and MRT23.

We preprocessed the raw data using MNE-Python (Gramfort et al., 2013). The MEG triggers were aligned to the exact presentation time on the screen, which was recorded using an optical sensor attached to the projection mirror. Trials were extracted from -200 to 1000 ms with respect to stimulus onset. The data were then bandpass filtered with lower and upper cutoffs set to 1 and 100 Hz, and a bandstop filter at 50 Hz was applied to remove line noise, including harmonics. Each trial was baseline-corrected by removing the mean activation from each MEG sensor between -200 ms and stimulus onset. To further increase the signal-to-noise ratio (SNR), empty room data were used to create signal space projectors, which were applied to the filtered raw data to remove

environmental artifacts. Independent component analysis (ICA) was also applied to decompose the MEG data into a set of independent components. The components that were likely to represent eyeblink artifacts were identified based on their spatial and temporal characteristics. Finally, trials with signals exceeding standard thresholds (magnetometer = $4e-12T$) in at least one channel were rejected, and the data were downsampled to 300 Hz (360 samples/trial) to reduce computational costs.

2.5. MEG source estimation

For each participant, we projected single-trial MEG data onto the cortical surface reconstructed from their anatomical MRI scan. This approach allowed us to combine each participant's data across blocks while restricting our analyses to signals presumed to originate in predetermined regions of interest on the cortical surface. These regions of interest included the lateral occipital cortex (LO), fusiform gyrus (FG), inferior parietal cortex (IP), posterior superior temporal sulcus (pSTS). The localization of these regions was achieved using a human brain anatomical atlas (Desikan et al., 2006).

Specifically, for each participant, we first acquired a T1-weighted anatomical MRI scan on a 3T GE scanner (Discovery MR750, GE Healthcare Systems, Milwaukee, WI) at the Institute of Psychology, Chinese Academy of Sciences to create individual head models. Anatomical reconstructions based on the T1-weighted anatomical scan were completed using the FreeSurfer toolbox (Fischl., 2012). Then, we used the MNE-Python related function "watershed_bem" and FreeSurfer participant reconstruction to create BEM surfaces. In MNE-Python, we used the digitizer data and the BEM to align each participant's MEG data to their anatomical MR image and computed a surface-based source space on the reconstructed cortical surface with source points spaced 5 mm apart. We then generated a forward solution, which mapped the MEG sensor space to the source space. Using this forward solution and a noise covariance matrix calculated from the baseline period (-200 to 0 ms), we created an MEG inverse operator. Finally, we obtained the MEG source estimation in the LO, IP, FG, and pSTS of single-trial data by performing dSPM source localization in MNE-Python and setting the signal-to-noise ratio to 3 .

2.6. Identification of face-selective regions

To localize source points in the LO, IP, FG, pSTS that respond selectively to faces, we employed a localizer task with a block design. This task included stimuli from four different categories: faces, houses, objects, and scrambled objects (Goesaert and Beeck, 2013; Nestor et al., 2016). Activations from this task, specifically those indicating a stronger response to faces compared to objects (faces > objects), were identified as face-selective source points within the LO, FG, IP, and pSTS regions. We designated these regions as lLO-faces, lIP-faces, lFG-faces, and lpSTS-faces in the left hemisphere, with prefix "l" denoting "left". Conversely, the prefix "r" was used for regions in the right hemisphere. We restricted all further analyses to these face-selective regions.

Specifically, each participant completed four runs, each corresponding to a different category of images (faces, objects, houses, or scrambled objects), with a fixation baseline of 10 s in between runs. Within each run of the MEG localizer, there were 12 blocks of one of the four categories. Twelve images from a single category were shown in a row in each block, in random order, for 1000 ms with a 100 ms ISI. Participants were instructed to press a button on the response glove with their right index finger to indicate that the color of the fixation cross changed to red within each block. In total, 144 trials per category were recorded. Face-selective source points were found in the source estimation of the LO, IP, FG, and pSTS of each participant. We used the EEGLAB (Delorme and Makeig., 2004) to perform a nonparametric, one-way permutation test on trials of source space for each source point and postbaseline time point ($10,000$ permutations per test) in order to

find face-selective source points. The detailed process is described in Supplementary Note 1.

2.7. Time-resolved multivariate pattern classification in face-selective regions

We conducted multivariate classification based on previous MEG decoding studies (Carlson et al., 2011; Cichy et al., 2014, 2016b). To improve the signal-to-noise ratio prior to decoding, we averaged every four trials belonging to the same facial expression for each face-selective region of each participant (Grootswagers et al., 2017; Isik et al., 2014). With 160 trials for each emotional category, after averaging every four trials, there are 40 trials available for multivariate classification for each emotional category. Second, a source point of the averaged trial was retained if the corresponding source point was identified as face-selective in the localization analysis. Following prior research that used a sliding window of temporal resolution (Ghuman et al., 2014; Ramkumar et al., 2013), time was expressed as the beginning of the 30 ms window with a sliding interval of 10 ms for classification. Finally, for each brain region of per participant, the vector of source data across all time points between the current time point and 30 ms after the current time point was extracted. To reduce computational costs and minimize noise, principal component analysis was applied to the MEG source space trials for each time point, retaining all components that explained 99.99 % of the variance in the data. These principal component scores were used for the subsequent classification in a time-resolved manner.

To obtain a dissimilarity measure for each pair of facial expressions, we computed the pairwise classification accuracy of a linear SVM with leave-one-out cross-validation (Chang and Lin, 2011). That is, a pairwise classification was conducted between each possible facial expression pair (e.g., anger vs. happy), resulting in a total of 21 pairwise classification accuracy values. The pairwise classification was repeated n times because there were $n-1 + n-1$ samples in the training set and $1 + 1$ in the test set with LOOCV. The representational dissimilarity matrix (RDM), a 7×7 decoding matrix, used the average accuracy across repetitions as its value (Fig. 1c). The diagonal of the RDM was undefined, and the RDM is asymmetrical. In each face-selective area and per participant, we thus averaged all pairwise decoding accuracy values in the lower triangular of each RDM, which resulted in one average decoding accuracy value at each time point. In each face-selective region, the final time course of neural decoding accuracy was computed by averaging data across all participants. To smooth the decoding accuracy, we averaged these accuracy values for every three neighboring time points.

2.8. A deep neural network for facial expression classification

CNNs have achieved remarkable performance in ImageNet localization and classification tasks (He et al., 2016). A common CNN model consists of multiple convolutional and pooling layers stacked on top of each other. Although a CNN does not explicitly model time, it has a clear sequential structure where information flows from one layer to the next in a feedforward manner. The visual information represented at each layer of a CNN follows a hierarchical organization, where the lower layers capture low-level features such as edges, colors, and orientations, while higher layers capture more complex features such as object and face parts. The top layers of the CNN represent high-level class-specific information, such as faces and objects (LeCun et al., 2015; Zeiler and Fergus, 2014).

Through a comprehensive evaluation of networks with increasing depth using an architecture with small convolution filters, VGGNets have been proven to be an effective method for face and facial expression recognition (Simonyan and Zisserman, 2015). Transfer learning is also commonly used for tasks where the dataset has too little data to train a full-scale model from scratch. Therefore, we fine-tuned a 16-layer VGGNet pretrained on ImageNet for facial expression recognition using backpropagation. This means that the network learned neuronal tuning

functions on its own. The images were obtained from JAFFE (Lyons et al., 1998), CK+ (Lucey et al., 2010), FACES (Ebner et al., 2010), and KDEF (Goeleven et al., 2008) datasets, which contained seven classes of facial expressions (anger, disgust, fear, happiness, neutral, sadness, and surprise) with approximately 700 face images in each expression category. To avoid overfitting, data augmentation techniques were employed during the training of the VGGNet model. The dataset was divided into 10 groups using random splitting, with one group used as the validation set and the remaining groups used for training. The network weight parameters were learned using mini-batch stochastic gradient descent (SGD) with momentum set to 0.9. Each batch of 256 facial expression images was fed into the network for 50 epochs. The base learning rate used for training was set to 0.001.

2.9. Correlation between constructed dissimilarity representations and neural response

We employed representational similarity analysis (Kriegeskorte et al., 2008; Norman et al., 2006) to compare the neural data in face-selective regions with the constructed dissimilarity representations of face stimuli.

According to the above description, we computed neural RDM for each time point, per face-selective region, and per participant based on pairwise SVM classification of the MEG data. Each cell in the neural RDM represented the classification accuracy between the two emotional categories of facial expression. The constructed dissimilarity representations included the behavioral RDM, image-based representations, and category-based representations, which were generated as follows.

First, we subtracted each behavioral similarity rating from 10 to create a dissimilarity rating. For example, if a participant evaluated the emotion similarity of a pair of facial expressions (e.g., happy vs. angry) as 1, the dissimilarity rating was computed by subtracting the similarity value from 10. The 7×7 behavioral RDM, which was indexed in rows and columns by the compared emotions, was constructed using the dissimilarity rating of 21 pairs of facial expressions. Each cell of the behavioral RDM reflects the averaged emotion dissimilarity judgement of a pair of facial expressions across all participants.

Second, we extracted the deep features of our stimulus from the first convolution layer in the first block and fifth block structure of the fine-tuned VGGNet. These two convolution layers were named the conv1-1 and conv5-1 layers, respectively. The conv1-1 layer extracted the image properties, while the conv5-1 layer extracted the category information of facial expressions. We then averaged the features of facial expressions with the same emotional category in the conv1-1 and conv5-1 layers and calculated the dissimilarities (1-Spearman's r) between each pair of facial expressions. This produced the image-based representation and category-based representation (Fig. 1d).

