



Cross-Subject Commonality of Emotion Representations in Dorsal Motion-Sensitive Areas

Yin Liang^{1*} and Baolin Liu^{2*}

¹ Faculty of Information Technology, College of Computer Science and Technology, Beijing Artificial Intelligence Institute, Beijing University of Technology, Beijing, China, ² School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China

OPEN ACCESS

Edited by:

Yu Zhang,
Stanford University, United States

Reviewed by:

Kaundinya S. Gopinath,
Emory University, United States
Zhuo Fang,
University of Ottawa, Canada

*Correspondence:

Yin Liang
yinliang@bjut.edu.cn
Baolin Liu
liubaolin@tsinghua.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 30 May 2020

Accepted: 22 September 2020

Published: 14 October 2020

Citation:

Liang Y and Liu B (2020)
Cross-Subject Commonality
of Emotion Representations in Dorsal
Motion-Sensitive Areas.
Front. Neurosci. 14:567797.
doi: 10.3389/fnins.2020.567797

Emotion perception is a crucial question in cognitive neuroscience and the underlying neural substrates have been the subject of intense study. One of our previous studies demonstrated that motion-sensitive areas are involved in the perception of facial expressions. However, it remains unclear whether emotions perceived from whole-person stimuli can be decoded from the motion-sensitive areas. In addition, if emotions are represented in the motion-sensitive areas, we may further ask whether the representations of emotions in the motion-sensitive areas can be shared across individual subjects. To address these questions, this study collected neural images while participants viewed emotions (joy, anger, and fear) from videos of whole-person expressions (contained both face and body parts) in a block-design functional magnetic resonance imaging (fMRI) experiment. Multivariate pattern analysis (MVPA) was conducted to explore the emotion decoding performance in individual-defined dorsal motion-sensitive regions of interest (ROIs). Results revealed that emotions could be successfully decoded from motion-sensitive ROIs with statistically significant classification accuracies for three emotions as well as positive versus negative emotions. Moreover, results from the cross-subject classification analysis showed that a person's emotion representation could be robustly predicted by others' emotion representations in motion-sensitive areas. Together, these results reveal that emotions are represented in dorsal motion-sensitive areas and that the representation of emotions is consistent across subjects. Our findings provide new evidence of the involvement of motion-sensitive areas in the emotion decoding, and further suggest that there exists a common emotion code in the motion-sensitive areas across individual subjects.

Keywords: functional magnetic resonance imaging, emotion perception, multivariate pattern analysis, motion-sensitive areas, cross-subject decoding

INTRODUCTION

The ability to understand emotions is a crucial social skill in humans. It has been proposed that body language plays an important role in conveying emotions (Calbi et al., 2017). Body language refers to the non-verbal signals in which physical behaviors, including facial expressions, body posture, gestures, eye movement, touch and the use of space, are used to express our true

feelings and emotions. According to experts, these non-verbal signals make up a huge part of our daily communication. Humans can easily recognize others' emotions from their whole-person expressions and perceive them in a categorical manner. Since the human brain can readily decode emotions, considerable functional magnetic resonance imaging (fMRI) studies have investigated the potential neural substrates and mechanisms underlying the perception of emotions.

Neuroimaging studies on emotion perception have used emotional faces or non-face bodies as stimuli and identified specific areas showing preferential activation patterns, respectively known as face-selective and body-selective areas. Classical face-selective areas mainly contain the fusiform face area (FFA), occipital face area (OFA), and superior temporal sulcus (STS), which are together considered the "core system" in Haxby's model (Haxby et al., 2000; Kanwisher and Yovel, 2006; Gobbini and Haxby, 2007; Pitcher, 2014; Henriksson et al., 2015). Emotional bodies are found to be represented in the extrastriate body area (EBA) and fusiform body area (FBA), and some similarities have been revealed between the processing of emotional bodies and faces (Minnebusch and Daum, 2009; de Gelder et al., 2010; Kret et al., 2011; Downing and Peelen, 2016). In addition, the STS, which acts as a crucial node for social information processing, has been found to be involved in the processing of emotions in both faces and bodies (Candidi et al., 2011; Zhu et al., 2013). Previous fMRI studies mainly assessed the perception of emotions using either isolated faces or non-face bodies as visual stimuli. However, behavioral studies have indicated that human brain prefers whole-person expressions which contain both the face and body parts, similar to that which we commonly perceive in real scenes, and encoding whole-person expressions in a holistic rather than part-based manner (Soria Bauser and Suchan, 2015). Therefore, it is essential to explore the neural representation of whole-person expressions individually rather than in an integrated manner based on the isolated emotional faces and bodies (Zhang et al., 2012; Soria Bauser and Suchan, 2015). Moreover, most previous studies used static emotional images as stimuli, but, considering that the emotions we mostly encounter in a natural context are dynamic, recent studies have proposed that dynamic stimuli are more ecologically valid than their static counterparts (Johnston et al., 2013; Yang et al., 2018). Thus, using dynamic emotional stimuli may be more appropriate to investigate the authentic mechanisms used to recognize emotions in daily life.

Compared to univariate analyses that estimate emotion-evoked responses, a multivariate pattern analysis (MVPA), as demonstrated by recent fMRI studies, can take advantage of distributed activation patterns in fMRI data, thus providing a more effective method to infer the functional roles of cortical regions in emotion perception (Mahmoudi et al., 2012). A growing number of studies have used ROI-based MVPA to explore emotion decoding performances in specific brain areas (Said et al., 2010; Harry et al., 2013; Wegrzyn et al., 2015). In addition, studies with dynamic stimuli have found that dorsal motion-sensitive areas within human motion complex (hMT) + /V5 and STS exhibited significant responses

to facial expressions (Furl et al., 2013, 2015). A macaque study identified motion-sensitive areas in the STS, which may be homologous to human STS, and found that facial expressions could be successfully decoded from motion-sensitive areas (Furl et al., 2012). Moreover, one of our recent studies has also identified the successful decoding of dynamic facial expressions in motion-sensitive areas (Liang et al., 2017). These findings suggest that motion-sensitive areas may transmit measurable quantities of expression information and may play an important role in emotion perception. However, these studies only used facial expressions as stimuli, and the full role of motion-sensitive areas in the decoding of whole-person expressions therefore remains unclear. Since we commonly perceive emotions from whole-person expressions in our daily lives, exploring the decoding performance of whole-person expressions in motion-sensitive areas may be meaningful in revealing the potential mechanisms by which the human brain efficiently recognizes emotions from body movements. Furthermore, if emotions are represented in the motion-sensitive areas, we may further ask whether emotion codes in the motion-sensitive areas can be shared across individual subjects. This would shed light on whether an individual's subjective emotion representation in motion-sensitive areas corresponds to those observed in others, which would be helpful in assessing the commonality and variability of emotion coding.

