

Posterior and Mid-Fusiform Contribute to Distinct Stages of Facial Expression Processing

Authors: Yuanning Li^{1,2,3*}, R. Mark Richardson^{1,3}, and Avniel Singh Ghuman^{1,2,3}

Affiliations:

¹Center for the Neural Basis of Cognition, Carnegie Mellon University and University of Pittsburgh, Pittsburgh, PA

²Program in Neural Computation, Carnegie Mellon University and University of Pittsburgh, Pittsburgh, PA

³Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA

***Correspondence to:**

Yuanning Li
S906 Scaife Hall
3550 Terrace Street
Pittsburgh, PA 15261
ynli@cmu.edu

Keywords: face perception; fusiform; intracranial electroencephalography; multivariate temporal pattern analysis; temporal dynamics

Abstract

Though the fusiform is well-established as a key node in the face perception network, its role in facial expression processing remains unclear, due to competing models and discrepant findings. To help resolve this debate, we recorded from 17 subjects with intracranial electrodes implanted in face sensitive patches of the fusiform. Multivariate classification analysis showed that facial expression information is represented in fusiform activity, in the same regions that represent identity, though with a smaller effect size. Examination of the spatiotemporal dynamics revealed a functional distinction between posterior and mid-fusiform expression coding, with posterior fusiform showing an early peak of facial expression sensitivity at around 180 ms after subjects viewed a face and mid-fusiform showing a later and extended peak between 230 – 460 ms. These results support the hypothesis that the fusiform plays a role in facial expression perception and highlight a qualitative functional distinction between processing in posterior and mid-fusiform, with each contributing to temporally segregated stages of expression perception.

Introduction

Face perception, including detecting a face, recognizing face identity, assessing sex, age, emotion, attractiveness, and other characteristics associated with the face, is critical to social communication. An influential cognitive model of face processing distinguishes processes associated with recognizing the identity of a face from those associated with recognizing expression ¹. A face sensitive region of the lateral fusiform gyrus, sometimes called the fusiform face area, is a critical node in the face processing network ²⁻⁵ that has been shown to be involved in identity perception ⁶⁻¹⁰. What role, if any, the fusiform plays in face expression processing continues to be debated, particularly given the hypothesized cognitive distinction between identity and expression perception.

Results demonstrating relative insensitivity of the fusiform to face dynamics ¹¹ and reduced fusiform activity for attention to gaze direction ¹² led to a model that proposed that this area was involved strictly in identity perception and not expression processing ². This model provided neuroscientific grounding for the earlier cognitive model that hypothesized a strong division between identity and expression perception ¹. However, some recent imaging studies directly probing whether the fusiform is sensitive to expression have shown mixed results ^{2,13-16}. Positive findings for fusiform sensitivity to expression have led to the competing hypothesis that the division of face processing is not for identity and expression, but rather form/structure and motion ³. Notably though, even studies with positive results have not examined whether the same patches of the fusiform that code for identity also code for expression (see ¹⁴ for a study that examined both, but saw negative results for expression coding). Furthermore, some studies show the fusiform has an expression-independent identity code ^{6,8}.

Beyond whether the fusiform responds differentially to expression, one key question is whether the fusiform intrinsically codes for expression or if differential responses are due to task-related and/or top-down modulation of fusiform activity¹⁷. Assessing this requires a method with high temporal resolution to distinguish between early, more bottom-up biased activity, and later activity that likely involved recurrent interactions. Furthermore, a passive viewing or incidental task is required to exclude biases introduced by variable task demands across stimuli. The low temporal resolution of fMRI makes it difficult to disentangle early bottom-up processing from later top-down and recurrent processing⁸. Some previous intracranial electroencephalography (iEEG) studies have used an explicit expression identification task, making task effects difficult to exclude¹⁸⁻²⁰. Those that have used an implicit task have shown mixed results regarding whether early fusiform response is sensitive to expression^{20,21}. Furthermore, iEEG studies often lack sufficient subjects and population-level analysis to allow for a generalizable interpretation.

To help mediate between these two models and clarify the role of the fusiform in facial expression perception, iEEG was recorded from 17 subjects with a total of 31 face sensitive electrodes in face sensitive patches of the fusiform gyrus while these subjects viewed faces with neutral, happy, sad, angry, and fearful expressions in a gender discrimination task. Multivariate temporal pattern analysis (MTPA) on the data from these electrodes was used to analyze the temporal dynamics of neural activity with respect to facial expression sensitivity in fusiform. In a subset of 7 subjects, identity coding was examined in the same electrodes also using MTPA. In addition to examining the overall patterns across all electrodes, the responses from and mid- and posterior

fusiform, as well as the left and right hemisphere, were compared. To supplement these iEEG results, a meta-analysis of 64 neuroimaging studies was done examining facial expression sensitivity in the fusiform. The results support the view that fusiform response is sensitive to facial expression and suggest that the posterior and mid-fusiform regions play a qualitatively different role in facial expression processing.

Results

Electrode selection and face sensitivity

The locations of the 31 fusiform electrodes from 17 participants sensitive to faces are shown in Figure 1A and Table 1. The averaged event-related potential (ERP) and event-related broadband gamma activity (ERBB) responses (see Methods for detailed definitions of ERP and ERBB) for each category across all channels are shown in Figure 1C and Figure 1D respectively. The averaged sensitivity index (d') for faces peaked at 160 ms ($d' = 1.22$, $p < 0.01$ in every channel, Figure 1B). Consistent with previous findings^{8,22-24}, a strong sensitivity for faces was observed in fusiform around 100-400 ms after stimulus onset.

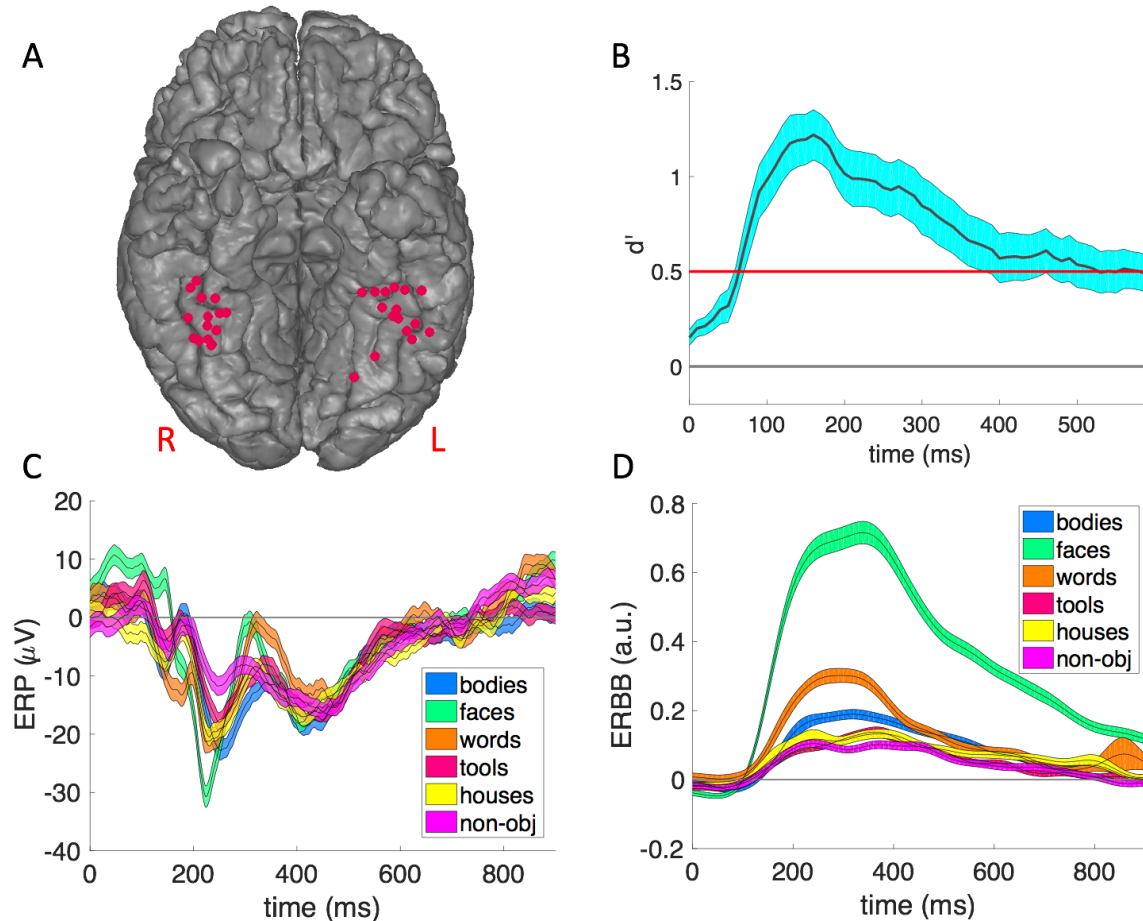


Figure 1. The face sensitive electrodes in the fusiform. **A)** The localization of the 31 face sensitive electrodes in (or close to) fusiform area, mapped onto a common space based on MNI coordinates. We moved depth electrode locations to the nearest location on the overlying cortical surface, in order to visualize all the electrodes. **B)** The timecourse of the sensitivity index (d') for faces versus the other categories in the six-way classification averaged across all 31 fusiform electrodes. The shaded areas indicate standard error of the mean. The red line corresponds to $p < 0.01$ with Bonferroni correction for multiple comparisons across 60 time points. **C)** The ERP for each category averaged across all face sensitive fusiform electrodes. The shaded areas indicate standard error of the mean. **D)** The ERBB for each category averaged across all face sensitive fusiform electrodes. The shaded areas indicate standard error of the mean.