In each face-selective region, the lower triangles of neural RDMs were then correlated (using Spearman's correlation) with the behavioral RDM and image- and category-based representations separately for each time point and per participant. In each face-selective region, the final time course of correlation was computed by averaging across all participants. To smooth the correlation, we averaged the correlation values of every three neighboring time points.

2.10. Statistical analysis

We employed non-parametric statistical tests for all data analyses, which were not predicated on any assumptions regarding the data distributions (Maris and Oostenveld, 2007; Pantazis et al., 2005). To perform statistical inference on decoding accuracy or correlation coefficient of RSA time series in all face-selective regions, we employed permutation-based cluster-size inference. The null hypothesis was established with a reference to a 50 % chance level for decoding accuracy values, and correlation coefficient or correlation differences were presumed to be 0. The significant temporal clusters were defined as

consecutive time points that exceeded a statistical threshold, also known as the cluster-inducing threshold. First, the conditional labels of the MEG data were permuted by randomly multiplying participants' results by +1 or -1 (i.e., sign permutation test); this procedure was repeated 1000 times resulting in a permutation distribution for every time point. Second, time points that exceeded the 95th percentile of the permutation distribution served as cluster-inducing time points (i.e., equivalent to $p < 0.05$; one-sided). Finally, clusters in time were determined using the 95th percentile of the maximum number of consecutive significant time points across all permutations (i.e., equivalent to $p < 0.05$; one-sided).

2.11. Onset, peak latency, and peak accuracy analysis

We employed bootstrap tests to examine the significant differences in the onset, peak latency, and peak accuracy for the time-resolved decoding accuracy among face-selective regions. Specifically, in each face-selective region, we first bootstrapped the participants' decoding accuracy at each time point 1000 times to obtain a three-dimensional matrix (participants \times time points \times 1000). We then computed the empirical distribution of the onset (i.e., minimum significant time point post-stimulus onset), peak (i.e., time point of maximum classification value between 80 and 220 ms post-stimulus onset) latency, and peak accuracy (i.e., maximum classification accuracy between 80 and 220 ms post-stimulus onset) of different face-selective regions. Because the peak latency and accuracy focused on the first peak that occurred after stimulus onset, we restricted the period to 220 ms post-stimulus onset to prevent confusion with stimulus offset responses (Carlson et al., 2011). To examine the differences in onset, peak latency, and peak accuracy among face-selective regions, we computed 1000 bootstrap samples of the difference, resulting in an empirical distribution of onset, peak latency, and peak accuracy differences. The number of differences smaller or larger than zero was divided by the number of permutations to calculate the p -value (i.e., two-sided testing). These p -values were then corrected for multiple comparisons using the false discovery rate (FDR) at a 0.05 level.

3. Results

3.1. Analysis of behavior

For the behavioral experiment, the average confusion matrix of the classification task and the behavioral RDM of the similarity rating task across participants are shown in Fig. 2a and 2b. The average classification accuracy is 75.21 % with a standard deviation of 5.46 % across all

participants.

For the classification task, a one-way repeated measures ANOVA revealed that the effect of emotion was highly significant, $F(6, 96) = 19.4$, $p < 0.001$, $\eta_p^2 = 0.55$. Follow-up paired sample t -tests with FDR correction revealed that happy faces were recognized significantly more accurately than angry, disgusted, fearful, and sad faces (all $t > 6.6$, all $p < 1.8e-5$, all $d > 1.6$). There was also a trend for happy faces to be better recognized than surprised faces ($t(16) = 3.44$, $p = 0.0066$, $d = 0.83$). Neutral faces were recognized significantly more accurately than angry, disgusted, fearful, and sad faces (all $t > 5.8$, all $p < 6.7e-5$, all $d > 1.4$) and displayed a trend for better recognition than surprised faces ($t(16) = 3.34$, $p = 0.0072$, $d = 0.81$), but neutral faces were not better recognized than happy faces ($t(16) = 0.80$, $p = 0.4524$, $d = 0.19$).

There was a trend that surprised faces are recognized more accurately than sad ($t(16) = 3.44$, $p = 0.0066$, $d = 0.83$), fearful ($t(16) = 2.76$, $p = 0.0206$, $d = 0.67$), and disgusted ($t(16) = 3.42$, $p = 0.0066$, $d = 0.83$) faces, while surprised faces were not recognized better than angry faces ($t(16) = 1.41$, $p = 0.2188$, $d = 0.34$). There were no significant differences in accuracy between fearful faces and sad ($t(16) = 0.81$, $p = 0.4428$, $d = 0.2095$), disgusted ($t(16) = 0.35$, $p = 0.7258$, $d = 0.0866$), and angry ($t(16) = 2.12$, $p = 0.0645$, $d = 0.5163$) faces. Furthermore, there was no significant distinction between sad and disgusted faces ($t(16) = 1.01$, $p = 0.38$, $d = 0.24$).

Overall, the behavioral classification performance was consistent with previous findings, indicating that happy faces are the most easily recognized facial expressions (Calvo and Beltrán, 2013). However, distinguishing between negative emotions, especially based on fearful, sad, and disgusted faces, was more challenging (Blair and Coles, 2000; Sullivan and Ruffman, 2004).

3.2. Time resolved decoding of facial expressions

To determine the time course at which neural representations in the human brain start to discriminate facial expressions, we employed linear SVM to classify each possible pair of facial expressions. We then computed the average of all 21 pairwise accuracy values at each time point, resulting in the time course of neural decoding accuracy for the seven categories of facial expressions. Decoding accuracy values were averaged across participants (Fig. 3a). Since it was a binary classification for each possible pair of expressions (e.g., anger and happy) selected from the seven expression categories, the chance level was 50 %. The results showed that facial expressions were rapidly discriminated from neural representations. The discrimination reached the level of significance at 100 ms after stimulus onset and peaked at 150 ms, with a decoding accuracy of 57.9 %. The accuracy remained significantly above

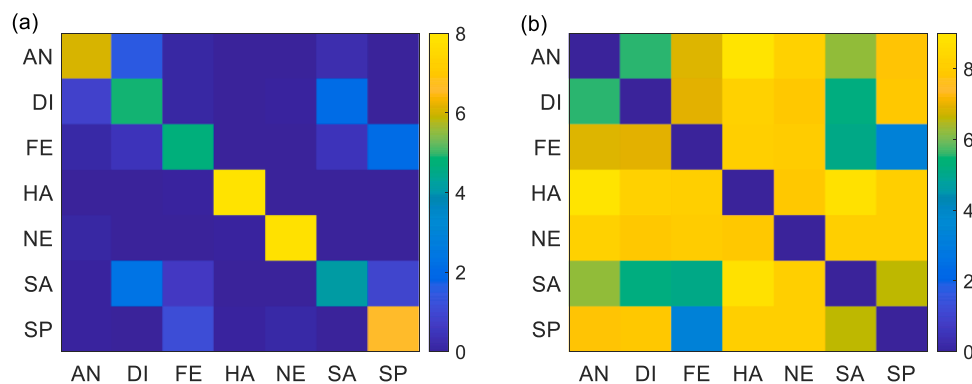


Fig. 2. Averaged facial expression classification and similarity rating performance across all participants are shown in Fig. 2a and Fig. 2b. (a) The confusion matrix is based on the facial expression classification task. Rows represent expression categories, and columns represent the responses chosen by participants (AH: anger, DI: disgust, FE: fear, HA: happy, NE: neutral, SA: sad, SP: surprise). The diagonal represents correct responses, and the off-diagonal represents errors. The color scale indicates the average number of times a particular facial expression and response pair were chosen across participants. (b) The RDM is based on the behavioral facial expression similarity rating task. Rows and columns represent expression categories (AH: anger, DI: disgust, FE: fear, HA: happy, NE: neutral, SA: sad, SP: surprise). The diagonal is undefined. The color scale indicates the dissimilarity of a pair of facial expressions, computed based on similarity ratings averaged across participants.