In this study, we conducted a regions of interest (ROI) MVPA to assess the potential role of dorsal motion-sensitive areas in emotion decoding. We performed a block-design fMRI experiment and collected neural images while participants viewed emotional videos expressed by whole-person expressions (joy, anger, and fear). Dynamic emotion stimuli were used in this study to enhance ecological validity and to assess the authentic mechanisms of emotion recognition in daily life. A separate localizer was used to identify individual-defined motion-sensitive ROIs. We first examined whether emotions could be decoded based on the activation patterns from motion-sensitive ROIs, after which we examined whether there exists a common representation of emotions in motion-sensitive areas across individuals.

MATERIALS AND METHODS

Participants

A total of 24 healthy, righted-handed college students participated in the experiment (12 males, ranging from 19–25-years-old). All subjects had normal or corrected-to-normal vision, with no history of neurological disorders, and signed informed written consent forms before the experiment. Experimental procedures were explained to them before the scanning. The threshold for head motion was framewise displacement (FD) < 0.5 mm (Power et al., 2012). Four subjects were discarded due to excessive head motion, and the final fMRI analysis was focused on the data of 20 subjects (10 males, mean age 21.8 ± 1.83 years old). This experiment was approved by the local Ethics Committee of Yantai Affiliated Hospital of Binzhou

Medical University. A separate group of subjects ($n = 18$, 8 females, mean age: 22.2 years old) participated in a preliminary behavioral experiment for the stimulus validation.

Experimental Procedures

The fMRI experiment was based on a block design, with four “main experiment” runs for the emotion perception task and one “localizer” run for the ROI identification. A separate localizer was used in our study to ensure that the data used for the ROI definition was independent of the data used for the classification in the main experiment analysis (Axelrod and Yovel, 2012, 2015; Furl et al., 2013; Harry et al., 2013).

Figure 1A shows the process of the main experiment. Each run began with a 10 s fixation cross followed by 18 stimulus blocks presented in a pseudo-random order (Axelrod and Yovel, 2012; Furl et al., 2013, 2015). Successive stimulus blocks were separated by 10 s intervals of a fixation cross. For the first three runs, three emotions (joy, anger, and fear) expressed by three stimulus types (facial, non-face bodily, and whole-person stimuli) were presented in different blocks, while for the fourth run, only three emotions expressed by the whole-person stimuli were presented. In each block, eight video clips of different examples per emotion category were displayed (each for 2000 ms), with an interstimulus interval (ISI) of 500 ms. At the end of each block, there was a 2 s button task instructing participants to indicate the emotion category they had seen by pressing a button. The emotion stimuli were taken from the Geneva Multimodal Emotion Portrayals (GEMEP) corpus (Banziger et al., 2012). Videos of eight individuals (four males and four females) displaying three emotions (joy, anger, and fear) were selected as whole-person emotion stimuli (Cao et al., 2018; Yang et al., 2018). Facial and bodily emotion stimuli were generated from the whole-person videos by cutting out and obscuring the irrelevant part with Gaussian blur masks using Adobe Premiere Pro CC (Kret et al., 2011). All video clips were cropped to 2,000 ms (25 frames/s) to retain the transition from a neutral expression to the emotion apex, and were converted into grayscale using MATLAB (Furl et al., 2012, 2013, 2015; Kaiser et al., 2014; Soria Bauser and Suchan, 2015). The resulting videos were resized to 720×576 pixels and presented on the center of the screen. All generated emotion stimuli were validated by another group of participants before scanning, confirming the validity of the stimuli in representing all expressions. **Figure 1B** shows the examples of whole-person emotion stimuli in the main experiment.

In the functional localizer run, participants viewed video clips or static images of four categories: faces, non-face bodies, whole-persons and objects. Each category appeared two times in a pseudo-random order, resulting in 16 blocks in total (4 categories \times video/image \times 2 repetitions). Each block contained 8 stimuli (7 novel and 1 repeated), and each was presented for 1,400ms, separated by an ISI of 100 ms. Participants performed a “one-back” task during the localizer run, that is, to press a button when they observed two identical stimuli appearing in consecutive trials.

The stimuli were presented using E-Prime 2.0 Professional (Psychology Software Tools, Pittsburgh, PA, United States) and

the behavioral results were collected using the response pad in the scanner. After scanning, participants were required to complete a questionnaire recording whether participants performed the experiment according to the instructions, their feelings during the fMRI experiment, and any difficulties they encountered.

Data Acquisition

Imaging data were acquired from Yantai Affiliated Hospital of Binzhou Medical University, using a 3.0-T SIEMENS MRI scanner with an eight-channel head coil. Acquisition parameters of task-related functional images and anatomical images were as follows: T2*-weighted functional images were collected using a gradient echo-planar imaging (EPI) sequence, with repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, voxel size = $3.1 \text{ mm} \times 3.1 \text{ mm} \times 4.0 \text{ mm}$, matrix size = 64×64 , slices = 33, slices thickness = 4 mm, slice gap = 0.6 mm (Yang et al., 2018). T1-weighted anatomical images were acquired using a three-dimensional magnetization-prepared rapid-acquisition gradient echo (3D MPRAGE) sequence, with TR = 1,900 ms, TE = 2.52 ms, time of inversion (TI) = 1100 ms, voxel size = $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$, matrix size = 256×256 . Participants viewed the emotion stimuli through the high-resolution stereo 3D glasses of the VisuaStim Digital MRI Compatible fMRI system. Foam pads and earplugs were used during scanning to reduce head motion and scanner noise.