Table 1 MNI coordinates and facial expression sensitivity (d') for all face sensitive electrodes. Electrode ID is labeled by subject number (SX) and electrode from that subject (a, b, etc.). Sensitivity to expression defined as $p < 0.05$ decoding accuracy corrected for multiple comparisons.

| Electrode ID | X (mm) | Y (mm) | Z (mm) | Peak time (ms) | Peak d' | Sensitive to expressions |
|--------------|--------|--------|--------|----------------|-----------|--------------------------|
| S1a | 35 | -59 | -22 | 260 | 0.29 | Y |
| S1b | 33 | -53 | -22 | 150 | 0.31 | Y |
| S1c | 42 | -56 | -26 | 200 | 0.20 | N |
| S2a | 40 | -57 | -23 | 170 | 0.34 | Y |
| S3a | -33 | -44 | -31 | 580 | 0.18 | N |
| S4a | -38 | -36 | -30 | 440 | 0.12 | N |
| S5a | -38 | -36 | -20 | 300 | 0.25 | Y |
| S5b | -42 | -37 | -19 | 330 | 0.25 | Y |
| S6a | 34 | -40 | -11 | 540 | 0.24 | Y |
| S6b | 39 | -40 | -10 | 490 | 0.33 | Y |
| S7a | 36 | -57 | -21 | 100 | 0.42 | Y |
| S8a | -22 | -72 | -9 | 100 | 0.23 | Y |
| S8b | -40 | -48 | -23 | 170 | 0.38 | Y |
| S9a | 32 | -46 | -7 | 180 | 0.34 | Y |
| S9b | 36 | -48 | -8 | 160 | 0.40 | Y |
| S10a | 29 | -46 | -15 | 310 | 0.31 | Y |
| S11a | -25 | -38 | -17 | 580 | 0.36 | Y |
| S11b | -34 | -38 | -18 | 400 | 0.46 | Y |
| S11c | -49 | -37 | -20 | 430 | 0.27 | Y |
| S12a | 41 | -33 | -19 | 70 | 0.06 | N |
| S12b | 37 | -51 | -9 | 70 | 0.22 | Y |
| S12c | 35 | -59 | -4 | 80 | 0.23 | Y |
| S13a | 43 | -36 | -13 | 400 | 0.11 | N |
| S13b | 44 | -48 | -11 | 190 | 0.10 | N |
| S14a | -52 | -54 | -17 | 30 | 0.14 | N |
| S15a | -37 | -47 | -10 | 180 | 0.64 | Y |
| S16a | -39 | -45 | -11 | 160 | 0.03 | N |
| S17a | -43 | -53 | -26 | 90 | 0.20 | N |
| S17b | -46 | -50 | -28 | 110 | 0.31 | Y |
| S17c | -30 | -63 | -20 | 120 | 0.27 | Y |
| S17d | -45 | -56 | -25 | 40 | 0.13 | N |

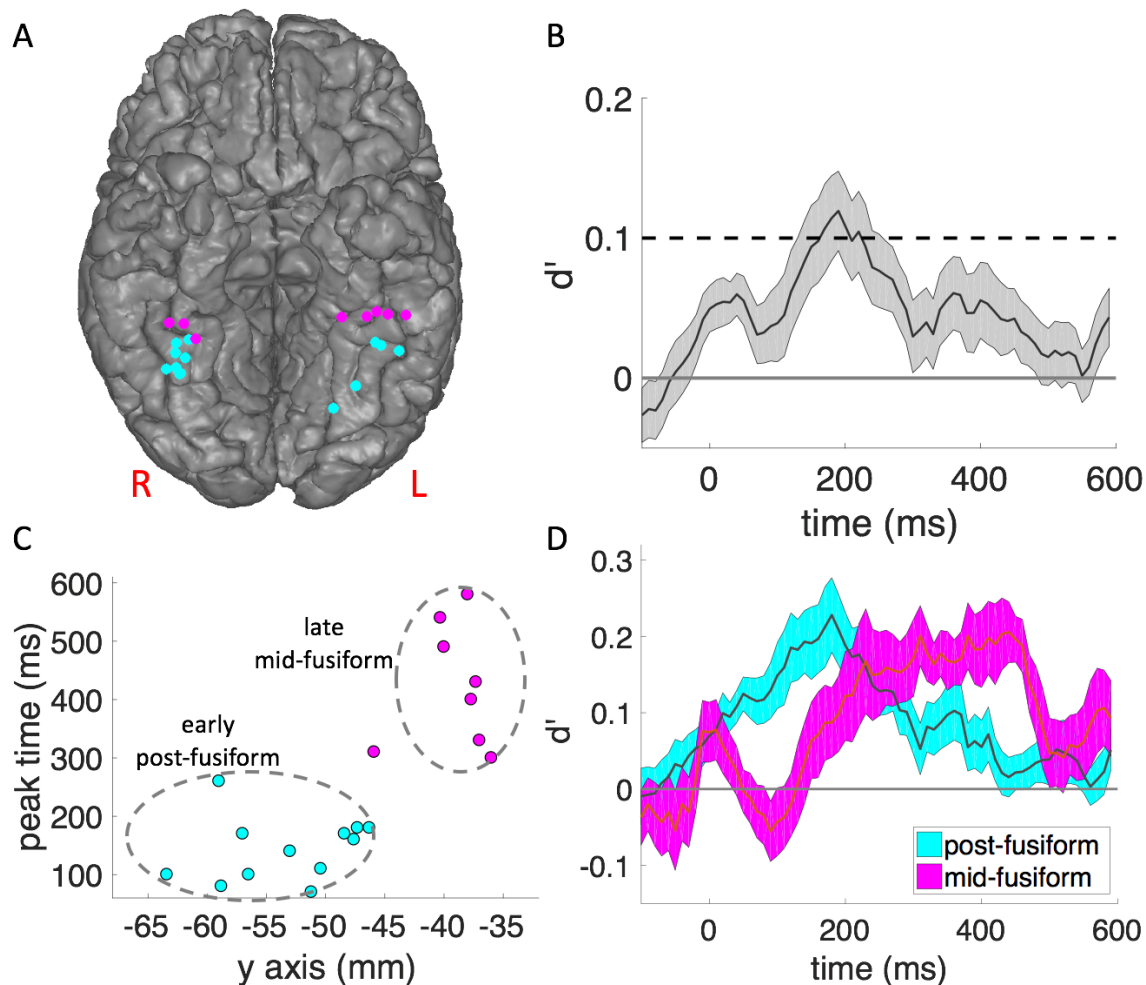


Figure 2. The timecourse of the facial expression classification in fusiform. **A)** The locations of the electrodes with significant face expression decoding accuracy, with the posterior fusiform group colored in cyan and the mid-fusiform group colored in magenta. **B)** The timecourse of mean and standard error for pairwise classification between different face expressions in all 31 fusiform electrodes. Dashed line: $p = 0.05$ threshold with Bonferroni correction for 60 time points [600 ms with 10 ms stepsize]). **C)** The time of the peak classification accuracy was plotted against the MNI y-coordinate for each single electrode with significant expression classification accuracy. *K*-means clustering partitions these electrodes into posterior and mid- fusiform groups. **D)** The mean and standard error for pairwise classification between different face expressions in posterior fusiform electrodes and mid- fusiform electrodes. The posterior group peaked at 180 ms after stimulus onset and the mid-fusiform group had an extended peak starting at 230 ms and extending to 450 ms (both $p < 0.05$, binomial test, Bonferroni corrected). See supplement for receiver operator characteristic (ROC) curves validating classification analysis (Figure S1).

Facial expression classification at the individual and group level

For each participant, the classification accuracy between each pair of facial expressions was estimated using 5-fold cross-validation (see Methods for details). As shown in Figure 2B, the averaged timecourse peaked at 190 ms after stimulus onset (average decoding at peak $d' = 0.12$, $p < 0.05$, Bonferroni corrected for multiple comparisons). In addition to the grand average, on the single electrode level, 21 out of the 31 electrodes from 12 out of 17 subjects showed a significant peak in their individual timecourses ($p < 0.05$, permutation test corrected for multiple comparisons). The locations of the significant electrodes are shown in Figure 2A and all electrodes are listed in Table 1.

The effect size for the mean peak expression classification is relatively low. This is in part because the electrodes consisted of two distinct populations with different timecourses (see below). Additionally, due to the variability in electrode position, iEEG effect sizes can be lower in some cases than what would be seen with electrodes optimally placed over face patches. To assess whether this was the case, we examined the correspondence between face category decoding and expression decoding based on the logic that placement closer to face patches should lead to higher face category decoding accuracy. A significant positive correlation between the decoding accuracy (d') for face category and the decoding accuracy (d') for facial expressions was seen (Pearson correlation $r = 0.57$, $N = 21$, $p = 0.007$). This suggests that electrode position relative to face patches in the fusiform can explain some of the effect size variability for expression classification. That suggests the true effect size for expression classification for optimal

electrode placement may be closer to what was seen for electrodes with higher accuracy (0.4-0.6, see Table 1) rather than the mean across all electrodes.

Spatiotemporal dynamics of facial expression decoding

The next question we addressed was whether spatiotemporal dynamics of facial expression representation in fusiform was location dependent. Specifically, we compared the dynamics of expression sensitivity between left and right hemispheres, and between posterior and mid- fusiform regions for electrodes showing significant expression sensitivity.

We first analyzed the lateralization effect for the expression coding in fusiform. The mean timecourses of decoding accuracy for left and right fusiform did not differ at the $p < 0.05$ uncorrected level at any time point (Figure S2).

In contrast substantial differences were seen in the timing and representation of expression coding between posterior and mid- fusiform. This was first illustrated by plotting the time of the peak decoding accuracy in each individual electrode against the corresponding MNI y-coordinate of the electrode (Figure 2C). A qualitative difference was seen between the peak times for electrodes posterior to approximately $y = -45$ compared to those anterior to that, rather than a continuous relationship between y-coordinate and peak time. This was quantified by a clustering analysis using both Bayesian information criterion (BIC)²⁵ and Silhouette analysis²⁶ (Figure S3), which both showed evidence for a cluster-structure in the data (Bayes factor > 20) with $k = 2$ as the optimal number of clusters (mean Silhouette coefficient = 0.59). The 2 clusters corresponded to the posterior and mid-fusiform (Figure 2C; see Supplemental

Information for detailed analysis on clustering and the selection of models with different values of k). The border between these data-driven clusters corresponds well with prior functional and anatomical evidence showing that the mid-fusiform face patch falls within a 1 cm disk centered around the anterior tip of mid-fusiform sulcus (MFS; which falls at $y = -40$ in MNI coordinates) with high probability²⁷. That would make the border between the mid-fusiform and posterior fusiform face patch approximately $y = -45$ in MNI coordinates, which is very close to the border produced by the clustering analysis ($y = -45.9$).

The timecourse of the posterior and mid- fusiform clusters were then examined in detail. As shown in Figure 2D, the timecourse of decoding accuracy in the posterior group peaked at 180 ms after stimulus onset and the timecourse of mid-fusiform group first peaked at 230 ms and the peak extended until approximately 450 ms after stimulus onset.

Representational Similarity Analysis

A recent meta-analysis suggests that fusiform is particularly sensitive to the contrast between specific pairs of expressions²⁸. To examine this in iEEG data, the representation dissimilarity matrices (RDMs) for facial expressions in the early and late activity in posterior and mid-fusiform were computed (Figure 3). No contrasts between expressions showed significant differences in posterior fusiform in the late window or in mid-fusiform in the early window ($p > 0.1$ in all cases, T-test), as expected due to the corresponding low overall classification accuracy. In the early posterior fusiform, expressions of negative emotions (fearful, angry) were dissimilar to happy and neutral

expressions ($p < 0.05$ in each case, T-test), but not very distinguishable from one another. In the late mid-fusiform activity, happy and neutral expressions were both distinguishable from expressions of negative emotions and from each other ($p < 0.05$ in each case, T-test). The results showed partial consistency with a previous meta-analysis based on neuroimaging studies (consistent in angry vs. neutral, fearful vs. neutral, fearful vs. happy, and fear vs sad)²⁸. However, the previous meta-analysis also reported significant contrast in fearful vs. angry and angry vs. sad, which were absent in our results.