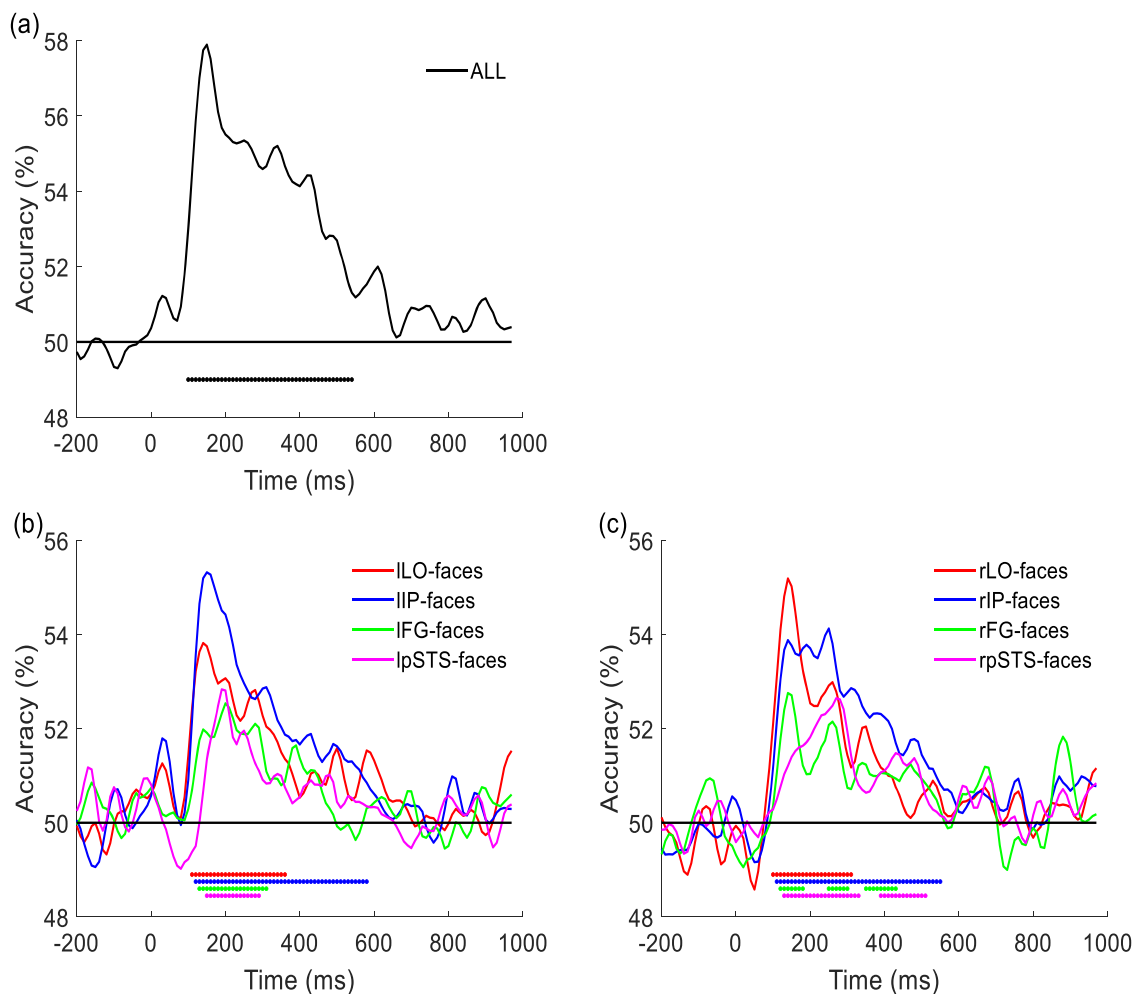


Fig. 3. The averaged time course of the facial expression decoding accuracy across participants was examined in three different contexts: (a) at the whole brain level, (b) in the face-selective regions of the left hemisphere, and (c) in the face-selective regions of the right hemisphere. The x-axis represents time, with 0 indicating the stimulus onset. The lower part of the plots includes dotted black and colored lines that indicate the range of time during which expression discrimination is significantly better than chance, as determined by a cluster-based sign permutation test (with a cluster-defining threshold of $p < 0.05$ and a corrected significance level of $p < 0.05$).

the chance level until 540 ms.

To determine when these face-selective regions encoded information for facial expressions discrimination, we examined the time course of neural representations for discriminating facial expressions in each of the four face-selective regions in the left hemisphere (lLO-faces, lIP-faces, lIFG-faces, and lpSTS-faces; Fig. 3b) and right hemisphere (rLO-faces, rIP-faces, rFG-faces, and rpSTS-faces; Fig. 3c) using the same decoding analysis at the whole-brain level. In the left hemisphere, the results showed that following the stimulus onset, facial expressions were discriminated at 110–390 ms in the lLO-faces, 120–580 ms in the lIP-faces, 130–310 ms in the lIFG-faces, and 150–290 ms in the lpSTS-faces. In the right hemisphere, facial expressions were discriminated at 100–320 ms in the rLO-faces, 110–540 ms in the rIP-faces, multiple ranges (120–180 ms, 250–300 ms, and 350–430 ms) in the rFG-faces, and 130–330 ms and 390–510 ms in the rpSTS-faces. For all cluster-corrected sign permutation tests, the cluster definition threshold and the cluster-corrected significance level were set at $p < 0.05$. Thus, the time-resolved decoding analysis suggested that starting at about 100 ms and 150 ms, all source points in the LO-faces, IP-faces, FG-faces, and pSTS-faces began to discriminate facial expressions. We also conducted facial expressions decoding in all regions to obtain a more comprehensive understanding of facial expression processing (Supplementary Note 2). These regions were also localized using an anatomical atlas of the human brain (Desikan et al., 2006).

For the time-resolved decoding analysis, the onset latency in the lLO-faces was earlier than that in the lIFG-faces ($p < 0.05$, two-sided bootstrap test, FDR corrected) and the lpSTS-faces ($p < 0.001$), but not earlier than that in the lIP-faces (Fig. 4a, $p = 0.1$). The onset latencies in the lIP-faces and lIFG-faces were earlier than in the lpSTS-faces (Fig. 4a, $p < 0.001$). Although the onset latency in the lIP-faces was earlier than that in the lIFG-faces, the significance was only marginal (Fig. 4a, $p = 0.0445$). In the rLO-faces, the onset latency occurred earlier than in the rFG-faces ($p < 0.001$, two-sided bootstrap test, FDR corrected) and the rpSTS-faces ($p < 0.001$), as well as in the rIP-faces (Fig. 4b, $p < 0.05$). The onset latencies in the rIP-faces and rFG-faces were earlier than in the rpSTS-faces (Fig. 4b, $p < 0.05$). Thus, these findings indicated that activation occurred earlier in the LO-faces than in FG-faces and pSTS-faces, and activation appeared to occur earlier in the IP-faces and FG-faces than in pSTS-faces.

In addition, the peak decoding accuracy appeared at about 150 ms in the lLO-faces and lIP-faces, and appeared at about 200 ms in the lIFG-faces and lpSTS-faces after stimulus onset (Fig. 4c). The decoding accuracy peaked at about 150 ms in rLO-faces, rIP-faces, and peaked at about 200 ms in the rpSTS-faces (Fig. 4d). There were two peak decoding accuracy values in rFG-faces, occurring at 150 ms and 200 ms. We noticed that the peak latencies were not significant different in these face-selective areas in the left (Fig. 4c) and right (Fig. 4d) hemispheres.

Aside from the differences in latencies in the four face-selective

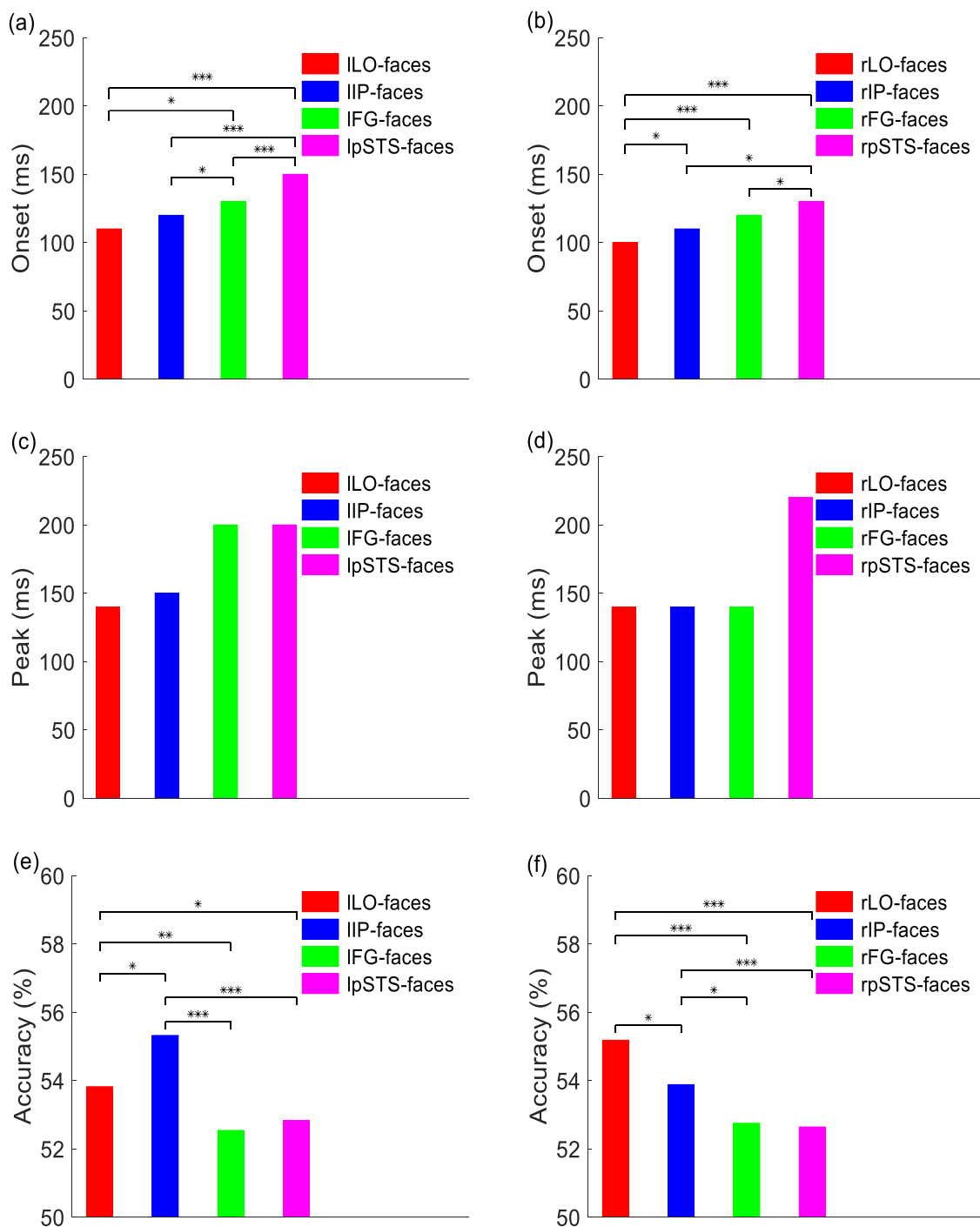


Fig. 4. (a) and (b) The onset latency, (c) and (d) peak latency, and (e) and (f) peak accuracy for decoding facial expressions in face-selective areas in the left and right hemisphere are shown. Stars above the bars indicate significant differences across regions (one-sample two-sided bootstrap test with FDR correction, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

regions, we also compared the peak accuracy in these regions. The peak accuracy of the ILO-faces was higher than that in the IFG-faces ($p < 0.005$) and lpSTS-faces (Fig. 4e, $p < 0.05$). The peak accuracy of the IIP-faces was also higher than that in the IFG-faces ($p < 0.001$) and lpSTS-faces (Fig. 4e, $p < 0.001$). Moreover, the peak accuracy of the rLO-faces outperformed that in the rFG-faces ($p < 0.001$, two-sided bootstrap test, FDR corrected) and rpSTS-faces (Fig. 4f, $p < 0.001$). The peak accuracy in the rIP-faces was also higher than that in the rFG-faces ($p < 0.05$) and rpSTS-faces (Fig. 4f, $p < 0.001$). Thus, the difference in peak accuracy showed that LO-faces and IP-faces exhibit better performance in discriminating facial expressions when compared with the FG-faces in

the ventral pathway and pSTS-faces in the lateral pathway.