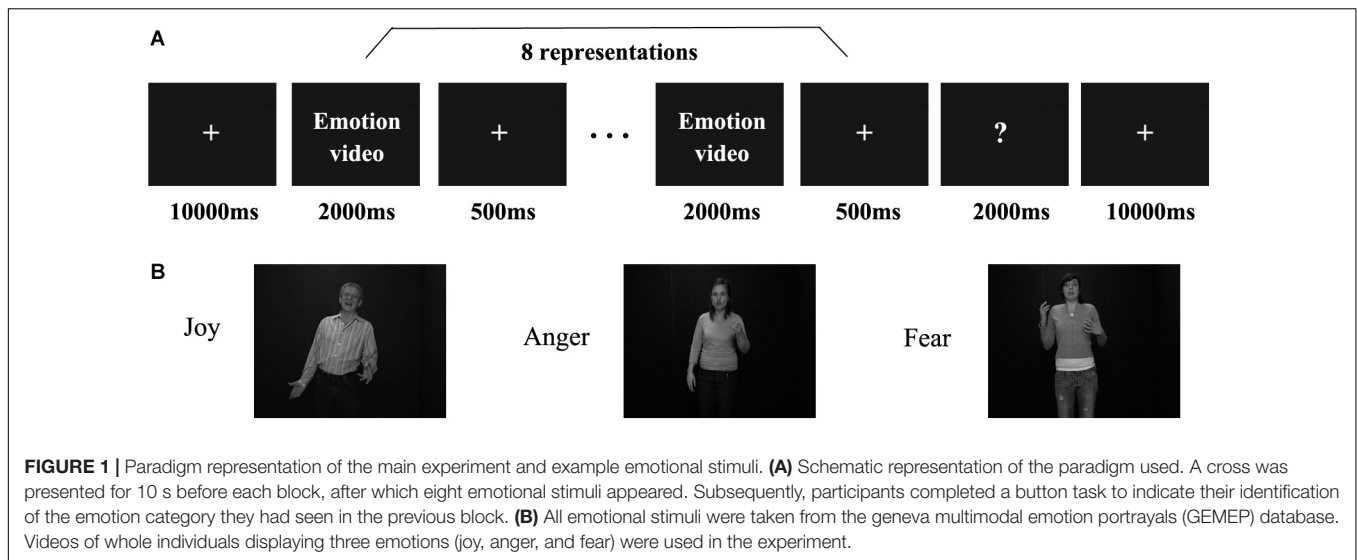
Preprocessing

Statistical parametric mapping 8 (SPM8) software¹ was used to preprocess the functional and structural images. For each functional run, the first five volumes were discarded to minimize the magnetic saturation effect. Slice-timing and head motion correction were performed for the remaining functional images. The threshold for head motion was $FD < 0.5 \text{ mm}$ (Power et al., 2012). Next, the structural images were co-registered to the mean functional image after motion correction, and were then unified segmented into gray matter, white matter (WM) and cerebrospinal fluid (CSF). The functional data were spatially normalized to the standard Montreal Neurological Institute (MNI) space using normalization parameters estimated from the unified segmentation, after which the voxel size was re-sampled into $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$. Subsequently, the normalized functional images of the localizer run were spatially smoothed with a 6-mm full-width at half-maximum Gaussian kernel to improve the signal-to-noise ratio.

Localization of Dorsal Motion-Sensitive Regions of Interest (ROIs)

Individual ROIs were defined using the localizer run data in which participants viewed static and dynamic faces, non-face bodies, whole persons and objects. At the first-level (within-subject) analysis, a general linear model (GLM) was constructed for each subject to estimate the task effect for each condition: dynamic face, static face, dynamic body, static body, dynamic whole-person, static whole-person, dynamic object and static

¹<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>



object. Each regressor was modeled by a boxcar function (representing the onsets and the durations of the stimulus blocks) convolved with a canonical hemodynamic response function (HRF). Several confounding nuisances were regressed out along with their temporal derivatives, including the realignment parameters from head motion correction and the physiological noise from WM and CSF were regressed using the CompCor (Behzadi et al., 2007; Whitfield-Gabrieli and Nieto-Castanon, 2012; Woo et al., 2014; Power et al., 2015; Xu et al., 2017; Geng et al., 2018). The low-frequency drifts of the time series were removed with a 1/128 Hz high-pass filter. The dorsal motion-sensitive ROIs were then identified by contrasting the average response to dynamic versus static conditions. The aim of using this contrast was to identify the motion-sensitive areas which are relatively domain-general, as both person and person parts, as well as those focused on non-person objects. We were especially interested in whether emotions perceived from whole-person expressions could be decoded from the relatively domain-general motion-sensitive areas, which are not specialized for representing only facial or bodily attributes. Thus, we chose to use a contrast which was expected to elicit motion areas to be domain general. Previous studies have showed that combined different types of stimuli together would be expected to localize motion-sensitive responses subsuming areas to be relatively domain-general (Furl et al., 2012, 2013, 2015; Liang et al., 2017). Therefore, to maximize the available data and to identify relatively domain-general motion-sensitive areas, we chose to average the results for ROI definition. We identified bilateral areas within human hMT + /V5 for all twenty subjects and bilateral STS areas for eighteen subjects, with two subjects only demonstrating a unilateral STS area in the left or right hemisphere. The ROIs were generated with a liberal threshold ($p < 0.05$; Skerry and Saxe, 2014; Miao et al., 2018; Yang et al., 2018). Individual subjects' motion-sensitive ROIs were defined as 9 mm spheres surrounding the peak coordinates. Subsequent emotion classification analyses were carried out based on these individually defined ROIs using the data from the main experiment runs. **Table 1** summarizes

TABLE 1 | Localization of motion-sensitive regions of interest (ROIs) used in the decoding analysis of main experiment data.

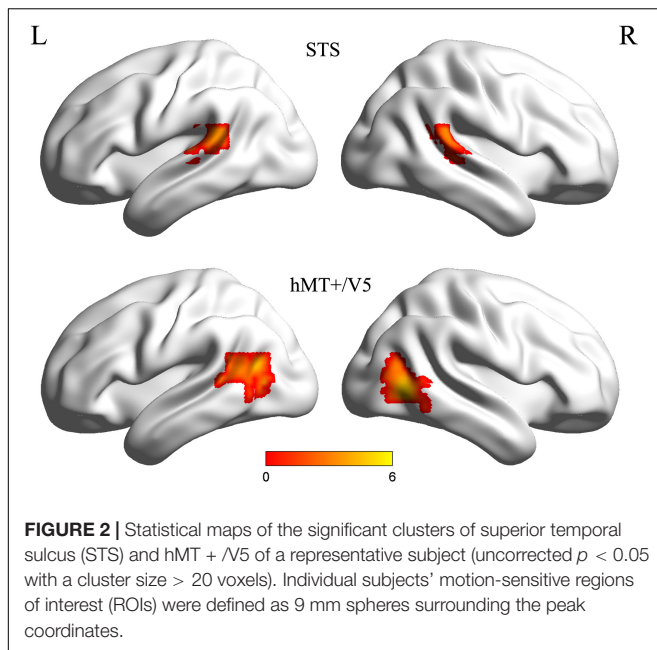
Functional ROIs	Hemisphere	Number of Subjects	MNI Coordinates		
			x	y	z
STS	L	19	-55 ± 10	-42 ± 5	14 ± 5
	R	19	57 ± 8	-39 ± 8	13 ± 6
hMT + /V5	L	20	-53 ± 5	-67 ± 5	6 ± 6
	R	20	53 ± 6	-64 ± 7	3 ± 5

Number of subjects in whom the ROIs were localized and the average Montreal Neurological Institute (MNI) coordinates (mean \pm standard deviation SD) are reported. STS, superior temporal sulcus.

the average MNI coordinates (mean \pm standard deviation SD) for each ROI, and **Figure 2** shows the statistical maps of the significant clusters in the ROI definition of a representative subject (uncorrected $p < 0.05$ with a cluster size > 20 voxels).