One question is the degree to which the representation in fusiform reflects the physical properties of the images subjects were viewing versus a more abstract representation of emotion. To examine this question, an 17-dimensional facial feature space was constructed based on a computer vision algorithm²⁹. The features characterize structural and spatial frequency properties of each image, e.g. eye width, eyebrow length, nose height, eye-mouth width ratio, skin tone, etc. An RDM was then built between the expressions in this feature space and compared to the neural feature spaces. There was a significant correlation between posterior fusiform representation space in the early time window (Spearman's $\rho = 0.24$, $p < 0.05$, permutation test). The correlation between mid- fusiform representation space in the late time window and the facial feature space was smaller and did not reach statistical significance (Spearman's $\rho = 0.15$, $p > 0.1$, permutation test). This suggests the earlier representation reflects the physical aspects of the images more closely whereas the later representation may also reflect the emotional content more abstractly.

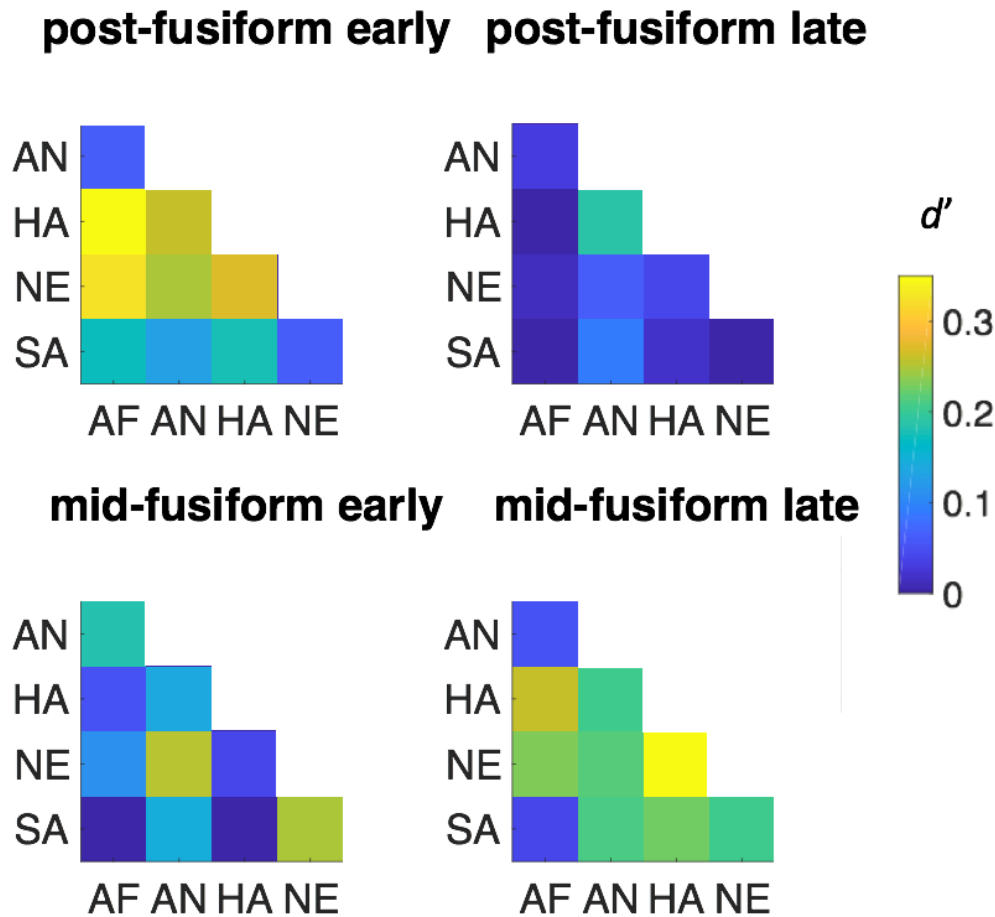


Figure 3. **Representational similarity analysis (RSA) between the facial feature space and the representational spaces of posterior and mid- fusiform at both early and late stages.** **Top row:** representational dissimilarity matrices (RDM) of posterior fusiform at early stage (left), RDM of posterior fusiform at late stage (right). **Bottom row:** RDM of mid-fusiform at early stage (left), RDM of mid-fusiform at late stage (right). Abbreviations: AF – fearful, AN – angry, HA – happy, NE – neutral, SA – sad.

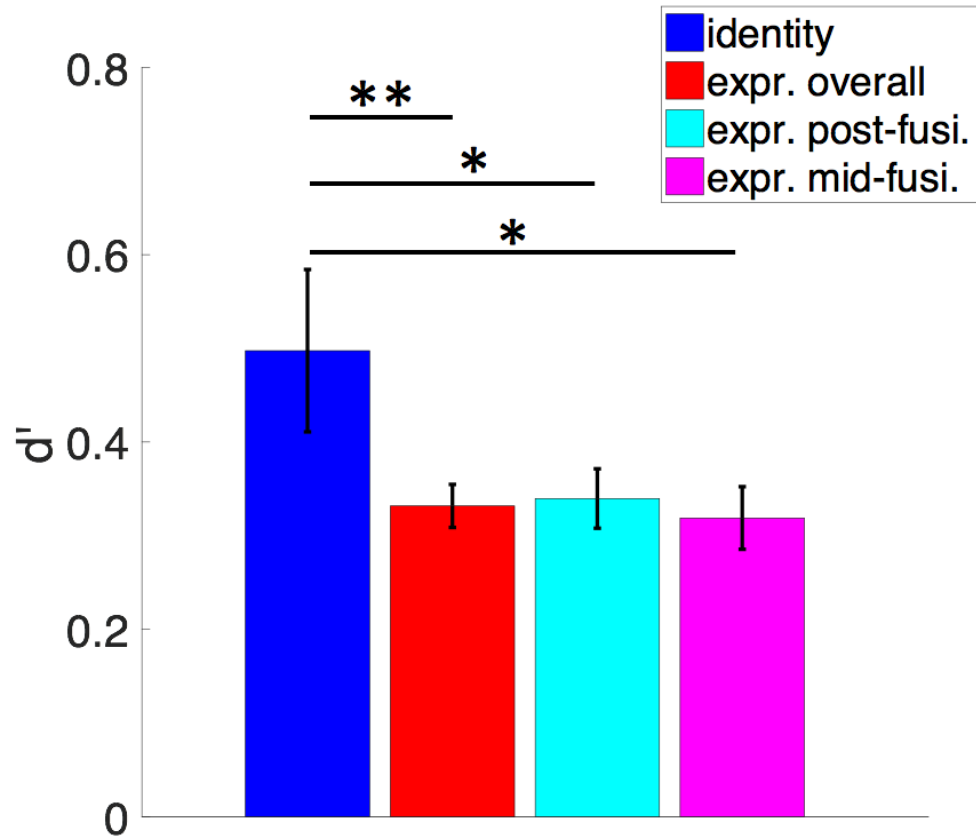


Figure 4. **The comparison between the decoding accuracy for face identities and facial expressions.** The average peak d' for face identity decoding (in blue) compared to the average d' for facial expression decoding overall (in red), in posterior fusiform (in cyan), and in anterior fusiform (in magenta). (error bar: standard error, ** $p < 0.01$, * $p < 0.05$, T-test).

Comparison to Facial Identity Classification

Given the strongly supported hypothesis the fusiform plays a central role in face identity recognition, the effect size of identity and expression coding in the fusiform was compared. Due to the relatively few repetitions of individual faces, individuation was examined in only the 7 subjects that had sufficient repetitions of each face identity allowing for multivariate classification of identity across expression; identity decoding was previously reported for 4 of these subjects in a recent study⁸. Across the 7 total subjects (3 here and 4 reported previously), the mean peak $d' = 0.50$ for face identity

decoding was significantly greater than the mean peak accuracy for facial expression decoding in the fusiform overall ($t(26) = 2.64, p = 0.0069$) as well as in both the posterior ($t(18) = 2.08, p = 0.026$) and mid-fusiform ($t(13) = 2.02, p = 0.032$) (Figure 4). With regards to the timing of identity (mean peak time = 314 ms) versus expression sensitivity, the posterior peak time for expression classification was significantly earlier than the peak time for identity ($t(18) = 4.45, p = 0.0003$), while the mid-fusiform extended peak time for expression classification overlapped with the peak time for identity.

Discussion

Multivariate classification methods were used to evaluate the encoding of facial expressions recorded from electrodes placed directly in face sensitive fusiform cortex. Though the effect size for expression classification is smaller than for identity classification, the results support a role for the fusiform in the processing of facial expressions. Electrodes that were sensitive to expression were also sensitive to identity, suggesting a shared neural substrate for identity and expression coding in the fusiform. The results also show that the posterior and mid- fusiform are dynamically involved in distinct stages of facial expression processing and have different representations of expressions. The differential representation and magnitude of the temporal displacement between the sensitivity in posterior and mid-fusiform suggests these are qualitatively distinct stages of facial expression processing and not merely a consequence of transmission or information processing delay along a feedforward hierarchy.

Fusiform is sensitive to facial expression

The results here show that the fusiform is sensitive to expression, though the effect size for classification of expression in the fusiform using iEEG is small-to-medium^{1 30}. The results also suggest that the same patches of the fusiform that are sensitive to expression are sensitive to identity as well. Given the variability of the effect size due to the proximity of electrode placement relative to face patches, the relative effect size may be more informative than the absolute effect size. The magnitude for facial expression classification is approximately half what was seen for face identity classification. This suggests that while fusiform contributes to facial expression perception, it is to a lesser degree than face identity processing. Greater involvement in identity than expression perception is expected for a region involved in structural processing of faces because identity relies on this information more than expression as expression perception also relies on facial dynamics. These results support models that hypothesize fusiform involvement in form/structural processing, at least for posterior fusiform (see discussion on spatially and temporally segregated stages of processing below), which can support facial expression processing^{3,5}. These results do not support models that hypothesize a strong division between facial identity and expression processing^{1,2}.