3.3. Temporal representation to different facial expressions

To explore the temporal representation of different facial expressions, we also performed a binary classification of neutral expression compared separately to each of the six emotional expressions at the brain level. The binary classification accuracy values were averaged across participants (Fig. 5a). Happy faces could be discriminated between 90 and 520 ms, with a peak of 120 ms and an accuracy of 61.15%. Angry faces were significantly discriminated between 80 and 500 ms,

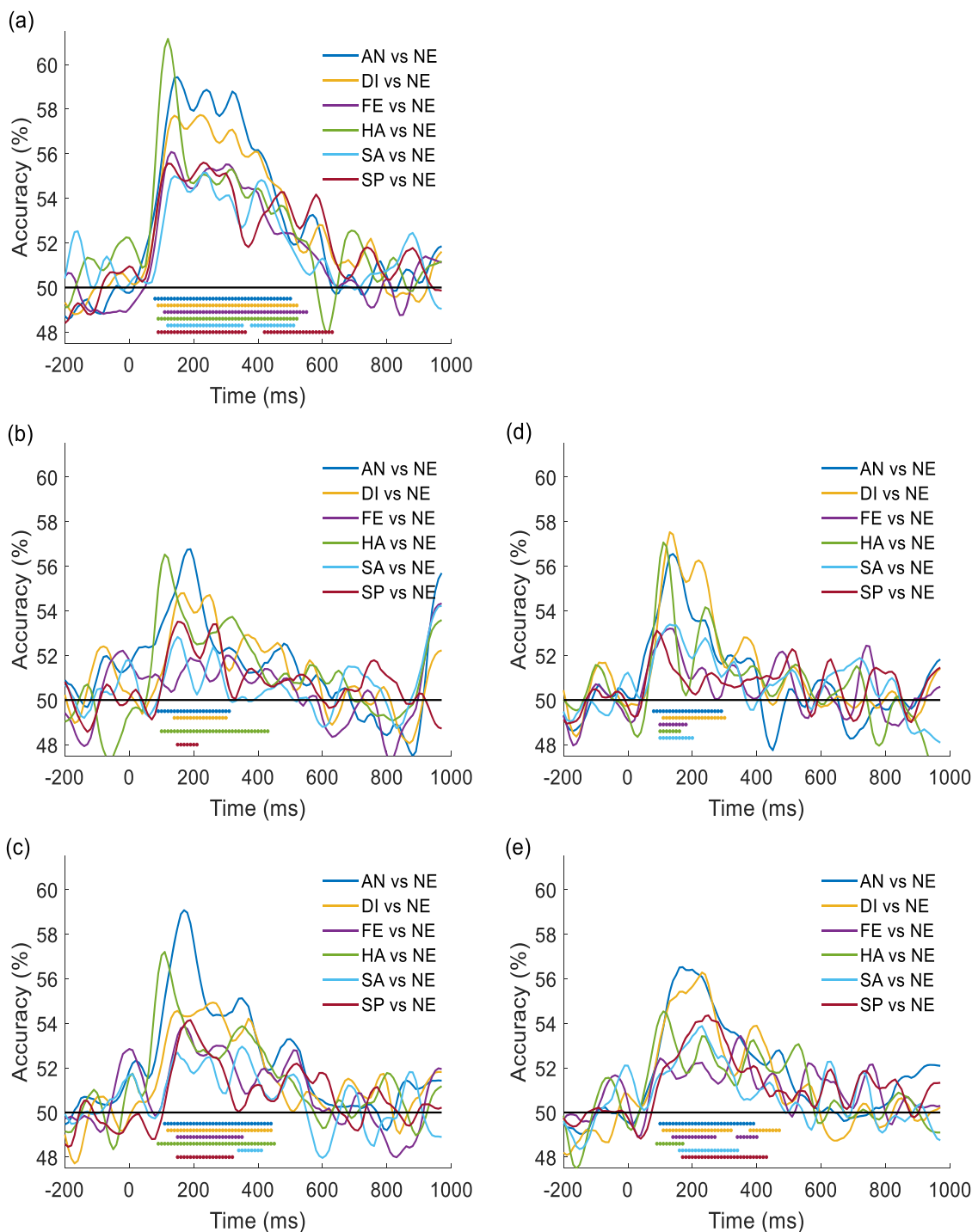


Fig. 5. Averaged time course of the classification accuracy values between neutral expression and each of the emotional facial expressions across participants. (a) Whole brain level, (b) lLO-faces, (c) lIP-faces, (d) rLO-faces, (e) rIP-faces. Abbreviations: HA: happiness, AN: anger, DI: disgust, FE: fear, SP: surprise, SA: sadness. The dotted lines below the plots indicate significant times according to a cluster-based sign permutation test (cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.05$).

peaking at 150 ms with an accuracy of 59.43 %. Disgusted faces were also distinguished between 90 and 520 ms, with a peak at 140 ms (57.70 %). The accuracy was significantly above the chance level at 110–550 ms of fearful faces, at 120 and 350 ms, and 380 and 510 ms of sad faces, and at 90–360 ms and 420–630 ms of surprised faces. Overall, these results revealed that the neural representation of each category of emotional face was rapidly discriminated from neutral expression at approximately 100–500 ms after the stimulus onset, with a peak at most 150 ms.

Furthermore, considering the function of LO-faces and IP-faces in discriminating facial expression, we also performed binary classification on these two face-selective regions. In the lLO-faces, apart from the fearful and sad faces, the remaining emotion faces can all be discriminated from neutral faces (Fig. 5b). In the lIP-faces, we found that the neural representation successfully discriminated all six emotional expressions from the neutral expression (Fig. 5c). In the right hemisphere, except for the surprise faces, the neural representation in rLO-faces successfully discriminated other emotional expressions from the

neutral expression (Fig. 5d). There was also a clear differentiation in rIP-faces between neutral expressions and each of the six facial expression categories (Fig. 5e). To test how persistent these neural representations of facial expressions were, we also performed temporal generalization analysis for each facial expression (Supplementary Note 3).

3.4. Correlation between behavioral similarity rating and neural response

To examine the extent to which the neural data in face-selective regions could correlate with behavior, we compared the pairwise structure of neural RDMs to the behavioral RDM. The behavioral RDM correlated significantly with neural RDMs in the lLO-faces (time periods: 100–190 and 270–380 ms) and lIP-faces (110–190 and 320–440 ms; Fig. 6a), but was not correlated with neural RDMs in the lFG-faces and lPSTs-faces in any time window (Fig. 6b). In the right hemisphere, the behavioral RDM correlated significantly with the neural RDMs in the rLO-faces (time periods: 90–160 ms) and rIP-faces (60–200 and 350–460 ms; Fig. 6c). There was also a significant correlation between the behavioral RDM and neural RDMs in the rpSTs-face at early time points (60–190 ms; Fig. 6d).

Thus, neural data in the LO-faces and IP-faces correlated with behavioral responses between ~100 and 450 ms after stimulus onset, indicating that these regions contain adequate information for discriminating facial expressions.

3.5. Comparison between deep feature-based representation with neural response

Using facial expression images from public datasets, we trained the 16-layer VGGNet pre-trained in ImageNet to perform facial expression recognition. When we used our emotional face stimuli as a test set for the fine-tuned VGGNet, the confusion matrix was constructed with a classification accuracy of 69.6 % (Table 1). The rows and columns of the confusion matrix represent the true expression category and the predicted class by VGGNet, respectively. For example, the cell in the second column of the first row reflects that the true emotional category of “AN”

Table 1

The confusion matrix of facial expressions classification using the fine-tuned VGGNet. The rows represent the true expression category, and the columns indicate the predicted class by VGGNet (AN: Anger, DI: Disgust, FE: Fear, HA: Happy, NE: Neutral, SA: Sad, SP: Surprise). The diagonal hence represents correct responses, and the off-diagonal represents errors.