Within-Subject and Cross-Subject Emotion Classifications

Emotion classification analyses were conducted on the unsmoothed data from the main experiment (Harry et al., 2013; Yang et al., 2018) using a MVPA. We carried out MVPA classifications within ROIs that were functionally localized based on individual subject localizer runs. Similar procedures as those in previous MVPA studies were used in this study. For each participant, raw intensity values for all voxels within an ROI were extracted and normalized using the z -score function. The MVPA classification was carried out based on the multi-voxel activation patterns. Feature selection was performed using an ANOVA, which yielded a p -value for each voxel to tell the probability that a given voxel's activity varied significantly between emotion conditions. Feature selection was executed only on the training set to avoid peeking, and the threshold for ANOVA was $p < 0.05$. Next, the data were classified using a linear support vector



machine (SVM) that was implemented in LIBSVM² (Chang and Lin, 2011; Skerry and Saxe, 2014). The activation patterns for each condition were used to train and test the SVM classifier to perform classification over emotions. **Figure 3** represents the framework of our emotion classification analyses. We conducted two types of classifications in this section to assess the potential role of motion-sensitive ROIs in emotion decoding: first, a classical within-subject emotion classification was carried out as implemented in previous MVPA studies (classifier was trained and tested within the same subject data); next, a cross-subject emotion classification was conducted (classifier was trained iteratively on all subjects but one and tested on the remaining one) to assess whether there is any commonality to emotional representations in motion-sensitive areas across individual subjects. The cross-subject classification was performed using a leave-one-subject-out cross-validation (LOOCV) scheme (Chikazoe et al., 2014). In each fold of LOOCV, we trained the classifier in all but one subject and the remaining one was used as the test set. The cross-validation procedure was repeated until each subject was used as the test set, and the classification performance was averaged over all folds. The cross-subject classification was used to further investigate whether emotion codes in the motion-sensitive areas can be shared across subjects.

We ran both three-way (joy vs. anger vs. fear) and two-way (joy versus anger/fear, which could be considered as positive vs. negative) emotion classifications. The three-way classification was implemented similarly as previous MVPA studies (Wang et al., 2016; Liang et al., 2017), using a one-against-one voting strategy. That is, we obtained classifiers for each pair of emotions and these pairwise classifiers were then added to yield the linear ensemble classifier for each emotion. Classifying positive versus negative emotions is essential since

these results basically demonstrate coarse-grained emotion codes which can clearly distinguish positive-to-negative valences in bipolar representations, all the while taking into account the fact that some regions may not classify specific emotions in a fine-grained way, but may be able to distinguish positive and negative valence emotion representations (Kim et al., 2017). Data were partitioned into multiple cross-validation folds and the classification accuracies were averaged across folds to yield a single classification accuracy in each ROI. For the within-subject emotion classification, a cross-validation was performed across blocks, while for the cross-subject emotion classification, the cross-validation folds were based on subjects (testing each participant's activation pattern by a classifier that was trained by all other participants). For the classification of positive versus negative emotions, half of the data from anger and fear conditions were randomly dropped for each cross-validation, equating the base rates and therefore generating a chance level of 0.5 (Kim et al., 2017; Cao et al., 2018). To evaluate the emotion decoding performance, the significance of the classification results was established as a group level one-sample t -test above chance level (with a chance of 0.33 for the classification of three emotions, and a chance of 0.5 for the classification of positive versus negative emotions; Wurm and Lingnau, 2015; Cao et al., 2018), and were subsequently corrected for multiple comparisons by false discovery rate (FDR) and Bonferroni corrections according to the number of ROIs.

RESULTS

Behavioral Results

Behavioral results of the emotion classification accuracies and the reaction times for each emotion (joy, anger, and fear) are summarized in **Table 2**. These results confirmed the validity of the emotion stimuli used in our experiment as all emotions were well recognized with a high level of accuracy. Paired t -tests for the classification accuracies and reaction times were performed among the three emotions. Results showed that the classification accuracy for joy was significantly higher than that for anger and fear and that there was no significant difference between the accuracies for anger and fear [joy vs. anger: $t_{(19)} = 1.831$, $p = 0.041$; joy vs. fear: $t_{(19)} = 2.333$, $p = 0.015$; anger vs. fear: $t_{(19)} = 1.286$, $p = 0.107$; one-tailed]. For the reaction times, participants showed a significantly quicker response to joy than to anger or fear, and the response time for anger was shorter than that for fear [joy vs. anger: $t_{(19)} = -3.514$, $p = 0.001$; joy vs. fear: $t_{(19)} = -6.180$, $p < 0.001$; anger vs. fear: $t_{(19)} = -3.161$, $p = 0.003$; one-tailed].

Within-Subject and Cross-Subject Emotion Decoding Results

In this section, we conducted MVPA emotion classifications based on the individually defined ROIs. Two types of classification analyses were performed to assess the potential role of the motion-sensitive ROIs in emotion decoding. The first one was a classical within-subject emotion classification which was implemented in a similar way as previous MVPA

