

Decoding Facial Expressions Based on Face-Selective and Motion-Sensitive Areas

Yin Liang,¹ Baolin Liu ^{1,2*} Junhai Xu,¹ Gaoyan Zhang,¹ Xianglin Li,³ Peiyuan Wang,⁴ and Bin Wang³

¹School of Computer Science and Technology, Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, 300350, People's Republic of China

²State Key Laboratory of Intelligent Technology and Systems, National Laboratory for Information Science and Technology, Tsinghua University, Beijing, 100084, People's Republic of China

³Medical Imaging Research Institute, Binzhou Medical University, Yantai, Shandong, 264003, People's Republic of China

⁴Department of Radiology, Yantai Affiliated Hospital of Binzhou Medical University, Yantai, Shandong, 264003, People's Republic of China

Abstract: Humans can easily recognize others' facial expressions. Among the brain substrates that enable this ability, considerable attention has been paid to face-selective areas; in contrast, whether motion-sensitive areas, which clearly exhibit sensitivity to facial movements, are involved in facial expression recognition remained unclear. The present functional magnetic resonance imaging (fMRI) study used multi-voxel pattern analysis (MVPA) to explore facial expression decoding in both face-selective and motion-sensitive areas. In a block design experiment, participants viewed facial expressions of six basic emotions (anger, disgust, fear, joy, sadness, and surprise) in images, videos, and eyes-obscured videos. Due to the use of multiple stimulus types, the impacts of facial motion and eye-related information on facial expression decoding were also examined. It was found that motion-sensitive areas showed significant responses to emotional expressions and that dynamic expressions could be successfully decoded in both face-selective and motion-sensitive areas. Compared with static stimuli, dynamic expressions elicited consistently higher neural responses and decoding performance in all regions. A significant decrease in both activation and decoding accuracy due to the absence of eye-related information was also observed. Overall, the findings showed that emotional expressions are represented in motion-sensitive areas in addition to conventional face-selective areas, suggesting that motion-sensitive regions may also effectively contribute to facial expression recognition. The results also suggested that facial motion and eye-related information played important roles by carrying considerable expression information that could facilitate facial expression recognition. *Hum Brain Mapp* 00:000–000, 2017. © 2017 Wiley Periodicals, Inc.

Contract grant sponsor: National Basic Research Program (973 Program) of China; Contract grant number: No. 2013CB329301; Contract grant sponsor: National Natural Science Foundation of China; Contract grant number: No. 61271128, No. 61571327; No. 61503278; Contract grant sponsor: Shandong Provincial Natural Science Foundation of China; Contract grant number: No. ZR2015HM081; Contract grant sponsor: Project of Shandong Province Higher Educational Science and Technology Program; Contract grant number: J15LL01

*Correspondence to: Baolin Liu, School of Computer Science and Technology, Tianjin University, Tianjin 300350, People's Republic of China. E-mail: liubaolin@tsinghua.edu.cn

Received for publication 9 June 2016; Revised 21 February 2017; Accepted 9 March 2017.

DOI: 10.1002/hbm.23578

Published online 00 Month 2017 in Wiley Online Library (wileyonlinelibrary.com).

Key words: facial expressions; fMRI; MVPA; face-selective areas; motion-sensitive areas; facial motion; eye-related information

INTRODUCTION

Facial expressions are important for social interactions as they convey information about others' emotions. Since humans can quickly and effortlessly recognize others' facial expressions, a key question is which brain regions are involved in facial expression recognition.