Actual Expression		AN	DI	FE	HA	NE	SA	SP
AN	7	1	0	0	0	0	0	0
DI	3	2	0	1	0	2	0	0
FE	0	0	6	0	1	1	0	0
HA	0	0	0	8	0	0	0	0
NE	1	0	0	0	7	0	0	0
SA	0	0	2	0	1	5	0	0
SP	0	0	3	0	0	1	4	0
		AN	DI	FE	HA	NE	SA	SP
		Predicted Expression						

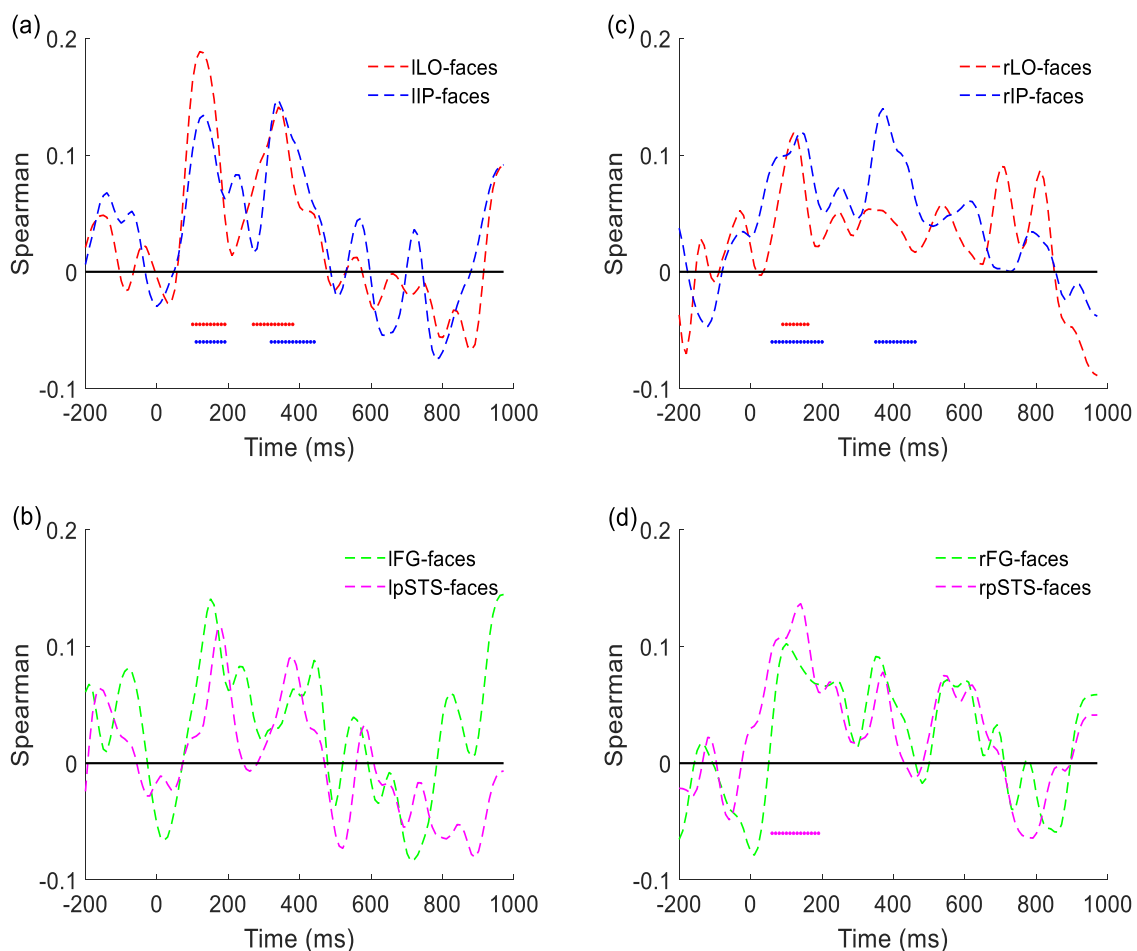


Fig. 6. Averaged correlations between behavioral RDM and neural RDMs in face-selective regions across participants are shown in Figure (a) for lLO-faces and lIP-faces, (b) for lFG-faces and lPSTs-faces, (c) for rLO-faces and rIP-faces, and (d) for rFG-faces and rpSTs-faces. Dotted horizontal lines below the plots indicate the time duration of significant activation according to a cluster-based sign permutation test (cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.05$).

was classified as “DI”.

The performance of VGGNet was comparable to the participants’ behavioral results, where the mean accuracy across participants was 75.2% (SD=5.5). Deep features of the test set were then extracted in the low layer (conv1–1) and high layer (conv5–1) of the fine-tuned VGGNet. We constructed the image-based representation using the features in the conv1–1 layer and computed the category-based representation utilizing the features in the conv5–1 layer. Each cell in the image- and category-based representation reflects the dissimilarity (1-Spearman’s r) between a pair of facial expressions. Thus, the image-based representation (Fig. 7a) mainly reflects image properties such as texture and edge, and the category-based representation (Fig. 7b) shows the category information of stimuli.

To understand when and what information was encoded in each face-selective region for facial expression discrimination, we examined the correlation of the neural RDMs within each region after stimulus onset with the image- and category-based representations. The RDMs in lLO-faces correlated with the image-based representation during the time periods of 120–160 ms, as well as the category-based representation during the time periods of 110–240 ms and 310–390 ms after stimulus onset (Fig. 8a). Similarly, the RDMs in lIP-faces also correlated with the image-based representation during the periods of 130–230 ms, and the category-based representation during the periods of 120–260 ms (Fig. 8b). In contrast, no correlation was found between the RDMs in lFG-faces and either image- or category-based representations (Fig. 8c), and the RDMs in lPSTS-faces only correlated with the category-based representation in a very short period between 180 and 230 ms (Fig. 8d). In the right hemisphere, the RDMs in rLO-faces (Fig. 8e), rIP-faces (Fig. 8f), and rpSTS-faces (Fig. 8h) had a significant correlation with the category-based representation but not with the image-based representation. The RDMs in rFG-faces had no correlation with either image- or category-based representation at any time points (Fig. 8g).

More importantly, we found that the RDMs in lLO-faces (time periods: 100–220 ms and 330–400 ms) and lIP-faces (time periods: 100–270 ms and 330–430 ms) were more strongly correlated with the category-based representation than the image-based representation. Similarly, compared to the image-based representations, the RDMs in rLO-faces (time periods: 130–180 ms and 200–300 ms) and rIP-faces (time periods: 130–220 ms and 390–480 ms) were also more strongly correlated with category-based representations. Thus, even from ~100 ms after stimulus onset, the neural data in the LO-faces and IP-faces encoded categorical information rather than image information of facial expressions. Due to the HMAX model being considered as stimulating the visual responses of simple cells in V1 (Riesenhuber and Poggio, 1999), we also computed the image-based representation based on activations for facial expression images from the HMAX C2 layer. We

then compared the neural RDMs in LO-faces and IP-faces with the image-based representation in the HMAX model. The result (Supplementary Note 4) may support the above finding that the neural data in IP-faces and LO-faces encoded the categorical information rather than the image information of facial expressions.

4. Discussion

The primary objective of this study was to investigate the temporal dynamics of neural representations in face-selective regions during the processing facial expression using high spatiotemporal resolution MEG techniques. The findings of this study revealed that within a time frame of 100 to 150 ms after the onset of the stimulus, face-selective regions such as LO-faces, IP-faces, FG-faces, and pSTS-faces already demonstrated the ability to differentiate between different facial expressions. Interestingly, the neural activation patterns observed in LO-faces and IP-faces indicated that these regions were primarily involved in the emotional categories rather than the image information processing. Additionally, a noteworthy correlation was observed between the neural representations in LO-faces and IP-faces and the participants’ performance in facial expression similarity tasks, within a time window of approximately 100 to 450 ms.

At the whole-brain level, the onset latency (at 100 ms) of facial expression discrimination was consistent with previous studies that have investigated the temporal dynamics of emotion face processing. For instance, multivariate analysis of EEG data has revealed that representations of facial expressions were extracted at approximately 100 ms (Li et al., 2022; Muukkonen et al., 2020; Smith and Smith, 2019). Decoding of each of the six basic expressions and neutral faces began at 80 and 120 ms in the human brain. These onsets of above-chance accuracy are similar to the binary classification for angry faces and neutral/happy faces in MEG studies (Dima et al., 2018), as well as the binary classification each of happy, fearful, angry faces, and neutral expressions in EEG studies (Muukkonen et al., 2020).

Crucially, our study utilizes time-resolved decoding analysis to investigate the time course of facial expression discrimination in multiple regions of interest within a distributed network. To the best of our knowledge, this was the first study to examine the time course of these face-selective regions in this specific context. In previous EEG study, Li et al. (2022) found that certain EEG channels located in the right occipital, temporal, parietal, and frontal regions achieved significant multiclass decoding accuracy for fearful, happy, and neutral faces in the time windows after 130 ms. A MEG study also demonstrated that responses from occipital sensors successfully discriminated angry and neutral faces at 93 ms, as well as angry and happy faces at 113 ms (Dima et al., 2018). However, the spatial information of facial expression

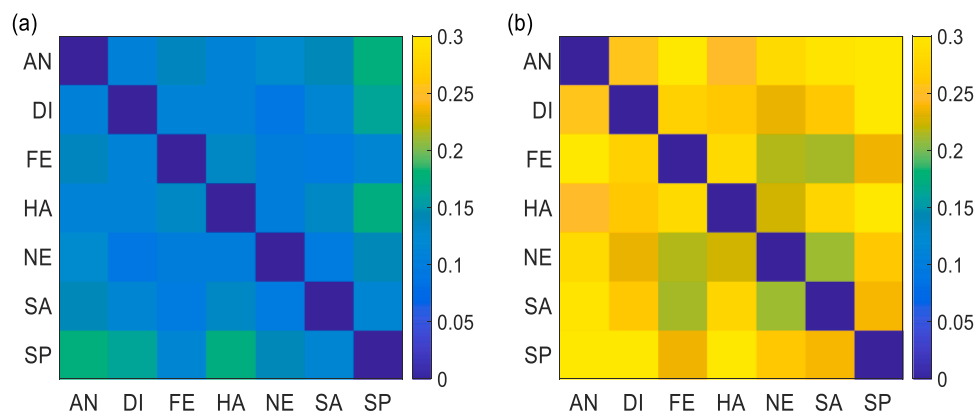


Fig. 7. RDMs were created based on deep features from different layers in the VGGNet. (a) Image-based representation. (b) Category-based representation. Rows and columns represent expression categories (AH: Anger, DI: Disgust, FE: Fear, HA: Happy, NE: Neutral, SA: Sad, SP: Surprise). The diagonal is undefined. The color scale indicates the dissimilarity (1-Spearman) of a pair of facial expressions, computed based on deep features extracted from VGGNet.