To test what a brain region codes for one must examine its response for early, bottom-up activation during an incidental task or passive viewing¹⁷, otherwise it is difficult to disentangle effects of task demands and top-down modulation. Indeed,

¹ d' is on the same scale as Cohen's d and they are equivalent when the data is univariate Gaussian, so d 's between 0.2-0.5 are "small" and 0.5-0.8 are medium.

previous studies have demonstrated that extended fusiform activity, particularly in the broadband gamma range, is modulated by task-related information^{8,31}. Some previous iEEG studies of expression coding in the fusiform have used an explicit expression judgment task and examined only broadband gamma activity, making it difficult to draw definitive conclusions about fusiform expression coding from these results^{18,19}. One previous study that used an implicit task did not show evidence of expression sensitivity during the early stage of activity in the fusiform²⁰; another did show evidence of expression sensitivity, though it reported results only from a single subject²¹. The results here show in a large iEEG sample that the early response of the fusiform most sensitive to bottom-up processing is modulated by expression, at least for the posterior fusiform.

The effect size for facial expression classification is consistent with mixed findings in the neuroimaging literature for expression sensitivity in the fusiform^{2,13-16,19}. IEEG generally has greater sensitivity and lower noise than non-invasive measures of brain activity. Methods with lower sensitivity, such as fMRI, would be expected to have a substantial false negative rate for facial expression coding in the fusiform. To quantify fMRI sensitivity to expression we performed a meta-analysis on 64 studies. Of these studies, 24 reported at least one expression sensitive loci in the fusiform. However, at the meta-analytic level, no significant cluster of expression sensitivity was seen in the fusiform after whole brain analysis (see Supplemental Table S1, S2, and Figure S4). Thus, consistent with the iEEG effect size for expression decoding in the fusiform seen here, there is some suggestion in the fMRI literature for expression sensitivity in the fusiform, but it is small and does not achieve statistical significance at the whole brain level.

Multiple, spatially and temporally segregated stages of face expression processing in the fusiform

Using a data-driven analysis, posterior and mid-fusiform face patches were shown to contribute differentially to expression processing. The dividing point between post-fusiform and mid-fusiform electrodes found in a data-driven manner is consistent with the anatomical border for the posterior and mid-fusiform face patches previously described, suggesting a strong coupling between the anatomical and functional divisions in fusiform²⁷. While posterior and mid-fusiform have been shown to be cytoarchitectonically distinct regions each with separate face sensitive patches^{27,32,33}, functional differences between these patches have remained elusive in the literature. The results here suggest that these anatomical and physiological distinctions correspond to functional distinctions in the role of these areas in face processing, as reflected in qualitatively different temporal dynamics in these regions for facial expression processing. Specifically, posterior fusiform participates in a relatively early stage of facial expression processing that may be related to structural encoding of faces while mid-fusiform demonstrates a distinct pattern of extended dynamics and participates in a later stage of processing that may be related to a more abstract and/or multifaceted representation of expression and emotion. These results support the revised model of fusiform function that posits the fusiform contributes to structural encoding of facial expression during the initial stages of processing^{3,5}, with the notable addition that it may be that only posterior fusiform contributes to this processing.

The early time period of expression sensitivity in posterior fusiform overlaps with strong face sensitive activity measured non-invasively around 170 ms after viewing a face, which is thought to reflect structural encoding of face information^{22,23,34-36}. Face sensitive activity in this time window has been shown to be insensitive to attention and is thought to reflect a “rapid, feed-forward phase of face-selective processing.”³⁷ Additionally, a face adaptation study showed that activity in this window reflects the actual facial expression rather than the perceived (adapted) expression³⁸. Consistent with these previous findings, the RSA results here show that the early posterior activity is more closely correlated to the physical features of the face that relate to facial expression compared to later, mid-fusiform activity.

The expression sensitivity in mid-fusiform onset began later than the posterior fusiform (around 230 ms), and remained active until ~450 ms after viewing a face. Face sensitive activity in this time window has been shown to be sensitive to face familiarity and to attention^{39,40}. Previous studies and the results presented here show that face identity can be decoded from the activity in this later time window in mid-fusiform^{8,41} and reflects a distributed code for identity among regions of the face processing network⁴². Additionally, the previously mentioned face adaptation study showed that activity in this window reflects the subjectively perceived facial expression after adaptation³⁸. The RSA analysis here showed that the activity in this time window in mid-fusiform was not significantly correlated with physical features of the face and therefore may reflect more subjective expression perception. Taken together, these results suggest the mid-fusiform expression sensitivity in this later window reflect a more abstract and subjective representation of expression and may be related to integration of multiple face cues,

including identity and expression. This abstract and multifaceted representation is likely to reflect processes arising from interactions across the face processing network ⁴.

To conclude, the results presented here support the hypothesis that the fusiform contributes to expression processing ^{3,5}. The finding that the same part of the fusiform is sensitive to both identity and expression contradicts models that hypothesize separate pathways for their processing ^{1,2} and instead supports the hypothesis that form and motion are the critical functional separation ³. The results also show there is a qualitative distinction between face processing in posterior and mid-fusiform, with each contributing to temporally and functionally distinct stages of expression processing. This distinct contribution of these two fusiform patches suggest that the structural and cytoarchitectonic differences between posterior and mid-fusiform are associated with functional differences between the contributions of these areas to face perception. Taken together, the results here illustrate the dynamic role the fusiform plays in multiple stages of facial expression processing.

Methods

Participants

The experimental protocols were approved by the Institutional Review Board of the University of Pittsburgh. Written informed consent was obtained from all participants.

17 human subjects (8 male, 9 female) underwent surgical placement of subdural electrocorticographic electrodes or stereoelectroencephalography (together

electrocorticography and stereoelectroencephalography are referred to here as iEEG) as standard of care for seizure onset zone localization. The ages of the subjects ranged from 19 to 65 years old (mean = 37.9, SD = 12.7). None of the participants showed evidence of epileptic activity on the fusiform electrodes used in this study nor any ictal events during experimental sessions.

Experiment design

In this study, each subject participated in two experiments. Experiment 1 was a functional localizer experiment and Experiment 2 was a face perception experiment. The experimental paradigms and the data pre-processing methods were similar to those described previously by Ghuman and colleagues ⁸.

Stimuli

In Experiment 1, 180 images of faces (50% male), bodies (50% male), words, hammers, houses, and phase scrambled faces were used as visual stimuli. Each of the six categories contained 30 images. Phase scrambled faces were created in MATLABTM by taking the 2-dimensional spatial Fourier spectrum of each of the face images, extracting the phase, adding random phases, recombining the phase and amplitude, and taking the inverse 2-dimensional spatial Fourier spectrum.

In Experiment 2, face stimuli were taken from the Karolinska Directed Emotional Faces stimulus set. Frontal views and 5 different facial expressions (fearful, angry, happy, sad, and neutral) from 70 faces (50% male) in the database were used, which yielded a total of 200 unique images.

Paradigms

In Experiment 1, each image was presented for 900 ms with 900 ms inter-trial interval during which a fixation cross was presented at the center of the screen ($\sim 10^\circ \times 10^\circ$ of visual angle). At random, 1/3 of the time an image would be repeated, which yielded 480 independent trials in each session. Participants were instructed to press a button on a button box when an image was repeated (1-back).

In Experiment 2, each face image was presented for 1500 ms with 500 ms inter-trial interval during which a fixation cross was presented at the center of the screen. This yielded 200 independent trials per session. Faces subtended approximately 5 degrees of visual angle in width. Subjects were instructed to report whether the face was male or female via button press on a button box.

Paradigms were programmed in MATLABTM using Psychtoolbox and custom written code. All stimuli were presented on an LCD computer screen placed approximately 150 cm from participants' heads.

All of the participants performed one session of Experiment 1. 9 of the subjects performed one session of Experiment 2, and the other 8 participants performed two or more sessions of Experiment 2.

Data analysis

Data preprocessing

The electrophysiological activity was recorded using iEEG electrodes at 1000 Hz. Single-trial potential signal was extracted by band-passing filtering the raw data between 0.2-115 Hz using a forth order Butterworth filter to remove slow and linear drift, and high frequency noise. The 60 Hz line noise was removed using a forth order Butterworth

filter with 55-65 Hz stop-band. Power spectrum density (PSD) at 2 – 100 Hz with bin size of 2 Hz and time-step size of 10 ms was estimated for each trial using multi-taper power spectrum analysis with Hann tapers, using FieldTrip toolbox⁴³. For each channel, the neural activity between 50 and 300 ms prior to stimulus onset was used as baseline, and the PSD at each frequency was then z-scored with respect to the mean and variance of the baseline activity to correct for the power scaling over frequency at each channel. The broadband gamma signal was extracted as mean z-scored PSD across 40-100 Hz. Event-related potential (ERP) and event-related broadband gamma signal (ERBB), both time-locked to the onset of stimulus from each trial, were used in the following data analysis.

To reduce potential artifacts in the data, raw data were inspected for ictal events, and none were found during experimental recordings. Trials with maximum amplitude 5 standard deviations above the mean across all the trials were eliminated. In addition, trials with a change of more than 25 μ V between consecutive sampling points were eliminated. These criteria resulted in the elimination of less than 1% of trials.

Electrode localization

Coregistration of grid electrodes and electrode strips was adapted from the method of Hermes, et al.⁴⁴. Electrode contacts were segmented from high resolution post-operative CT scans of patients coregistered with anatomical MRI scans before neurosurgery and electrode implantation. The Hermes method accounts for shifts in electrode location due to the deformation of the cortex by utilizing reconstructions of the cortical surface with FreeSurferTM software and co-registering these reconstructions with a high-resolution

post-operative CT scan. SEEG electrodes were localized with Brainstorm software⁴⁵ using post-operative MRI co-registered with pre-operative MRI images.

Electrode selection

Face sensitive electrodes were selected based on both anatomical and functional constraints. Anatomical constraint was based upon the localization of the electrodes on the reconstruction using post-implantation MRI. In addition, multivariate temporal pattern analysis (MTPA) was used to functionally select the electrodes that showed sensitivity to faces, comparing to other conditions in the localizer experiment (see below for MTPA details). Specifically, three criteria were used to screen and select the electrodes of interest: (1) electrodes of interest were restricted to those that were located in or near the fusiform gyrus; (2) electrodes were selected such that their peak 6-way classification d' score for faces (see below for how this was calculated) exceeded 0.5 ($p < 0.01$ based on a permutation test, as described below); (3) electrodes were selected such that the peak amplitude of the mean event related potential (ERP) and/or mean event related broadband gamma signal (ERBB) for faces was larger than the peak of mean ERP and/or ERBB for the other non-face object categories in the time window of 0 – 500 ms after stimulus onset.