²<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

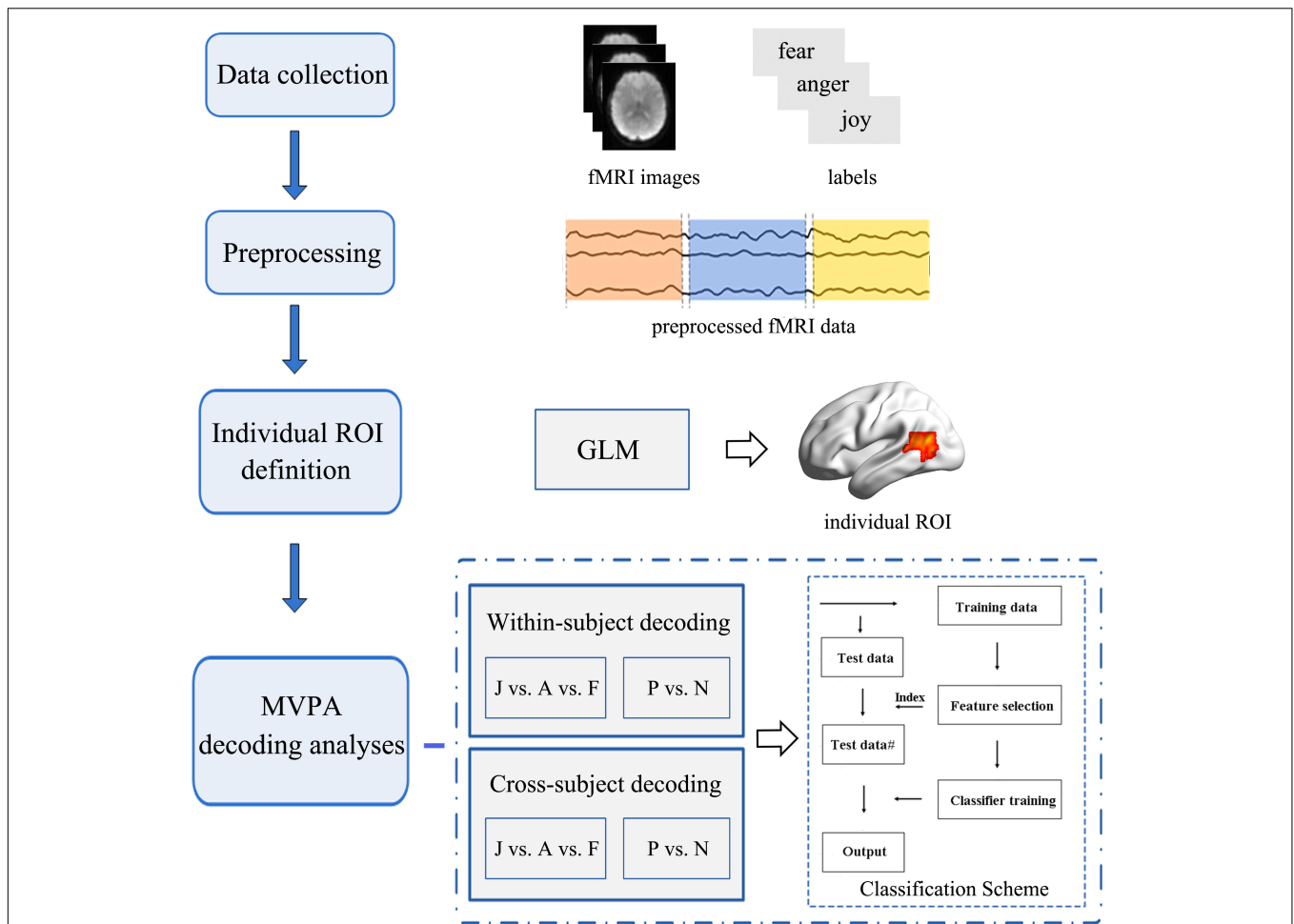


FIGURE 3 | Flowchart of the data analysis procedure. Multivariate pattern analysis (MVPA) emotion decoding analyses were conducted based on the individually defined regions of interest (ROIs). Two types of classification analyses (within-subject classification and cross-subject classification) were performed to examine the potential role of the motion-sensitive ROIs in the emotion decoding. Both three-way (joy vs. anger vs. fear: J vs. A vs. F) and two-way (positive vs. negative: P vs. N) emotion classifications were performed.

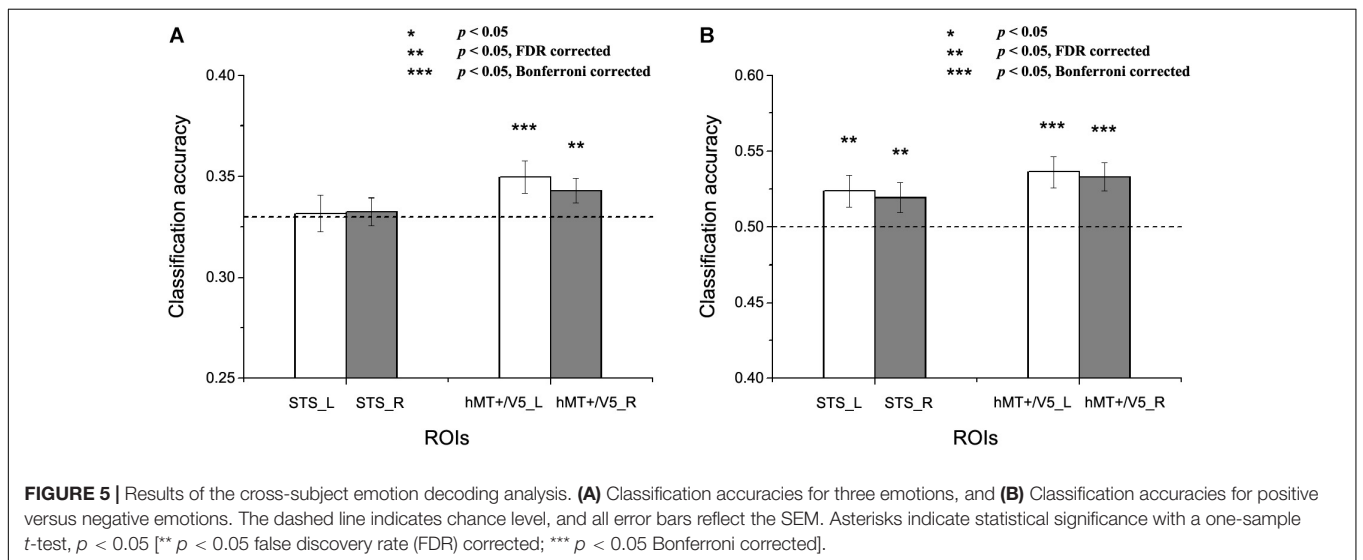
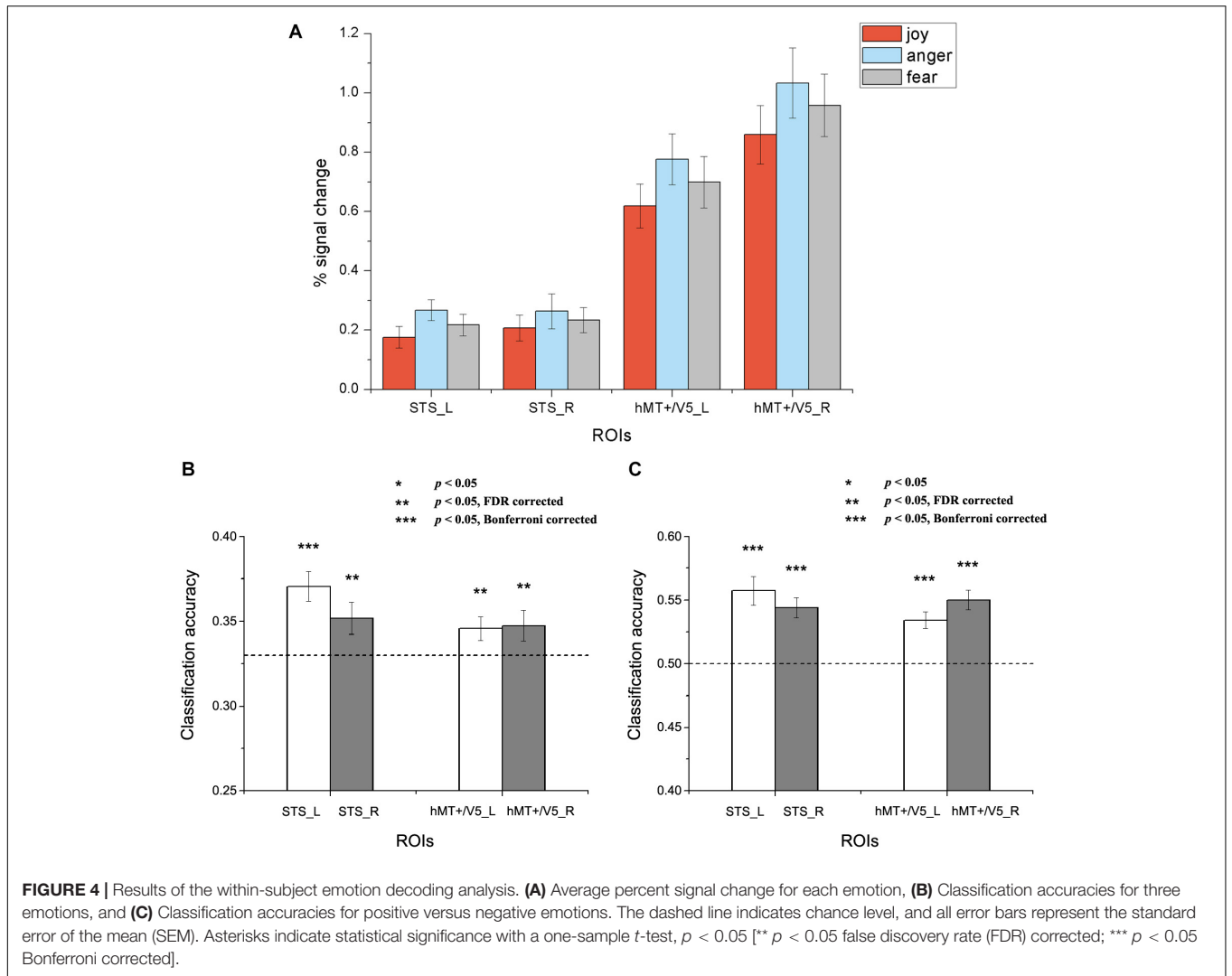
TABLE 2 | Behavioral results (mean % and standard deviations SD).