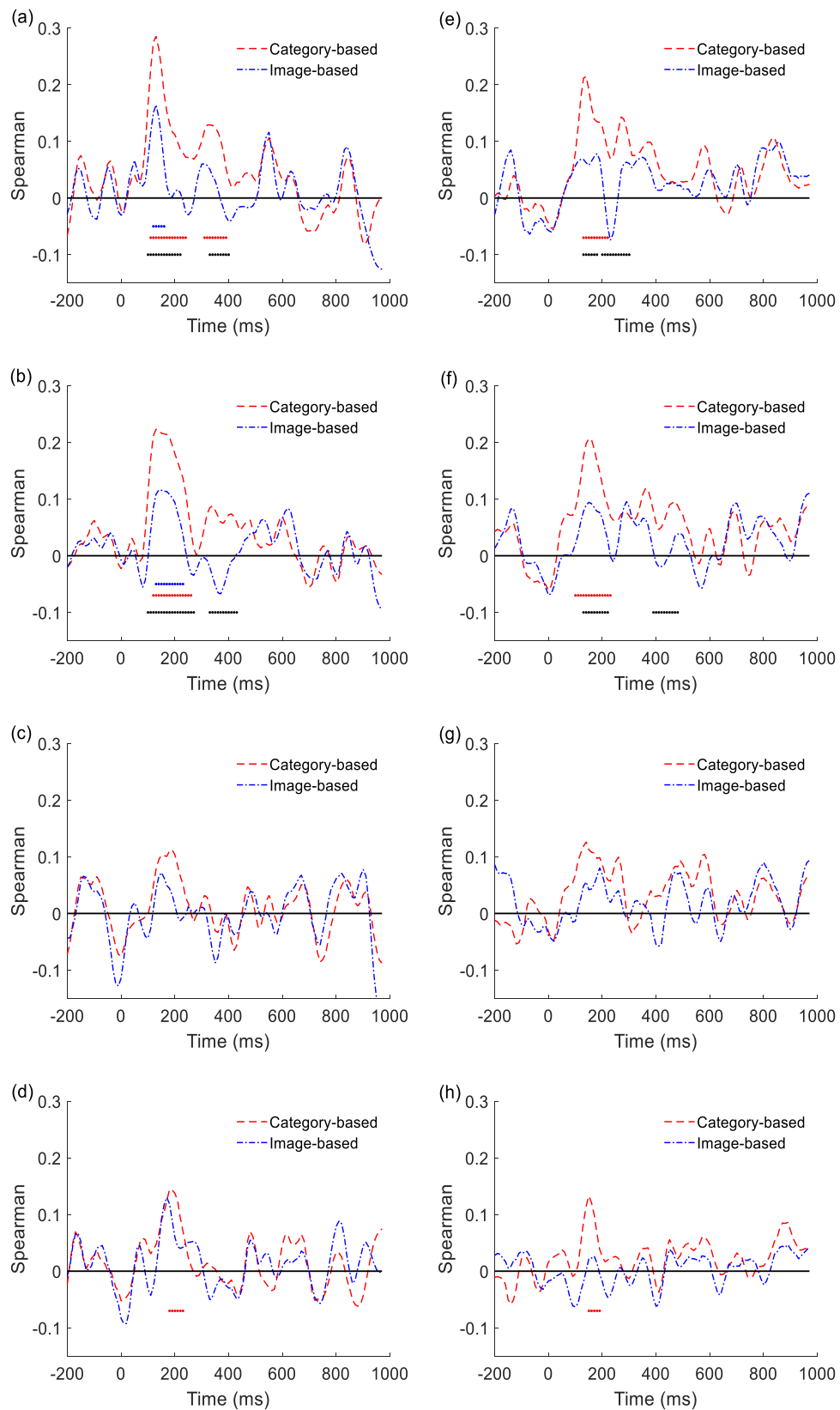


Fig. 8. Averaged correlations between image- and category-based representation and neural RDMs in face-selective regions across participants are shown in (a) lLO-faces, (b) lIP-faces, (c) lFG-faces, (d) lpSTS-faces, (e) rLO-faces, (f) rIP-faces, (g) rFG-faces, and (h) rpSTS-faces. Dotted lines under the plots indicate time intervals where the category-based (red dotted line) or image-based (blue dotted line) representation was significantly correlated with neural RDMs. The dotted black lines indicate time intervals during which the correlations differed significantly between the image- and category-based representation (both according to the cluster-based sign permutation test, cluster definition threshold $p < 0.05$, and corrected significance level $p < 0.05$).

decoding in M/EEG sensor space cannot be localized to these face-selective areas in the distributed network. Our results precisely and clearly illustrated when all regions of face-selective areas began to encode information for discriminating facial expressions, which suggested rapid emotional face processing in the ventral, lateral, and dorsal visual pathways.

We also observed that the decoding accuracy initially peaked between 140 and 200 ms in these regions, except for the rpSTS-faces condition (which peaked at 270 ms). Interestingly, the latencies of peak decoding accuracy in these regions were similar to the time point at which the N170 component, evoked by face stimuli, typically appears. However, most studies have identified early components (i.e., 100–250 ms post-stimulus presentation) as markers that carry information about differences between expressive and neutral faces (Batty and Taylor, 2003; Caharel et al., 2005; Wronka and Walentowska, 2011), while the subsequent component (i.e., after 250 ms post-stimulus) is seen as a marker that differentiates various expressions (Rellecke et al., 2012, 2013). The findings in our study demonstrate that facial expressions are successfully discriminated in regions of interest at a much earlier stage (approximately 100 ms), which is in contrast to traditional component analysis in ERP studies. This difference is likely due to the method of analysis. Multivariate analysis techniques consider the relationship among multiple variables (such as channels in M/EEG), which can enhance the sensitivity of identifying differences among multiple experimental conditions (Grootswagers et al., 2017; Norman et al., 2006). As emotional faces are represented in the brain in a complex and network-based manner (Hamann, 2012), multivariate analysis is more advantageous for addressing the temporal dynamics of information processing in the human brain.

Our findings provide empirical support for the significance of face-selective regions in the decoding of facial expression. Prior studies utilizing functional magnetic resonance imaging (fMRI) have demonstrated that the FFA (Greening et al., 2018; Liang et al., 2017; Wegrzyn et al., 2015) and pSTS (Said et al., 2010; Wegrzyn et al., 2015; H. Zhang et al., 2016;) regions possess the capability to accurately predict facial expressions through the utilization of multivariate analysis techniques, surpassing chance levels. Our results aligned with these previous findings, as we also observed substantial decoding accuracy within the FG-faces and pSTS-faces regions for facial expressions. However, our investigation revealed that the LO-faces and IP-faces regions exhibited even greater accuracy in decoding facial expressions compared to the FG-faces and pSTS-faces regions. Additionally, we discovered a significant correlation between behavioral responses and neural activity within the LO-faces and IP-faces regions during a specific time window of approximately 100–400 ms, which was not observed within the FG-faces and pSTS-faces regions. These findings suggested that the face-selective regions within the occipital-parietal pathway (LO and IP) may play a pivotal role in the discrimination of facial expressions, an aspect that has been overlooked in previous studies. Consequently, our results offered a novel perspective on the processing of facial expressions within the human brain.

One particularly noteworthy finding of our study was that the responses in the LO-faces and IP-faces regions exhibited a category-based representation rather than an image-based representation, even at an early stage of processing (~100 ms). This study is the first to investigate time course of the nature of representations in these specific regions of interest for the discrimination of facial expressions. Previous studies have primarily focused on object and face identity discrimination. For example, Cichy et al. (2016) Cichy et al., 2016a compared the similarity of representations between fMRI data and deep neural networks and found that the occipital lobe primarily processes low-level image properties, while the more anterior regions in the parietal and temporal lobes are responsible for abstract and category information in object recognition. In the case of face identity, the occipital lobe was predominantly associated with low-level image properties captured by GIST descriptors (Tsantani et al., 2021; Weibert et al., 2018) or Gabor-Jet models

(Tsantani et al., 2021), while the FFA primarily processes high-level properties such as perceived visual similarity and social traits (Tsantani et al., 2021) and high-level deep features extracted using deep neural networks (Cichy, Khosla, et al., 2016). These studies support a hierarchical organization of face-selective regions in existing face processing models (Duchaine and Yovel, 2015; Haxby et al., 2000), with the OFA specializing in low-level image properties and the FFA processing higher-level visual features and other social information. Additionally, Vida et al. (2017) demonstrated that LO-faces and FG-faces encode image-based and identity-based representation at different stages of processing. However, our findings did not support a hierarchical model for facial expression discrimination, as the data from the LO-faces and IP-faces regions consistently exhibited a category-based representation rather than an image-level representation between ~100 and 400 ms.