Multivariate temporal pattern analysis (MTPA)

Multivariate methods were used instead of traditional univariate statistics because of their superior sensitivity^{8,46-48}. In this study, MTPA was applied to decode the coding of stimulus condition in the recorded neural activity. The timecourse of the decoding

accuracy was estimated by classification using a sliding time window of 100 ms. Both ERP and ERBB signals in the time window are combined as input features for the MTPA classifier, which yields 110 temporal features in each case (100 voltage potentials for ERP and 10 normalized mean power-spectrum density for ERBB). The 110 dimensional data were then used as input for the classifier. The goal of the classifier was to learn the patterns of the data distributions in such 110-dimensional space for different conditions and to decode the conditions of the corresponding stimuli from the testing trials. The classifier was trained on each electrode of each subject separately to assess the electrode sensitivity to faces and facial expressions. For Experiment 1, it was a 6-way classification problem and we specifically focused on the sensitivity of face category against other non-face categories. Therefore, we used the sensitivity index (d') for face category against all other non-face category as the metric of face sensitivity. d' was calculated as $Z(\text{true positive rate}) - Z(\text{false positive rate})$ where Z is the inverse of the Gaussian cumulative distribution function. d' was used because it is an unbiased measure of effect size and one that takes into both the true positive and false positive rates. It also has the advantage that it is an effect size measure that has similar interpretation as Cohen's d ^{30,49} while also being applicable to multivariate classification. In addition, we provide full receiver-operator characteristic (ROC) curves for completeness and as validation of d' values. For Experiment 2, averaged pair-wise classification between every possible pair of facial expressions (10 pairs in total) was used.

The choice of the classifier is an empirical problem. The performance of the classifier depends on whether the assumptions of the classifier approximate the underlying truth of the data. Additionally, the complexity of the model and the size of the dataset affect

performance (bias-variance trade-off). In this study, we employed Naïve Bayes (NB) classifiers, which assumes that each of the input features are conditionally independent from one another, and are Gaussian distributed. The classification accuracy of the classifier was estimated through 5-fold cross-validation. Specifically, all the trials were randomly and evenly spited into five folds. In each cross-validation loop, the classifier was trained based on four folds and the performance was evaluated on the left out fold. The overall performance was estimated by averaging cross all the 5 cross-validation loops. In general, different classifiers gave similar results. Specifically, we evaluated the performance of different classifiers (NB, support vector machines, and random forests) on a small subset of the data, and NB classifier tended to perform better than other commonly used classifiers in the current experiment, but other classifiers also gave similar results. In addition, our previous experience⁴⁸ with similar datasets also suggested that NB performed reasonably well in such classification analysis. We therefore used NB throughout the work presented here. The advantage of the Naïve Bayes classifier in the current study is likely due to intrinsic properties of the high dimensional problem⁵⁰ that make a high-bias low-variance classifier (i.e. NB classifier) preferable compared to the low-bias high-variance classifiers (i.e. support vector machines).

Permutation testing

Permutation testing was used to determine the significance of the sensitivity index d' . For each permutation, the condition labels of all the trials were randomly permuted and the same procedure as described above was used to calculate the d' for each permutation.

The permutation was repeated for a total of 1000 times. The d' of each permutation was used as the test statistic and the null distribution of the test statistic was estimated using the histogram of the permutation test.

K-means clustering

K-means clustering was used to cluster the electrodes into groups based on both functional and anatomical features²⁶. Specifically, we applied *k*-means clustering algorithm to the electrodes in a 2D feature space of MNI y-coordinate and the peak classification accuracy time. Note that each dimension was normalized through z-scoring in order to account for different scales in space and time. See Supplemental Information for detailed analysis using Bayesian information criterion and Silhouette analysis for model selection.

Facial feature analysis

The facial features from the stimulus images were extracted following the similar process as⁸. Anatomical landmarks for each picture were first determined by IntraFace²⁹, which marks 49 points on the face along the eyebrows, down the bridge of the nose, along the base of the nose, and outlining the eyes and mouth. Based on these landmarks we calculated 17 facial feature dimensions listed in Table S3. The values for these 17 feature dimensions were normalized by subtracting the mean and dividing by the standard deviation across the all the pictures. The mean representation of each expression in facial feature space was computed by averaging across all 70 faces of the same expression.

Representational similarity analysis (RSA)

RSA was used to analyze the neural representational space for expressions⁵¹. With pair-wise classification accuracy between each pair of facial expressions, we constructed the representational dissimilarity matrix (RDM) of the neural representation of facial expressions, with the element in the *i*-th column of the *j*-th row in the matrix corresponding to the pairwise classification accuracy between the *i*-th and *j*-th facial expressions. The corresponding RDM in the facial feature space was constructed by assessing the Euclidean distance between the *i*-th and *j*-th facial expressions in the 17-dimensional facial feature space.

Acknowledgements

We thank the patients for participating in the iEEG experiments and the UPMC Presbyterian epilepsy monitoring unit staff and administration for their assistance and cooperation with our research. We thank Michael Ward, Ellyanna Kessler, Vincent DeStefino, Shawn Walls, Roma Konecky, Nicolas Brunet, and Witold Lipski for assistance with data collection and thank Ari Kappel and Matthew Boring for assistance with electrode localization. This work was supported by the National Institute on Drug Abuse under award NIH R90DA023420 (to YL), the National Institute of Mental Health under award NIH R01MH107797 and NIH R21MH103592 (to ASG), and the National Science Foundation under award 1734907 (to ASG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Reference:

- 1 Bruce, V. & Young, A. Understanding face recognition. *Br J Psychol* **77** (3), 305-327 (1986).
- 2 Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends Cogn Sci* **4**, 223-233, doi:Doi 10.1016/S1364-6613(00)01482-0 (2000).
- 3 Duchaine, B. & Yovel, G. A Revised Neural Framework for Face Processing. *Annu Rev Vis Sc* **1**, 393-416, doi:10.1146/annurev-vision-082114-035518 (2015).
- 4 Ishai, A. Let's face it: it's a cortical network. *Neuroimage* **40**, 415-419, doi:10.1016/j.neuroimage.2007.10.040 (2008).
- 5 Calder, A. J. & Young, A. W. Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci* **6**, 641-651, doi:10.1038/nrn1724 (2005).
- 6 Nestor, A., Plaut, D. C. & Behrmann, M. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *P Natl Acad Sci USA* **108**, 9998-10003, doi:10.1073/pnas.1102433108 (2011).
- 7 Goesaert, E. & Op de Beeck, H. P. Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *J Neurosci* **33**, 8549-8558, doi:10.1523/JNEUROSCI.1829-12.2013 (2013).
- 8 Ghuman, A. S. *et al.* Dynamic encoding of face information in the human fusiform gyrus. *Nat Commun* **5**, 5672, doi:10.1038/ncomms6672 (2014).
- 9 Barton, J. J., Press, D. Z., Keenan, J. P. & O'Connor, M. Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology* **58**, 71-78 (2002).
- 10 Barton, J. J. Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *J Neuropsychol* **2**, 197-225 (2008).
- 11 Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C. & Kanwisher, N. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* **56**, 2356-2363, doi:10.1016/j.neuroimage.2011.03.067 (2011).
- 12 Hoffman, E. A. & Haxby, J. V. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience* **3**, 80-84, doi:Doi 10.1038/71152 (2000).
- 13 Skerry, A. E. & Saxe, R. A common neural code for perceived and inferred emotion. *J Neurosci* **34**, 15997-16008, doi:10.1523/JNEUROSCI.1676-14.2014 (2014).
- 14 Zhang, H. *et al.* Face-selective regions differ in their ability to classify facial expressions. *Neuroimage* **130**, 77-90, doi:10.1016/j.neuroimage.2016.01.045 (2016).

- 15 Harry, B., Williams, M. A., Davis, C. & Kim, J. Emotional expressions evoke a differential response in the fusiform face area. *Front Hum Neurosci* **7**, 692, doi:10.3389/fnhum.2013.00692 (2013).
- 16 Harris, R. J., Young, A. W. & Andrews, T. J. Brain regions involved in processing facial identity and expression are differentially selective for surface and edge information. *Neuroimage* **97**, 217-223, doi:10.1016/j.neuroimage.2014.04.032 (2014).
- 17 Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends Cogn Sci* **15**, 254-262, doi:10.1016/j.tics.2011.04.003 (2011).
- 18 Kawasaki, H. *et al.* Processing of facial emotion in the human fusiform gyrus. *J Cogn Neurosci* **24**, 1358-1370, doi:10.1162/jocn_a_00175 (2012).
- 19 Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M. A., III & Adolphs, R. Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS One* **3**, e3892, doi:10.1371/journal.pone.0003892 (2008).
- 20 Musch, K. *et al.* Selective attention modulates high-frequency activity in the face-processing network. *Cortex* **60**, 34-51, doi:10.1016/j.cortex.2014.06.006 (2014).
- 21 Pourtois, G., Spinelli, L., Seeck, M. & Vuilleumier, P. Modulation of face processing by emotional expression and gaze direction during intracranial recordings in right fusiform cortex. *J Cogn Neurosci* **22**, 2086-2107, doi:10.1162/jocn.2009.21404 (2010).
- 22 Eimer, M. The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport* **11**, 2319-2324 (2000).
- 23 Eimer, M. in *The Oxford handbook of face perception* (eds Andrew J. Calder, Gillian Rhodes, Mark H. Johnson, & J. V. Haxby) 329-344 (Oxford University Press, 2011).
- 24 Allison, T., Puce, A., Spencer, D. D. & McCarthy, G. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* **9**, 415-430 (1999).
- 25 Kass, R. E. & Wasserman, L. A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion. *J Am Stat Assoc* **90**, 928-934, doi:10.2307/2291327 (1995).
- 26 Kaufman, L. & Rousseeuw, P. J. *Finding groups in data: an introduction to cluster analysis*. (John Wiley & Sons, 2009).
- 27 Weiner, K. S. *et al.* The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* **84**, 453-465, doi:10.1016/j.neuroimage.2013.08.068 (2014).
- 28 Vytal, K. & Hamann, S. Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *J Cogn Neurosci* **22**, 2864-2885, doi:10.1162/jocn.2009.21366 (2010).
- 29 Xiong, X. & De la Torre, F. Supervised descent method and its application to face alignment. *IEEE CVPR* (2013).
- 30 Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd edn, (L. Erlbaum Associates, 1988).