	Classification Accuracy (%)		Reaction Time (ms)	
	Mean	SD	Mean	SD
Joy	100	0	675.25	155.35
Anger	98.75	3.05	767.05	224.22
Fear	97.08	5.59	836.10	210.54

studies (Axelrod and Yovel, 2012, 2015; Wurm and Lingnau, 2015; Liang et al., 2017). In addition, we conducted a cross-subject emotion classification to assess whether there is any commonality in emotion representations in motion-sensitive areas across individual subjects. Both three-way (joy vs. anger vs. fear) and two-way (joy versus anger/fear, which could be considered as positive vs. negative) emotion classifications were performed. Feature selection was conducted using ANOVA which was executed only on the training data, with a threshold

of $p < 0.05$. SVM classifier was trained and tested with cross-validation scheme to perform classification analysis over emotion categories. The classification accuracies for each ROI and subject were entered into one-tailed one-sample t -tests against the chance levels (Wurm and Lingnau, 2015), and the statistical results were corrected for multiple comparisons by FDR and Bonferroni corrections according to the number of ROIs. **Figures 4, 5** separately show the results for within-subject and cross-subject emotion classifications and the statistical significances for multiple comparisons correction results are indicated by asterisks.

Results for the within-subject emotion decoding analysis are shown in **Figure 4**, which illustrates the average percent signal change for each emotion (**Figure 4A**) and the classification accuracies for three emotions (**Figure 4B**) and for positive versus negative emotions (**Figure 4C**) in all ROIs. We found that the classification accuracies for three emotions and for positive versus negative emotions were significantly higher than chance in all ROIs [For three emotions classification: left STS: $t_{(18)} = 4.692$,



$p < 0.001$; right STS: $t_{(18)} = 2.336$, $p = 0.016$; left hMT + /V5: $t_{(19)} = 2.294$, $p = 0.018$; right hMT + /V5: $t_{(19)} = 1.950$, $p = 0.033$. For positive versus negative emotions classification: left STS: $t_{(18)} = 5.149$, $p < 0.001$; right STS: $t_{(18)} = 5.478$, $p < 0.001$; left hMT + /V5: $t_{(19)} = 5.202$, $p < 0.001$; right hMT + /V5: $t_{(19)} = 6.548$, $p < 0.001$).

We next assessed whether a person's emotion representations in the motion-sensitive areas could be predicted by others' emotion representations. **Figure 5** shows the results for the cross-subject emotion classifications in all motion-sensitive ROIs (**Figure 5A** shows the classification results for three emotions and **5B** shows the classification results for positive versus negative emotions). When classifying emotions from the classifiers trained by the activation patterns of other subjects, we found that classification accuracies were significantly higher than chance in hMT + /V5 both for the three emotions and for the positive versus negative emotions [classification of three emotions: left hMT + /V5: $t_{(19)} = 2.483$, $p = 0.01$; right hMT + /V5: $t_{(19)} = 2.116$, $p = 0.024$; classification of positive vs. negative emotions: left hMT + /V5: $t_{(19)} = 3.510$, $p = 0.001$; right hMT + /V5: $t_{(19)} = 3.523$, $p = 0.001$]. In the STS, although the classification accuracies for the three emotions did not achieve significance [left STS: $t_{(18)} = 0.174$, $p = 0.432$; right STS: $t_{(18)} = 0.351$, $p = 0.365$], we did find successful cross-subject positive-to-negative emotion decoding [left STS: $t_{(18)} = 2.199$, $p = 0.021$; right STS: $t_{(18)} = 1.995$, $p = 0.031$].

DISCUSSION

In this study, we performed a block-design fMRI experiment and collected neural data while participants viewed emotions (joy, anger, and fear) from videos representing whole-person expressions. Both within-subject and cross-subject MVPA emotion classification analyses were performed to examine the decoding performance of individual-defined motion-sensitive ROIs. We ran both three-way (joy vs. anger vs. fear) and two-way (positive vs. negative) emotion classifications. Our results showed that emotions could be successfully decoded based on the activation patterns in dorsal motion-sensitive areas. Moreover, results from the cross-subject classification analysis showed that motion-sensitive areas supported the classification of individual emotion representation across subjects.

Emotions Perceived From Whole-Person Expressions Are Represented in Dorsal Motion-Sensitive Areas

We obtained significant classification results for both the classification of the three emotions and the positive versus negative emotions, indicating that emotions perceived from whole-person expressions are represented in the motion-sensitive areas.