Due to the rapid and robust representation of facial expression in the LO-faces and IP-faces, it suggests that the occipital-parietal pathways play a role in facial expression discrimination. Recognizing facial expression involves specific processing that is, to some extent, separable from other aspects of face recognition (Bruce and Young, 1986). A previous study has also demonstrated that the ability to recognize facial expressions is preserved in patients with both prosopagnosia and cortical blindness ((Duchaine and Nakayama, 2003)). Researches have proposed that the pSTS along the lateral visual pathway and the FFA within the ventral pathway, which receives inputs from the OFA, play roles in recognizing and decoding facial expressions (Liang et al., 2017; Wegrzyn et al., 2015). Our findings indicated that, in addition to the ventral and lateral pathways, the IP cortex, which receives input from the occipito-parietal cortex, also plays a role in recognizing facial expressions as part of the dorsal pathway. These findings further supported previous empirical studies, such as Baroni et al.'s research using intracranial electrodes to decode faces, which found that certain electrodes in the bilateral IP were capable of successfully classifying faces (Baroni et al., 2017). There is also evidence that the damage to the IP affected the performance of facial expression recognition (Adolphs et al., 1996, 2000). At the same time, it is crucial to understand the general role of the IP in recognizing expressions. Previous research on the inferior parietal cortex has primarily focused on its role in representing actions rather than decoding facial expressions (Ester et al., 2020; Freud et al., 2020; Kravitz et al., 2011). Our study highlighted the unique role of the IP region in decoding facial expressions, which has been overlooked. We postulated that facial expression can be considered as a distinct form of action, characterized by facial expression activities that primarily manifest in the facial region. Consequently, we propose that the IP region may encode patterns of these expression actions. However, it is imperative to validate this hypothesis through further comprehensive investigations.

While our study provided valuable insights, it is important to acknowledge its limitations. Firstly, one may question whether the neural responses observed in the face-selective regions of our study align with those typically identified in fMRI studies of face processing. Although the source points within LO-faces, IP-faces, FG-faces, and pSTS-faces in our study are located within the same anatomical sub-regions as OFA, IP, FFA, and pSTS, respectively, there may still be variations in their neural representations. Several factors can contribute to this variability. The projection of magnetic field data into brain space during MEG source estimation is a mathematically ill-posed problem, which inherently limits the spatial accuracy. Additionally, the spatial signature of neural current changes across the cortex may depend on the specific method used to compute the inverse model. Furthermore, precise spatial alignment between MEG data and MRI structural imaging is crucial for accurate forward solution computation. However, movement of the markers on the head during MRI scans can introduce ambiguity in spatial comparisons. Second, while multivariate decoding techniques offer greater sensitivity to condition-specific differences in neural responses compared to traditional methods, it is important to consider whether the information used for classification is essential for brain

computation. Machine-learning algorithms employed in pattern classification do not aim to simulate the brain's information processing mechanisms. Instead, they assess statistical dependencies between stimuli and response patterns. Moreover, successful classifications using linear and nonlinear classifiers require sufficient input from different conditions to prevent overfitting. However, small sample sizes are often sufficient in human brain computation. Therefore, it remains unclear whether decoding-based inferences truly reflect the computational mechanisms in the human brain.

5. Conclusions

In conclusion, we utilized MVPA and a trained CNN to investigate the temporal representation of emotional faces in face-selective regions. Our findings demonstrated that neural data in the LO-faces, IP-faces, FG-faces, and pSTS-faces exhibit early discrimination of facial expression, occurring within approximately 100–150 ms after the onset of the stimulus. Importantly, even at this early time point (~100 ms), LO-faces and IP-faces predominantly processed emotional category information rather than spatial information of facial expressions. Furthermore, our results highlighted the ability of LO-faces and IP-faces in discriminating facial expressions compared to STS-faces and FG-faces. This study presented novel insights and challenges the existing understanding of facial expression processing in the brain.

Ethics statement

This study was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences. Prior to participating in the study, each participant provided written informed consent.

CRedit authorship contribution statement

Zhihao Zhang: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Tong Chen:** Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Ye Liu:** Methodology, Visualization, Writing – original draft, Writing – review & editing. **Chongyang Wang:** Data curation, Formal analysis, Writing – original draft. **Ke Zhao:** Conceptualization, Data curation, Supervision, Writing – original draft, Writing – review & editing. **Chang Hong Liu:** Writing – original draft, Writing – review & editing. **Xiaolan Fu:** Conceptualization, Supervision, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing interests.

Data availability

The available data and code can be found at: <https://osf.io/ha5ye/>.

Acknowledgements

This research was financially supported in part by the National Natural Science Foundation of China (62061136001, 32071055). We would like to thank our three reviewers for insightful comments that have greatly improved the manuscript. We express our sincere gratitude for the professional guidance provided by Dr. Yiheng Tu and Dr. Chao Jiang in the field of data analysis.

Supplementary materials

Supplementary material associated with this article can be found, in

the online version, at [doi:10.1016/j.neuroimage.2023.120442](https://doi.org/10.1016/j.neuroimage.2023.120442).

References

- Adolphs, R., Damasio, H., Tranel, D., Cooper, G., Damasio, A.R., 2000. A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J. Neurosci.* 20 (7), 2683–2690. <https://doi.org/10.1523/JNEUROSCI.20-07-02683.2000>.
- Adolphs, R., Damasio, H., Tranel, D., Damasio, A.R., 1996. Cortical systems for the recognition of emotion in facial expressions. *J. Neurosci.* 16 (23), 7678–7687. <https://doi.org/10.1523/JNEUROSCI.16-23-07678.1996>.
- Ambrus, G.G., Kaiser, D., Cichy, R.M., Kovács, G., 2019. The neural dynamics of familiar face recognition. *Cereb. Cortex* 29 (11), 4775–4784. <https://doi.org/10.1093/cercor/bhz010>.
- Baroni, F., van Kempen, J., Kawasaki, H., Kovach, C.K., Oya, H., Howard, M.A., Adolphs, R., Tsuchiya, N., 2017. Intracranial markers of conscious face perception in humans. *Neuroimage* 162, 322–343. <https://doi.org/10.1016/j.neuroimage.2017.08.074>.
- Batty, M., Taylor, M.J., 2003. Early processing of the six basic facial emotional expressions. *Cogn. Brain Res.* 17 (3), 613–620. [https://doi.org/10.1016/S0926-6410\(03\)00174-5](https://doi.org/10.1016/S0926-6410(03)00174-5).
- Bernstein, M., Yovel, G., 2015. Two neural pathways of face processing: a critical evaluation of current models. *Neurosci. Biobehav. Rev.* 55, 536–546. <https://doi.org/10.1016/j.neubiorev.2015.06.010>.
- Blair, R.J.R., Coles, M., 2000. Expression recognition and behavioural problems in early adolescence. *Cogn. Dev.* 15 (4), 421–434. [https://doi.org/10.1016/S0885-2014\(01\)00039-9](https://doi.org/10.1016/S0885-2014(01)00039-9).
- Bruce, V., Young, A., 1986. Understanding face recognition. *Br. J. Psychol.* 77 (3), 305–327. <https://doi.org/10.1111/j.2044-8295.1986.tb02199.x>.
- Caharel, S., Courtay, N., Bernard, C., Lalonde, R., Rebai, M., 2005. Familiarity and emotional expression influence an early stage of face processing: an electrophysiological study. *Brain Cogn.* 59 (1), 96–100. <https://doi.org/10.1016/j.bandc.2005.05.005>.
- Calvo, M.G., Beltrán, D., 2013. Recognition advantage of happy faces: tracing the neurocognitive processes. *Neuropsychologia* 51 (11), 2051–2061. <https://doi.org/10.1016/j.neuropsychologia.2013.07.010>.
- Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011. High temporal resolution decoding of object position and category. *J. Vis.* 11 (10), 9. <https://doi.org/10.1167/11.10.9>.
- Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3), 1–27. <https://doi.org/10.1145/1961189.1961199>.
- Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A., 2016a. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* 6 (1) <https://doi.org/10.1038/srep27755>.
- Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. *Nat. Neurosci.* (3), 17. <https://doi.org/10.1038/nn.3635>.
- Cichy, R.M., Pantazis, D., Oliva, A., 2016b. Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cereb. Cortex* 26 (8), 3563–3579. <https://doi.org/10.1093/cercor/bhw135>.
- 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. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31 (3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>.
- 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. <https://doi.org/10.1002/hbm.24226>.
- Duchaine, B., Yovel, G., 2015. A revised neural framework for face processing. *Ann. Rev. Vis. Sci.* 1 (1), 393–416. <https://doi.org/10.1146/annurev-vision-082114-035518>.
- Duchaine, B., Parker, H., Nakayama, K., 2003. Normal recognition of emotion in a prosopagnosic. *Perception* 32 (7), 827–838. <https://doi.org/10.1068/p5067>.
- Ebner, N.C., Riediger, M., Lindenberger, U., 2010. FACES—a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* 42 (1), 351–362. <https://doi.org/10.3758/BRM.42.1.351>.
- Ekman, P., 1993. Facial expression and emotion. *Am. Psychol.* 48 (4), 384–392. <https://doi.org/10.1037/0003-066X.48.4.384>.
- Ester, E.F., Sprague, T.C., Serences, J.T., 2020. Categorical biases in human occipitoparietal cortex. *J. Neurosci.* 40 (4), 917–931. <https://doi.org/10.1523/JNEUROSCI.2700-19.2019>.
- Fischl, B., 2012. FreeSurfer. *Neuroimage* 62 (2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>.
- Freud, E., Behrmann, M., Snow, J.C., 2020. What does dorsal cortex contribute to perception? *Open Mind* 4, 40–56. https://doi.org/10.1162/opmi_a.00033.
- Ghuman, A.S., Brunet, N.M., Li, Y., Konecky, R.O., Pyles, J.A., Walls, S.A., Destefino, V., Wang, W., Richardson, R.M., 2014. Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* 5 (1) <https://doi.org/10.1038/ncomms6672>.
- Goeleven, E., De Raedt, R., Leyman, L., Verschuere, B., 2008. The karolinska directed emotional faces: a validation study. *Cogn. Emot.* 22 (6), 1094–1118. <https://doi.org/10.1080/02699930701626582>.