- 31 Engell, A. D. & McCarthy, G. The relationship of gamma oscillations and face-specific ERPs recorded subdurally from occipitotemporal cortex. *Cereb Cortex* **21**, 1213-1221, doi:10.1093/cercor/bhq206 (2011).
- 32 Weiner, K. S. *et al.* The Cytoarchitecture of Domain-specific Regions in Human High-level Visual Cortex. *Cereb Cortex*, doi:10.1093/cercor/bhw361 (2016).
- 33 Freiwald, W. A. & Tsao, D. Y. Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science* **330**, 845-851, doi:10.1126/science.1194908 (2010).
- 34 Eimer, M. Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Res Cogn Brain Res* **10**, 145-158 (2000).
- 35 Bentin, S. & Deouell, L. Y. Structural encoding and identification in face processing: erp evidence for separate mechanisms. *Cogn Neuropsychol* **17**, 35-55, doi:10.1080/026432900380472 (2000).
- 36 Blau, V. C., Maurer, U., Tottenham, N. & McCandliss, B. D. The face-specific N170 component is modulated by emotional facial expression. *Behav Brain Funct* **3**, 7, doi:10.1186/1744-9081-3-7 (2007).
- 37 Furey, M. L. *et al.* Dissociation of face-selective cortical responses by attention. *Proc Natl Acad Sci U S A* **103**, 1065-1070, doi:10.1073/pnas.0510124103 (2006).
- 38 Furl, N., van Rijsbergen, N. J., Treves, A., Friston, K. J. & Dolan, R. J. Experience-dependent coding of facial expression in superior temporal sulcus. *Proc Natl Acad Sci U S A* **104**, 13485-13489, doi:10.1073/pnas.0702548104 (2007).
- 39 Eimer, M., Holmes, A. & McGlone, F. P. The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. *Cogn Affect Behav Neurosci* **3**, 97-110 (2003).
- 40 Eimer, M. Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology* **111**, 694-705 (2000).
- 41 Vida, M. D., Nestor, A., Plaut, D. C. & Behrmann, M. Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proc Natl Acad Sci U S A* **114**, 388-393, doi:10.1073/pnas.1614763114 (2017).
- 42 Li, Y., Richardson, R. M. & Ghuman, A. S. Multi-Connection Pattern Analysis: Decoding the representational content of neural communication. *Neuroimage* **162**, 32-44, doi:10.1016/j.neuroimage.2017.08.033 (2017).
- 43 Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* **2011**, 156869, doi:10.1155/2011/156869 (2011).
- 44 Hermes, D., Miller, K. J., Noordmans, H. J., Vansteensel, M. J. & Ramsey, N. F. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods* **185**, 293-298, doi:10.1016/j.jneumeth.2009.10.005 (2010).
- 45 Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci* **2011**, 879716, doi:10.1155/2011/879716 (2011).

- 46 Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annu Rev Neurosci* **37**, 435-456, doi:10.1146/annurev-neuro-062012-170325 (2014).
- 47 Miller, K. J., Schalk, G., Hermes, D., Ojemann, J. G. & Rao, R. P. Spontaneous Decoding of the Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals Complementary Information in the Event-Related Potential and Broadband Spectral Change. *PLoS Comput Biol* **12**, e1004660, doi:10.1371/journal.pcbi.1004660 (2016).
- 48 Hirshorn, E. A. *et al.* Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc Natl Acad Sci U S A* **113**, 8162-8167, doi:10.1073/pnas.1604126113 (2016).
- 49 Sawilowsky, S. S. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* **8**, 597-599 (2009).
- 50 Bickel, P. J. & Levina, E. Some theory for Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. *Bernoulli* **10**, 989-1010 (2004).
- 51 Kriegeskorte, N. & Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* **17**, 401-412, doi:10.1016/j.tics.2013.06.007 (2013).

Supplement Information

Methods

Clustering analysis

We applied k -means clustering to the electrodes in the 2D space of MNI y -coordinate and peak classification time for facial expressions, with different values of k , and evaluated the model performance by computing the Bayes information criterion (BIC) and the mean Silhouette coefficient (SC) across all points.

Following ^{1,2}, the BIC was estimated using Schwartz criterion. Specifically, $BIC = l(D|\hat{\theta}) - \frac{p}{2} \log N$, where $l(D|\hat{\theta})$ is the log-likelihood of the data under the assumption of k -means (spherical Gaussian) taken at the maximum likelihood estimation of parameters $\hat{\theta}$, p is the total number of parameters in the model, and N is the total number of data points.

Following ³, the Silhouette value for the i -th point was computed as $S_i = (b_i - a_i) / \max(a_i, b_i)$, where a_i is the average within cluster distance for the i -th point, and b_i is the minimum average between cluster distance for the i -th point (minimized over all other clusters). The mean SC was then estimated by averaging the Silhouette value over all data points.

Meta-analysis

Activation likelihood estimation (ALE, ^{4,5}) was used for the meta-analysis of the neuroimaging literature. We first searched the online database of neuroimaging studies on Neurosynth.org and found around 300 imaging studies with the keyword “facial expressions”. We then further narrowed the list down to 64 fMRI by only including the

studies that had a direct full brain mapping by contrasting between emotional facial expressions, e.g. fear vs neutral, happy vs sad, etc. We only took into account the reported activation foci for the contrast between facial expressions. Then all of the activation foci in those relevant full brain map results were collected and extracted as 3D coordinates in MNI space. In the ALE, each of the extracted foci was assigned as the center of a Gaussian distribution, whose variance was scaled by the number of subjects in the corresponding experiment. These Gaussian distributions were then combined to build a full brain map of ALE. The ALE map was corrected for multiple comparison using cluster-based permutation test. Then we performed a spatial permutation test with 1000 permutations to construct a null distribution of the full brain activation. The ALE and the corresponding statistical analysis were performed based on GingerALE 2.3.6^{6,7}.

Results

Selection of models for *k*-means clustering

We applied *k*-means clustering to the electrodes in the 2D space of MNI y-coordinate and peak classification time for facial expressions, with different values of *k*, and evaluate the model performance by computing the Bayes information criterion (BIC) and the mean Silhouette coefficient (SC) across all points.

As shown in Figure S3, for *k* = 1, BIC = -61.28; for *k* = 2, BIC = -54.63. Therefore, Bayes factor between the hypothesis (H1) that there is a cluster structure (*k* = 2) and the null hypothesis (H0) that there is no cluster structure (*k* = 1) can be approximated as $BF \approx \exp(\frac{BIC_1 - BIC_0}{2})$. This approximation yields a $BF > 20$, which

suggests a strong evidence of H1 over H0. In other words, there is a strong clustering structure in the data.

Moreover, for $k = 2$, BIC = -54.63, the mean SC = 0.601; for $k = 3$, BIC = -56.29, the mean SC = 0.490; for $k = 4$, BIC = -58.54, mean SC = 0.428. Both BIC and mean SC suggest that $k = 2$ is the optimal number of clusters. Therefore, $k = 2$ was used in the study.

Meta-analysis of the neuroimaging literature

In the broad neuroimaging literature, we found 64 fMRI studies with full brain contrasts between face expressions (See Table S1). Among the 64 studies, 24 studies report at least one significant focus of fusiform sensitivity to differences in expressions (See Figure S4 for activation map). A total of 999 significant foci were reported in those experiments for contrasts between different facial expressions (Figure S4). A full brain activation likelihood estimation (ALE) was performed and significance was assessed using a cluster-based permutation test. 4 significant clusters were found at the $p < 0.01$ threshold, none of which included the fusiform. The MNI coordinates for the center and the corresponding label names of the 4 clusters are shown in Table S2.

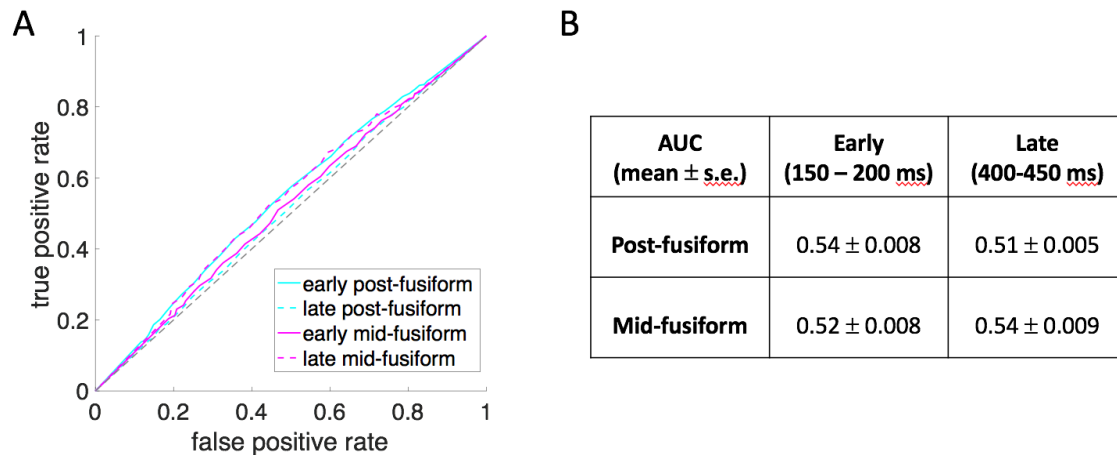


Figure S1. The mean ROC curve and area-under-curve (AUC) for posterior fusiform electrodes and mid-fusiform electrodes at early (150-200 ms after stim onset) and late stage (400-450 ms after stim onset).

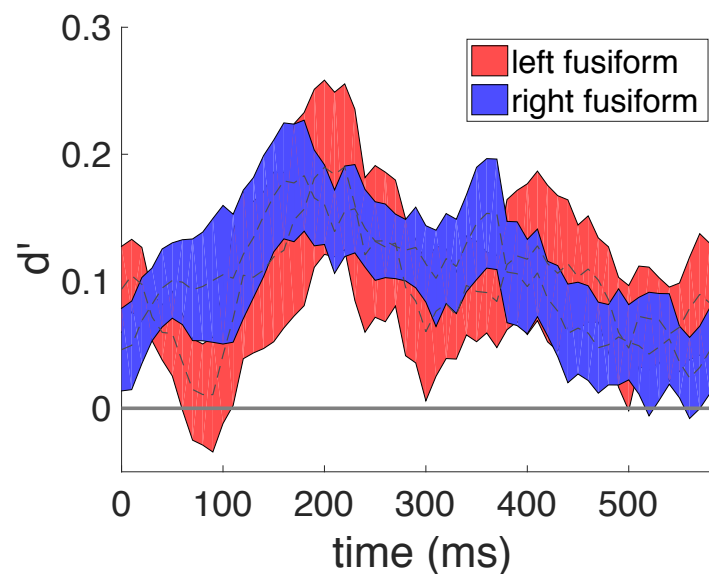


Figure S2. The mean and standard error for classification between different face expressions in left and right fusiform electrodes. The timecourse of the left fusiform peaked at 220 ms after stimulus onset with mean $d' = 0.19$, and the timecourse of the right fusiform peaked at 180 ms after stimulus onset with mean $d' = 0.18$ (both $p < 0.05$, binomial test, Bonferroni corrected).