Previous studies on facial expressions with dynamic stimuli have revealed a certain degree of sensitivity in dorsal temporal areas, showing that motion-sensitive areas within hMT + /V5 and STS exhibited strong responses to dynamic facial emotions (Foley et al., 2011; Furl et al., 2013, 2015). Considering that the

results of the average response from the univariate analysis alone are insufficient to reveal the potential role of a specific brain area underlying decoding (Axelrod and Yovel, 2012; Mahmoudi et al., 2012), recent fMRI studies used ROI-based MVPA to examine the decoding performance of motion-sensitive areas. Furl et al. (2012) used macaque STS as a model system and revealed the successful decoding of facial emotions in motion-sensitive areas. Similar results were obtained in one of our recent studies (Liang et al., 2017). These studies suggest that motion-sensitive areas may transmit measurable quantities of expression information and may be involved in the processing of emotional information. In this study, we defined individual motion-sensitive ROIs and found that emotions perceived from whole-person expressions could be successfully decoded from motion-sensitive areas. Our results are consistent with previous findings, and provide new evidence that emotions perceived from whole-person expressions are represented in the motion-sensitive areas. It should be noted that our results revealed the emotion decoding performance of the relatively domain-general motion-sensitive areas, as the localization contrast we used contained both person and person parts, as well as non-person objects, which was expected to reflect all responses to visual motion (Furl et al., 2012, 2015). Therefore, our results suggest that motion sensitive voxels which respond to various motions, not only specific to facial or bodily attributes, may make a significant contribution to emotion decoding.

Taken together, our findings provide new evidence that emotions are represented in dorsal motion-sensitive areas, pointing to the role of dorsal motion-sensitive areas as key regions in the processing of emotional information in daily communication.

Commonality of Emotion Representations in Motion-Sensitive Areas Across Individuals

Furthermore, we assessed whether an individual's emotion representation in the motion-sensitive areas corresponds to that observed in others by conducting a cross-subject emotion classification analysis (classifier was trained iteratively on all subjects but one and tested on the remaining one). This may provide evidence of whether an individual's subjective emotion representation in the motion-sensitive areas corresponds to that observed in others, which may be helpful in evaluating the commonality and variability in emotion coding (Haxby et al., 2011; Raizada and Connolly, 2012; Chikazoe et al., 2014). We obtained statistically significant results for both the cross-subject classification of three emotions and positive versus negative emotions in the hMT + /V5, indicating that the hMT + /V5 code may reflect experienced emotions in the same way across participants. In addition, although much less significant emotion classification results were identified for the three emotions, we revealed the successful cross-subject classification of positive versus negative emotions in the STS. This reveals that population codes in the STS were less able to decode a specific emotion in a fine-grained way, but demonstrated that the similarity in emotion representations among people may allow for the robust distinction of coarsely defined positive-to-negative

emotional valences in the context of bipolar representations (Kim et al., 2017). Our results also suggest that subjective emotion representations are more similarly structured across individual subjects in the hMT + /V5 than in the STS, since hMT + /V5 supported the cross-subject classification of both fine-grained three emotions and coarse-grained positive-to-negative emotions, while the STS only supported the coarse-grained classification in a significant way.

Overall, our study indicates that the representation of emotions in motion-sensitive areas may be similar across participants. This may provide evidence that even in the most subjective perception of an individual's emotion experience, its internal emotion coding can be predicted on the basis of the patterns observed in others in the motion-sensitive areas. This finding is important, since such cross-subject commonality may allow for the common scaling of the valence of emotional experiences across participants. In summary, we show that a person's emotional representations in motion-sensitive areas may be predicted by others' emotional representations, suggesting that there exists a common emotion code in the motion-sensitive areas across individuals.

In the present study, different types of emotional stimuli (facial, bodily, and whole-person expressions) were contained in the main experiment. Future studies with whole-person stimuli separately may further improve the implementation of the classification scheme and lead to better understanding of the whole-person expressions decoding. In addition, compared with ROI-based analyses, whole-brain group-level analyses would provide more informative results. Future studies combine both whole-brain activation-based and FC-based analyses would further enrich our findings about the neural substrates and the mechanisms for the quick and effortless recognition of whole-person emotions.

CONCLUSION

Our results showed that emotions perceived from whole-person expressions can be robustly decoded in dorsal motion-sensitive areas. Moreover, successful cross-subject emotion decoding suggests that the emotion representations in motion-sensitive areas could be shared across participants. This study extends previous MVPA studies to the emotion perception of whole-person expressions, which are more frequently perceived in daily life, and may further our understanding of the potential neural substrates underlying the efficient recognition of emotions from body language. Our findings provide new evidence that emotions are represented in dorsal motion-sensitive areas, underscoring

REFERENCES

- Axelrod, V., and Yovel, G. (2012). Hierarchical processing of face viewpoint in human visual cortex. *J. Neurosci.* 32, 2442–2452. doi: 10.1523/jneurosci.4770-11.2012
- Axelrod, V., and Yovel, G. (2015). Successful decoding of famous faces in the fusiform face area. *PLoS One* 10:e0117126. doi: 10.1371/journal.pone.0117126

the important role of the motion-sensitive areas in the emotion perception. Our study also suggests that emotion representations in motion-sensitive areas are similar across individuals.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by local Ethics Committee of Yantai Affiliated Hospital of Binzhou Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YL and BL designed the study. YL performed the experiments, analyzed results, and wrote the manuscript. Both authors have approved the final manuscript.

FUNDING

This work was supported by the Beijing Natural Science Foundation (No. 4204089), National Natural Science Foundation of China (Nos. 61906006 and U1736219), China Postdoctoral Science Foundation funded project (No. 2018M641135), Beijing Postdoctoral Research Foundation (No. zz2019-74), Beijing Chaoyang District Postdoctoral Research Foundation (zz2019-02), National Key Research and Development Program of China (No. 2018YFB0204304), and Fundamental Research Funds for the Central Universities of China (No. FRF-MP-19-007).

ACKNOWLEDGMENTS

We would like to thank the Medical Imaging Research Institute of Binzhou Medical University, Yantai Affiliated Hospital of Binzhou Medical University, as well as the volunteers for the assistance in the data acquisition. We also thank Prof. Irene Rotondi (Campus Biotech, University of Geneva, Switzerland) for supplying the GEMEP Corpus.

- Banziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827
- Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37, 90–101. doi: 10.1016/j.neuroimage.2007.04.042