- Goesaert, E., Beeck, O., 2013. Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *J. Neurosci.* 33 (19), 8549–8558. <https://doi.org/10.1523/JNEUROSCI.1829-12.2013>.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7. <https://www.frontiersin.org/articles/10.3389/fnins.2013.00267>.
- Greening, S.G., Mitchell, D.G.V., Smith, F.W., 2018. Spatially generalizable representations of facial expressions: decoding across partial face samples. *Cortex* 101, 31–43. <https://doi.org/10.1016/j.cortex.2017.11.016>.
- 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. https://doi.org/10.1162/jocn_a.01068.
- Gross, J., 2019. Magnetoencephalography in cognitive neuroscience: a primer. *Neuron* 104 (2), 189–204. <https://doi.org/10.1016/j.neuron.2019.07.001>.
- Hamann, S., 2012. Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends Cogn. Sci. (Regul. Ed.)* 16 (9), 458–466. <https://doi.org/10.1016/j.tics.2012.07.006>.
- Han, S., Hu, J., Li, W., Zhao, S., Chen, M., Xu, P., Luo, Y., 2021. From structure to concepts: the two stages of facial expression recognition. *Neuropsychologia* 150, 107700. <https://doi.org/10.1016/j.neuropsychologia.2020.107700>.
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. <https://doi.org/10.1016/j.neuroimage.2013.10.067>.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends Cogn. Sci. (Regul. Ed.)* 4 (6), 223–233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.
- Hinojosa, J.A., Mercado, F., Carretié, L., 2015. N170 sensitivity to facial expression: a meta-analysis. *Neurosci. Biobehav. Rev.* 55, 498–509. <https://doi.org/10.1016/j.neubiorev.2015.06.002>.
- Isik, L., Meyers, E.M., Leibo, J.Z., Poggio, T., 2014. The dynamics of invariant object recognition in the human visual system. *J. Neurophysiol.* 111 (1), 91–102. <https://doi.org/10.1152/jn.00394.2013>.
- Johnston, P., Mayes, A., Hughes, M., Young, A.W., 2013. Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex* 49 (9), 2462–2472. <https://doi.org/10.1016/j.cortex.2013.01.002>.
- King, D.E., 2009. Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* 10, 1755–1758.
- Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* 12 (4) <https://doi.org/10.1038/nrn3008>.
- Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2. <https://www.frontiersin.org/articles/10.3389/fnins.2008.0004.2008>.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90. <https://doi.org/10.1145/3065386>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* (7553), 521. <https://doi.org/10.1038/nature14539>.
- Li, J., Liu, J., Liang, J., Zhang, H., Zhao, J., Huber, D.E., Rieth, C.A., Lee, K., Tian, J., Shi, G., 2009. A distributed neural system for top-down face processing. *Neurosci. Lett.* 451 (1), 6–10. <https://doi.org/10.1016/j.neulet.2008.12.039>.
- Li, Y., Zhang, M., Liu, S., Luo, W., 2022. EEG decoding of multidimensional information from emotional faces. *Neuroimage* 258, 119374. <https://doi.org/10.1016/j.neuroimage.2022.119374>.
- Liang, Y., Liu, B., Xu, J., Zhang, G., Li, X., Wang, P., Wang, B., 2017. Decoding facial expressions based on face-selective and motion-sensitive areas: decoding facial expressions. *Hum. Brain Mapp.* 38 (6), 3113–3125. <https://doi.org/10.1002/hbm.23578>.
- Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I., 2010. The extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp. 94–101. <https://doi.org/10.1109/CVPRW.2010.5543262>.
- Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J., 1998. Coding facial expressions with Gabor wavelets. In: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205. <https://doi.org/10.1109/AFGR.1998.670949>.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Muukkonen, I., Ölander, K., Numminen, J., Salmela, V.R., 2020. Spatio-temporal dynamics of face perception. *Neuroimage* 209, 116531. <https://doi.org/10.1016/j.neuroimage.2020.116531>.
- Nestor, A., Plaut, D.C., Behrmann, M., 2016. Feature-based face representations and image reconstruction from behavioral and neural data. *Proc. Natl. Acad. Sci.* 113 (2), 416–421. <https://doi.org/10.1073/pnas.1514551112>.
- 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. (Regul. Ed.)* 10 (9), 424–430. <https://doi.org/10.1016/j.tics.2006.07.005>.
- Pantazis, D., Nichols, T.E., Baillet, S., Leahy, R.M., 2005. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *Neuroimage* 25 (2), 383–394. <https://doi.org/10.1016/j.neuroimage.2004.09.040>.
- Peirce, J.W., 2007. PsychoPy—psychophysics software in python. *J. Neurosci. Methods* 162 (1), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>.
- Pitcher, D., Ungerleider, L.G., 2021. Evidence for a third visual pathway specialized for social perception. *Trends Cogn. Sci. (Regul. Ed.)* 25 (2), 100–110. <https://doi.org/10.1016/j.tics.2020.11.006>.
- Ramkumar, P., Jas, M., Pannasch, S., Hari, R., Parkkonen, L., 2013. Feature-specific information processing precedes concerted activation in human visual cortex. *J. Neurosci.* 33 (18), 7691–7699. <https://doi.org/10.1523/JNEUROSCI.3905-12.2013>.
- Rellecke, J., Sommer, W., Schacht, A., 2012. Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biol. Psychol.* 90 (1), 23–32. <https://doi.org/10.1016/j.biopsycho.2012.02.002>.
- Rellecke, J., Sommer, W., Schacht, A., 2013. Emotion effects on the N170: a question of reference? *Brain Topogr.* 26 (1), 62–71. <https://doi.org/10.1007/s10548-012-0261-y>.
- Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2 (11) <https://doi.org/10.1038/14819>.
- Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V., 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vis.* 10 (5), 11. <https://doi.org/10.1167/10.5.11>. –11.
- Sarkheil, P., Goebel, R., Schneider, F., Mathiak, K., 2013. Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. *Soc. Cogn. Affect. Neurosci.* 8 (8), 950–957. <https://doi.org/10.1093/scan/nss092>.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556). arXiv: 10.48550/arXiv.1409.1556.
- Smith, F.W., Smith, M.L., 2019. Decoding the dynamic representation of facial expressions of emotion in explicit and incidental tasks. *Neuroimage* 195, 261–271. <https://doi.org/10.1016/j.neuroimage.2019.03.065>.
- Sullivan, S., Ruffman, T., 2004. Emotion recognition deficits in the elderly. *Int. J. Neurosci.* 114 (3), 403–432. <https://doi.org/10.1080/00207450490270901>.
- Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D. J., Westerlund, A., Casey, B.J., Nelson, C., 2009. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168 (3), 242–249. <https://doi.org/10.1016/j.psychres.2008.05.006>.
- Tracy, J.L., Robins, R.W., 2008. The automaticity of emotion recognition. *Emotion* 8 (1), 81–95. <https://doi.org/10.1037/1528-3542.8.1.81>.
- Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A.L., McGettigan, C., Garrido, L., 2021. FFA and OFA encode distinct types of face identity information. *J. Neurosci.* 41 (9), 1952–1969. <https://doi.org/10.1523/JNEUROSCI.1449-20.2020>.
- Vida, M.D., Nestor, A., Plaut, D.C., Behrmann, M., 2017. Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proc. Natl. Acad. Sci.* 114 (2), 388–393. <https://doi.org/10.1073/pnas.1614763114>.
- Wegrzyn, M., Riehle, M., Labudde, K., Woermann, F., Baumgartner, F., Pollmann, S., Bien, C.G., Kissler, J., 2015. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* 69, 131–140. <https://doi.org/10.1016/j.cortex.2015.05.003>.
- Weibert, K., Flack, T.R., Young, A.W., Andrews, T.J., 2018. Patterns of neural response in face regions are predicted by low-level image properties. *Cortex* 103, 199–210. <https://doi.org/10.1016/j.cortex.2018.03.009>.
- Wronka, E., Walentowska, W., 2011. Attention modulates emotional expression processing. *Psychophysiology* 48 (8), 1047–1056. <https://doi.org/10.1111/j.1469-8986.2011.01180.x>.
- Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.). *Computer Vision – ECCV 2014*. Springer International Publishing, pp. 818–833. https://doi.org/10.1007/978-3-319-10590-1_53.
- Zhang, H., Japee, S., Nolan, R., Chu, C., Liu, N., Ungerleider, L.G., 2016. Face-selective regions differ in their ability to classify facial expressions. *Neuroimage* 130, 77–90. <https://doi.org/10.1016/j.neuroimage.2016.01.045>.
- Zhang, L., Song, Y., Liu, L., Liu, J., 2016. Dissociable roles of internal feelings and face recognition ability in facial expression decoding. *Neuroimage* 132, 283–292. <https://doi.org/10.1016/j.neuroimage.2016.02.049>.