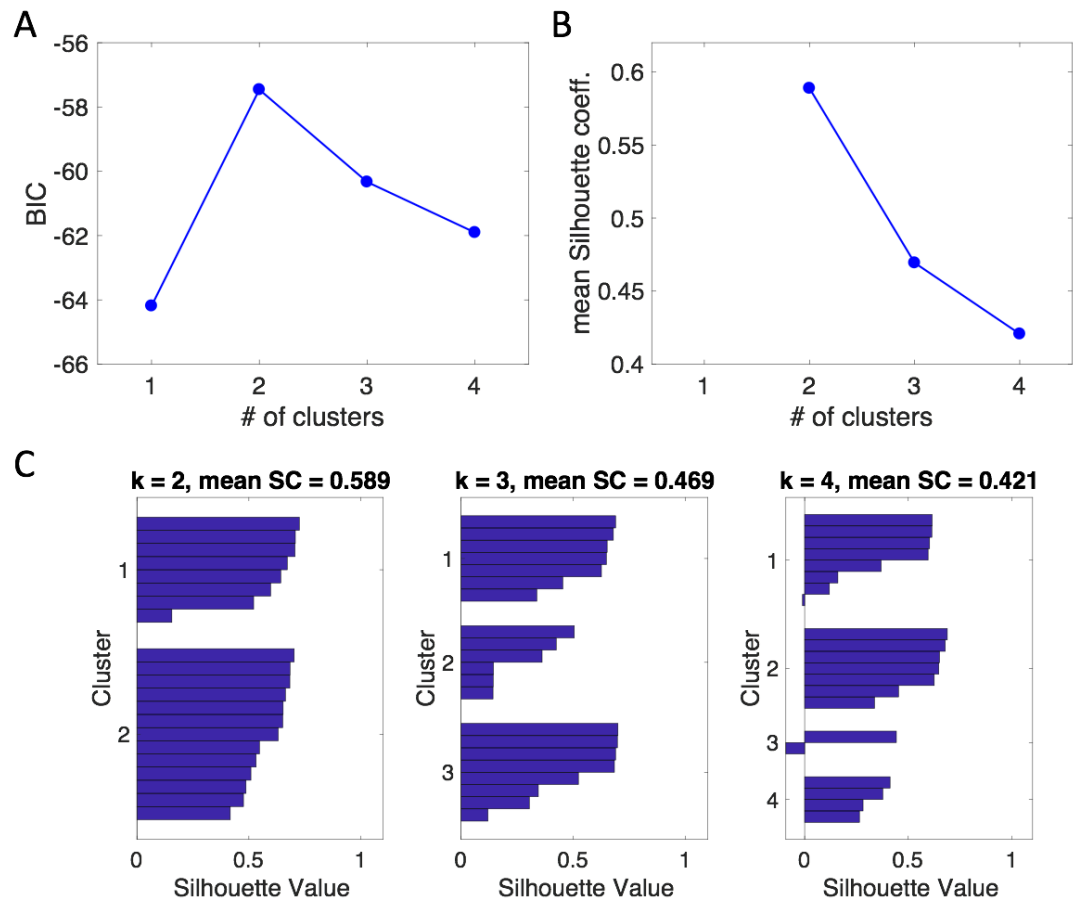


Figure S3. Clustering analysis. **A)** BIC of k -means models with different values of k ($k = 1, 2, 3, 4$). **B)** Mean SC of k -means models with different values of k ($k = 2, 3, 4$, note that SC is not applicable for $k = 1$). **C)** The distribution of Silhouette Coefficients (SC) with different values of k in k -means clustering. From left to right, $k = 2, 3$, and 4.

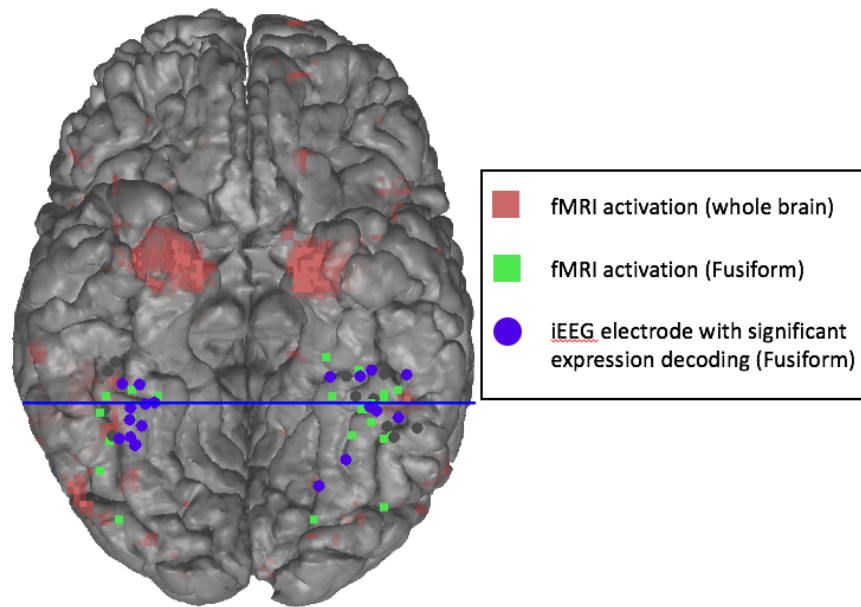


Figure S4. Activation map for facial expressions. (Red) Whole brain activation map from all 64 relevant fMRI studies. (Green square) Voxels in fusiform reported in 24/64 of the fMRI studies that have significant contrast between facial expressions. (Blue dots) iEEG electrodes in fusiform that have significant facial expression decoding. (Blue line) the border between posterior and mid-fusiform clusters based upon clustering analysis in the iEEG electrodes.

Table S1 A summary list for the 64 neuroimaging studies included in the meta-analysis

| | title | authors | journal | year |
|---|---|---|--|------|
| 1 | A common neural code for perceived and inferred emotion. | Skerry AE, Saxe R | The Journal of neuroscience : the official journal of the Society for Neuroscience | 2014 |
| 2 | A left amygdala mediated network for rapid orienting to masked fearful faces. | Carlson JM, Reinke KS, Habib R | Neuropsychologia | 2009 |
| 3 | A neural network reflecting individual differences in cognitive processing of emotions during perceptual decision making. | Meriau K, Wartenburger I, Kazzner P, Prehn K, Lammers CH, van der Meer E, Villringer A, Heekeren HR | NeuroImage | 2006 |
| 4 | Affect-specific activation of shared networks for perception and execution of facial expressions. | Kircher T, Pohl A, Krach S, Thimm M, Schulte-Ruther M, Anders S, Mathiak K | Social cognitive and affective neuroscience | 2013 |

| | | | | |
|----|--|---|--|------|
| 5 | Amygdala activation at 3T in response to human and avatar facial expressions of emotions. | Moser E, Derntl B, Robinson S, Fink B, Gur RC, Grammer K | Journal of neuroscience methods | 2007 |
| 6 | Amygdala integrates emotional expression and gaze direction in response to dynamic facial expressions. | Sato W, Kochiyama T, Uono S, Yoshikawa S | NeuroImage | 2010 |
| 7 | Amygdala reactivity predicts automatic negative evaluations for facial emotions. | Dannlowski U, Ohrmann P, Bauer J, Kugel H, Arolt V, Heindel W, Suslow T | Psychiatry research | 2007 |
| 8 | Amygdala response to facial expressions in children and adults. | Thomas KM, Drevets WC, Whalen PJ, Eccard CH, Dahl RE, Ryan ND, Casey BJ | Biological psychiatry | 2001 |
| 9 | Amygdala response to facial expressions reflects emotional learning. | Hooker CI, Germine LT, Knight RT, D'Esposito M | The Journal of neuroscience : the official journal of the Society for Neuroscience | 2006 |
| 10 | Anxiety predicts a differential neural response to attended and unattended facial signals of anger and fear. | Ewbank MP, Lawrence AD, Passamonti L, Keane J, Peers PV, Calder AJ | NeuroImage | 2009 |
| 11 | Automatic emotion processing as a function of trait emotional awareness: an fMRI study. | Lichev V, Sacher J, Ihme K, Rosenberg N, Quirin M, Lepsien J, Pampel A, Rufer M, Grabe HJ, Kugel H, Kersting A, Villringer A, Lane RD, Suslow T | Social cognitive and affective neuroscience | 2014 |
| 12 | Beyond threat: amygdala reactivity across multiple expressions of facial affect. | Fitzgerald DA, Angstadt M, Jelsone LM, Nathan PJ, Phan KL | NeuroImage | 2006 |
| 13 | Binding action and emotion in social understanding. | Ferri F, Ebisch SJ, Costantini M, Salone A, Arciero G, Mazzola V, Ferro FM, Romani GL, Gallese V | PloS one | 2013 |
| 14 | Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. | Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G | Neuron | 2003 |
| 15 | Brain networks involved in haptic and visual identification of facial expressions of emotion: an fMRI study. | Kitada R, Johnsrude IS, Kochiyama T, Lederman SJ | NeuroImage | 2010 |
| 16 | Brain responses to dynamic facial expressions of pain. | Simon D, Craig KD, Miltner WH, Rainville P | Pain | 2006 |
| 17 | Brain responses to facial expressions of pain: emotional or | Budell L, Jackson P, Rainville P | NeuroImage | 2010 |

| | | | | |
|----|---|--|--|------|
| | motor mirroring? | | | |
| 18 | Cerebral integration of verbal and nonverbal emotional cues: impact of individual nonverbal dominance. | Jacob H, Kreifelts B, Bruck C, Erb M, Hosl F, Wildgruber D | NeuroImage | 2012 |
| 19 | Cerebral regulation of facial expressions of pain. | Kunz M, Chen JI, Lautenbacher S, Vachon-Preseau E, Rainville P | The Journal of neuroscience : the official journal of the Society for Neuroscience | 2011 |
| 20 | Classification images reveal the information sensitivity of brain voxels in fMRI. | Smith FW, Muckli L, Brennan D, Pernet C, Smith ML, Belin P, Gosselin F, Hadley DM, Cavanagh J, Schyns PG | NeuroImage | 2008 |
| 21 | Converging evidence for the advantage of dynamic facial expressions. | Arsalidou M, Morris D, Taylor MJ | Brain topography | 2011 |
| 22 | Decoding of affective facial expressions in the context of emotional situations. | Sommer M, Dohnel K, Meinhardt J, Hajak G | Neuropsychologia | 2008 |
| 23 | Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. | Foley E, Rippon G, Thai NJ, Longe O, Senior C | Journal of cognitive neuroscience | 2012 |
| 24 | Dynamic stimuli demonstrate a categorical representation of facial expression in the amygdala. | Harris RJ, Young AW, Andrews TJ | Neuropsychologia | 2014 |
| 25 | Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. | Trautmann SA, Fehr T, Herrmann M | Brain research | 2009 |
| 26 | Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. | Sato W, Kochiyama T, Yoshikawa S, Naito E, Matsumura M | Brain research. Cognitive brain research | 2004 |
| 27 | Facial emotion modulates the neural mechanisms responsible for short interval time perception. | Tipples J, Brattan V, Johnston P | Brain topography | 2015 |
| 28 | Facial expression and gaze-direction in human superior temporal sulcus. | Engell AD, Haxby JV | Neuropsychologia | 2007 |
| 29 | Facial expressions and complex IAPS pictures: common and | Britton JC, Taylor SF, Sudheimer KD, Liberzon I | NeuroImage | 2006 |