- Calbi, M., Angelini, M., Gallese, V., and Umiltà, M. A. (2017). "Embodied Body Language": an electrical neuroimaging study with emotional faces and bodies. *Sci. Rep.* 7:6875.
- Candidi, M., Stienen, B. M., Aglioti, S. M., and De Gelder, B. (2011). Event-related repetitive transcranial magnetic stimulation of posterior superior temporal sulcus improves the detection of threatening postural changes in human bodies. *J. Neurosci.* 31, 17547–17554. doi: 10.1523/jneurosci.0697-11.2011
- Cao, L. J., Xu, J. H., Yang, X. L., Li, X. L., and Liu, B. L. (2018). Abstract representations of emotions perceived from the face, body, and whole-person expressions in the left postcentral gyrus. *Front. Hum. Neurosci.* 12:419. doi: 10.3389/fnhum.2018.00419
- Chang, C., and Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intellig. Syst. Tech.* 2, 1–27. doi: 10.1145/1961189.1961199
- Chikazoe, J., Lee, D., Kriegeskorte, N., and Anderson, A. (2014). Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* 17, 1114–1122. doi: 10.1038/nn.3749
- de Gelder, B., Van Den Stock, J., Meeren, H. K., Sinke, C. B., Kret, M. E., and Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neurosci. Biobehav. Rev.* 34, 513–527. doi: 10.1016/j.neubiorev.2009.10.008
- Downing, P. E., and Peelen, M. V. (2016). Body selectivity in occipitotemporal cortex: causal evidence. *Neuropsychologia* 83, 138–148. doi: 10.1016/j.neuropsychologia.2015.05.033
- Foley, E., Rippon, G., Thai, N., Longe, O., and Senior, C. (2011). Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. *J. Cogn. Neurosci.* 24, 507–520. doi: 10.1162/jocn_a_00120
- Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B. B., and Ungerleider, L. G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J. Neurosci.* 32, 15952–15962. doi: 10.1523/jneurosci.1992-12.2012
- Furl, N., Henson, R. N., Friston, K. J., and Calder, A. J. (2013). Top-down control of visual responses to fear by the amygdala. *J. Neurosci.* 33, 17435–17443. doi: 10.1523/jneurosci.2992-13.2013
- Furl, N., Henson, R. N., Friston, K. J., and Calder, A. J. (2015). Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cereb. Cortex* 25, 2876–2882. doi: 10.1093/cercor/bhu083
- Geng, X., Xu, J., Liu, B., and Shi, Y. (2018). Multivariate classification of major depressive disorder using the effective connectivity and functional connectivity. *Front. Neurosci.* 12:38. doi: 10.3389/fnins.2018.00038
- Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi: 10.1016/j.neuropsychologia.2006.04.015
- Harry, B., Williams, M. A., Davis, C., and Kim, J. (2013). Emotional expressions evoke a differential response in the fusiform face area. *Front. Hum. Neurosci.* 7:692. doi: 10.3389/fnhum.2013.00692
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., et al. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416. doi: 10.1016/j.neuron.2011.08.026
- Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0
- Henriksson, L., Mur, M., and Kriegeskorte, N. (2015). Faciotopy-A face-feature map with face-like topology in the human occipital face area. *Cortex* 72, 156–167. doi: 10.1016/j.cortex.2015.06.030
- Johnston, P., Mayes, A., Hughes, M., and Young, A. W. (2013). Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex* 49, 2462–2472. doi: 10.1016/j.cortex.2013.01.002
- Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., and Peelen, M. V. (2014). Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *J. Neurophysiol.* 111, 82–90. doi: 10.1152/jn.00371.2013
- Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 2109–2128.
- Kim, J., Shinkareva, S., and Wedell, D. (2017). Representations of modality-general valence for videos and music derived from fMRI data. *Neuroimage* 148, 42–54. doi: 10.1016/j.neuroimage.2017.01.002
- Kret, M. E., Pichon, S., Grezes, J., and De Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage* 54, 1755–1762. doi: 10.1016/j.neuroimage.2010.08.012
- Liang, Y., Liu, B. L., Xu, J. H., Zhang, G. Y., Li, X. L., Wang, P. Y., et al. (2017). Decoding facial expressions based on face-selective and motion-sensitive areas. *Hum. Brain Mapp.* 38, 3113–3125. doi: 10.1002/hbm.23578
- Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., and Brovelli, A. (2012). Multivoxel pattern analysis for fMRI data: a review. *Comput. Math Methods Med.* 2012:961257.
- Miao, Q. M., Zhang, G. Y., Yan, W. R., and Liu, B. L. (2018). Investigating the brain neural mechanism when signature objects were masked during a scene categorization task using functional MRI. *Neuroscience* 388, 248–262. doi: 10.1016/j.neuroscience.2018.07.030
- Minnebusch, D. A., and Daum, I. (2009). Neuropsychological mechanisms of visual face and body perception. *Neurosci. Biobehav. Rev.* 33, 1133–1144. doi: 10.1016/j.neubiorev.2009.05.008
- Pitcher, D. (2014). Facial expression recognition takes longer in the posterior superior temporal sulcus than in the occipital face area. *J. Neurosci.* 34, 9173–9177. doi: 10.1523/jneurosci.5038-13.2014
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018
- Power, J. D., Schlaggar, B. L., and Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 105, 536–551. doi: 10.1016/j.neuroimage.2014.10.044
- Raizada, R. D., and Connolly, A. C. (2012). What makes different people's representations alike: neural similarity space solves the problem of across-subject fMRI decoding. *J. Cogn. Neurosci.* 24, 868–877. doi: 10.1162/jocn_a_00189
- Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., and Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vis.* 10, 71–76.
- Skerry, A., and Saxe, R. (2014). A common neural code for perceived and inferred emotion. *J. Neurosci.* 34, 15997–16008. doi: 10.1523/jneurosci.1676-14.2014
- Soria Bauser, D., and Suchan, B. (2015). Is the whole the sum of its parts? Configural processing of headless bodies in the right fusiform gyrus. *Behav. Brain Res* 281, 102–110. doi: 10.1016/j.bbr.2014.12.015
- Wang, X. S., Fang, Y. X., Cui, Z. X., Xu, Y. W., He, Y., Guo, Q. H., et al. (2016). Representing object categories by connections: evidence from a multivariate connectivity pattern classification approach. *Hum. Brain Mapp.* 37, 3685–3697. doi: 10.1002/hbm.23268
- Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., et al. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* 69, 131–140. doi: 10.1016/j.cortex.2015.05.003
- Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect* 2, 125–141. doi: 10.1089/brain.2012.0073
- Woo, C. W., Koban, L., Kross, E., Lindquist, M. A., Banich, M. T., Ruzic, L., et al. (2014). Separate neural representations for physical pain and social rejection. *Nat. Commun.* 5:5380.
- Wurm, M., and Lingnau, A. (2015). Decoding actions at different levels of abstraction. *J. Neurosci.* 35, 7727–7735. doi: 10.1523/jneurosci.0188-15.2015
- Xu, J., Yin, X., Ge, H., Han, Y., Pang, Z., Liu, B., et al. (2017). Heritability of the effective connectivity in the resting-state default mode network. *Cereb. Cortex* 27, 5626–5634. doi: 10.1093/cercor/bhw332
- Yang, X. L., Xu, J. H., Cao, L. J., Li, X. L., Wang, P. Y., Wang, B., et al. (2018). Linear representation of emotions in whole persons by combining facial and

- bodily expressions in the extrastriate body area. *Front. Hum. Neurosci.* 11:653. doi: 10.3389/fnhum.2017.00653
- Zhang, J., Li, X., Song, Y., and Liu, J. (2012). The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PLoS One* 7:e40390. doi: 10.1371/journal.pone.0040390
- Zhu, Q., Nelissen, K., Van Den Stock, J., De Winter, F. L., Pauwels, K., De Gelder, B., et al. (2013). Dissimilar processing of emotional facial expressions in human and monkey temporal cortex. *Neuroimage* 66, 402–411. doi: 10.1016/j.neuroimage.2012.10.083

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.