| | | | | |
|----|--|--|--|------|
| | differential networks. | | | |
| 30 | Frontal lobe networks for effective processing of ambiguously expressed emotions in humans. | Nomura M, Iidaka T, Kakehi K, Tsukiura T, Hasegawa T, Maeda Y, Matsue Y | Neuroscience letters | 2003 |
| 31 | Functional imaging of face and hand imitation: towards a motor theory of empathy. | Leslie KR, Johnson-Frey SH, Grafton ST | NeuroImage | 2004 |
| 32 | Functional neuroanatomy of perceiving surprised faces. | Schroeder U, Hennenlotter A, Erhard P, Haslinger B, Stahl R, Lange KW, Ceballos-Baumann AO | Human brain mapping | 2004 |
| 33 | Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. | Ethofer T, Breitscher J, Wiethoff S, Bisch J, Schlipf S, Wildgruber D, Kreifelts B | NeuroImage | 2013 |
| 34 | Incongruence effects in crossmodal emotional integration. | Muller VI, Habel U, Derntl B, Schneider F, Zilles K, Turetsky BI, Eickhoff SB | NeuroImage | 2011 |
| 35 | Integration of cross-modal emotional information in the human brain: an fMRI study. | Park JY, Gu BM, Kang DH, Shin YW, Choi CH, Lee JM, Kwon JS | Cortex; a journal devoted to the study of the nervous system and behavior | 2010 |
| 36 | Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. | Wegrzyn M, Riehle M, Labudda K, Woermann F, Baumgartner F, Pollmann S, Bien CG, Kissler J | Cortex; a journal devoted to the study of the nervous system and behavior | 2015 |
| 37 | Is a neutral expression also a neutral stimulus? A study with functional magnetic resonance. | Carvajal F, Rubio S, Serrano JM, Rios-Lago M, Alvarez-Linera J, Pacheco L, Martin P | Experimental brain research | 2013 |
| 38 | Leaving a bad taste in your mouth but not in my insula. | von dem Hagen EA, Beaver JD, Ewbank MP, Keane J, Passamonti L, Lawrence AD, Calder AJ | Social cognitive and affective neuroscience | 2009 |
| 39 | Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. | Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA | The Journal of neuroscience : the official journal of the Society for Neuroscience | 1998 |
| 40 | Mind your left: spatial bias in subcortical fear processing. | Siman-Tov T, Papo D, Gadoth N, Schonberg T, Mendelsohn A, Perry D, Hendler T | Journal of cognitive neuroscience | 2009 |
| 41 | Multiple mechanisms of consciousness: the neural correlates of emotional | Amting JM, Greening SG, Mitchell DG | The Journal of neuroscience : the official journal of the | 2010 |

| | | | | |
|----|---|---|---|------|
| | awareness. | | Society for Neuroscience | |
| 42 | Neural mechanism for judging the appropriateness of facial affect. | Kim JW, Kim JJ, Jeong BS, Ki SW, Im DM, Lee SJ, Lee HS | Brain research. Cognitive brain research | 2005 |
| 43 | Neural mechanism of unconscious perception of surprised facial expression. | Duan X, Dai Q, Gong Q, Chen H | NeuroImage | 2010 |
| 44 | Neural responses to ambiguity involve domain-general and domain-specific emotion processing systems. | Neta M, Kelley WM, Whalen PJ | Journal of cognitive neuroscience | 2013 |
| 45 | Nonconscious emotional processing involves distinct neural pathways for pictures and videos. | Faivre N, Charron S, Roux P, Lehericy S, Kouider S | Neuropsychologia | 2012 |
| 46 | Orbitofrontal and hippocampal contributions to memory for face-name associations: the rewarding power of a smile. | Tsukiura T, Cabeza R | Neuropsychologia | 2008 |
| 47 | Orbitofrontal Cortex Reactivity to Angry Facial Expression in a Social Interaction Correlates with Aggressive Behavior. | Beyer F, Munte TF, Gottlich M, Kramer UM | Cerebral cortex (New York, N.Y. : 1991) | 2014 |
| 48 | Positive facial affect - an fMRI study on the involvement of insula and amygdala. | Pohl A, Anders S, Schulte-Ruther M, Mathiak K, Kircher T | PloS one | 2013 |
| 49 | Preferential amygdala reactivity to the negative assessment of neutral faces. | Blasi G, Hariri AR, Alce G, Taurisano P, Sambataro F, Das S, Bertolino A, Weinberger DR, Mattay VS | Biological psychiatry | 2009 |
| 50 | Pupillary contagion: central mechanisms engaged in sadness processing. | Harrison NA, Singer T, Rotshtein P, Dolan RJ, Critchley HD | Social cognitive and affective neuroscience | 2006 |
| 51 | Reduced emotion processing efficiency in healthy males relative to females. | Weisenbach SL, Rapport LJ, Briceno EM, Haase BD, Vederman AC, Bieliauskas LA, Welsh RC, Starkman MN, McInnis MG, Zubieta JK, Langenecker SA | Social cognitive and affective neuroscience | 2014 |
| 52 | Segregating intra-amygdalar responses to dynamic facial emotion with cytoarchitectonic maximum probability maps. | Hurlemann R, Rehme AK, Diessel M, Kukolja J, Maier W, Walter H, Cohen MX | Journal of neuroscience methods | 2008 |
| 53 | Similarities and differences in perceiving threat from dynamic | Kret ME, Pichon S, Grezes J, de Gelder B | NeuroImage | 2011 |

| | | | | |
|----|---|--|---|------|
| | faces and bodies. An fMRI study. | | | |
| 54 | Stop looking angry and smile, please: start and stop of the very same facial expression differentially activate threat- and reward-related brain networks. | Muhlberger A, Wieser MJ, Gerdes AB, Frey MC, Weyers P, Pauli P | Social cognitive and affective neuroscience | 2011 |
| 55 | Temporal pole activity during perception of sad faces, but not happy faces, correlates with neuroticism trait. | Jimura K, Konishi S, Miyashita Y | Neuroscience letters | 2009 |
| 56 | The amygdala and FFA track both social and non-social face dimensions. | Said CP, Dotsch R, Todorov A | Neuropsychologia | 2010 |
| 57 | The amygdala processes the emotional significance of facial expressions: an fMRI investigation using the interaction between expression and face direction. | Sato W, Yoshikawa S, Kochiyama T, Matsumura M | NeuroImage | 2004 |
| 58 | The behavioral and neural effect of emotional primes on intertemporal decisions. | Luo S, Ainslie G, Monterosso J | Social cognitive and affective neuroscience | 2014 |
| 59 | The changing face of emotion: age-related patterns of amygdala activation to salient faces. | Todd RM, Evans JW, Morris D, Lewis MD, Taylor MJ | Social cognitive and affective neuroscience | 2011 |
| 60 | The functional correlates of face perception and recognition of emotional facial expressions as evidenced by fMRI. | Jehna M, Neuper C, Ischebeck A, Loitfelder M, Ropele S, Langkammer C, Ebner F, Fuchs S, Schmidt R, Fazekas F, Enzinger C | Brain research | 2011 |
| 61 | The highly sensitive brain: an fMRI study of sensory processing sensitivity and response to others' emotions. | Acevedo BP, Aron EN, Aron A, Sangster MD, Collins N, Brown LL | Brain and behavior | 2014 |
| 62 | The Kuleshov Effect: the influence of contextual framing on emotional attributions. | Mobbs D, Weiskopf N, Lau HC, Featherstone E, Dolan RJ, Frith CD | Social cognitive and affective neuroscience | 2006 |
| 63 | The stimuli drive the response: an fMRI study of youth processing adult or child emotional face stimuli. | Marusak HA, Carre JM, Thomason ME | NeuroImage | 2013 |
| 64 | Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. | Botvinick M, Jha AP, Bylsma LM, Fabian SA, Solomon PE, Prkachin KM | NeuroImage | 2005 |

Table S2 the MNI coordinates for the weighted center, volume, and the corresponding label name of the significant clusters in the ALE map from meta-analysis

| Cluster # | X | Y | Z | Volume (mm ³) | Lateralization | Label |
|-----------|-----|-----|-----|---------------------------|----------------|--------------------------------|
| 1 | 23 | -3 | -18 | 6088 | right | amygdala |
| 2 | -22 | -4 | -18 | 4640 | left | amygdala |
| 3 | 56 | -42 | 5 | 2520 | right | middle/superior temporal gyrus |
| 4 | 3 | 11 | 53 | 1464 | right | superior frontal gyrus |

Table S3 17 features used for the facial feature space

| Feature # | Feature name |
|-----------|--|
| 1 | eyebrow length |
| 2 | inter-eyebrow distance |
| 3 | eye width |
| 4 | inter-eyes distance |
| 5 | vertical distance between eyes and nosetip |
| 6 | horizontal length of the nose |
| 7 | distance between nose and upper lip |
| 8 | face height |
| 9 | face width |
| 10 | eye height |
| 11 | width of the mouth |
| 12 | intense of red on cheeks |
| 13 | intense of green on cheeks |
| 14 | intense of blue on cheeks |
| 15 | contrast polarity between eyes and nose |
| 16 | eye area |
| 17 | eye mouth ratio |

Reference:

- 1 Kass, R. E. & Wasserman, L. A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion. *J Am Stat Assoc* **90**, 928-934, doi:10.2307/2291327 (1995).
- 2 Pelleg, D. & Moore, A. W. in *International Conference on Machine Learning*.
- 3 Kaufman, L. & Rousseeuw, P. J. *Finding groups in data: an introduction to cluster analysis*. (John Wiley & Sons, 2009).
- 4 Laird, A. R. *et al.* ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Hum Brain Mapp* **25**, 155-164, doi:10.1002/hbm.20136 (2005).
- 5 Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F. & Fox, P. T. Activation likelihood estimation meta-analysis revisited. *Neuroimage* **59**, 2349-2361, doi:10.1016/j.neuroimage.2011.09.017 (2012).
- 6 Eickhoff, S. B. *et al.* Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum Brain Mapp* **30**, 2907-2926, doi:10.1002/hbm.20718 (2009).
- 7 Turkeltaub, P. E. *et al.* Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. *Hum Brain Mapp* **33**, 1-13, doi:10.1002/hbm.21186 (2012).