Functional magnetic resonance imaging (fMRI) studies have identified a network of cortical areas involved in face processing. Among these areas, the occipital face area (OFA) in the inferior occipital gyrus [Gauthier et al., 2000; Ishai et al., 2005], the fusiform face area (FFA) in the fusiform gyrus [Grill-Spector et al., 2004; Haxby et al., 2000; Rotshtein et al., 2005; Yovel and Kanwisher, 2004], and the face-selective area in the posterior superior temporal sulcus (pSTS) [Allison et al., 2000; Gobbini et al., 2011; Haxby et al., 2000; Lee et al., 2010; Winston et al., 2004] together constitute the "core system" [Gobbini and Haxby, 2007; Haxby et al., 2000]. These three face-selective areas have been widely studied for their role in facial expression perception and are considered key regions for the processing of facial expressions [Fox et al., 2009b; Ganel et al., 2005; Harry et al., 2013; Hoffman and Haxby, 2000; Kawasaki et al., 2012; Monroe et al., 2013; Said et al., 2010; Sato et al., 2004]. Previous fMRI studies mainly used static expression images as stimuli [Adolphs, 2002; Andrews and Ewbank, 2004; Gur et al., 2002; Murphy et al., 2003; Phan et al., 2002]. However, facial expressions are dynamic in natural social contexts, and facial motion conveys additional information about temporal and structural facial properties [Harwood et al., 1999; Sato et al., 2004]. Recent studies have suggested that the use of dynamic stimuli may be more appropriate when investigating the "authentic" mechanisms of facial expression recognition that people use in their daily life [Johnston et al., 2013; Trautmann et al., 2009]. These studies showed enhanced brain activation patterns and found that in addition to the conventional face-selective areas, motion-sensitive areas in the human motion complex (V5f) and pSTS are also sensitive to facial motion [Foley et al., 2011; Furl et al., 2013, 2015; Pitcher et al., 2011; Schultz et al., 2013; Schultz and Pilz, 2009]. These findings provided the neural substrates for further study of facial expression recognition.

The identification of these face-selective and motion-sensitive areas has led to the question of which brain regions are involved in facial expression recognition as well as to explore the roles of these regions in expression decoding. Multi-voxel pattern analysis (MVPA) provides a method to examine whether expression information could be decoded from the distributed patterns of activity

evoked in a specific brain area [Mahmoudi et al., 2012; Norman et al., 2006; Williams et al., 2007]. Using MVPA, Said et al. [2010] found the successful classification of facial expressions in the STS, while the role of the FFA was tested by Harry et al. [2013]. As these studies focused only on single regions, Wegrzyn et al. [2015] directly compared classification rates across the brain areas proposed by Haxby's model [Haxby et al., 2000]. However, the study used only a subset of the basic emotions and only images as stimuli; the role of motion-related cues in facial expression decoding was not considered. Consequently, comprehensively evaluating the decoding performance for both static and dynamic facial expressions of the full set of six basic emotions would be meaningful. Moreover, emotional expressions could be represented in the motion-sensitive areas (Mf areas) of macaques [Furl et al., 2012], suggesting that human motion-sensitive areas might also represent expression information. Hence, it is necessary to investigate the decoding performance of motion-sensitive areas in addition to that of the conventional face-selective areas. Additionally, some psychologists who have investigated gaze fixation and the weights of different facial components in facial expression perception found a preference for the expressive information conveyed by the eyes in Asians [Jack et al., 2009, 2012; Kret et al., 2013; Yuki et al., 2007]. Their findings motivated us to explore the particular impact of the eyes in facial expression decoding.

The present fMRI study explored whether face-selective and motion-sensitive areas effectively contributed to facial expression recognition. We hypothesized that because motion-sensitive areas clearly respond to facial motion, facial expression information may also be represented in these regions. To address this question, we used a block design experiment and collected neural activity while participants viewed facial expressions of the six basic emotions (anger, disgust, fear, joy, sadness, and surprise) expressed in images, videos, and obscured-eye videos. Because multiple stimulus types were included, we also investigated the impacts of facial motion and eye-region information on facial expression decoding. Regions of interest (ROIs) were identified using a separate localizer. We conducted univariate analysis and MVPA to examine neural responses and decoding performance for facial expressions in all ROIs.

MATERIALS AND METHODS

Participants

Twenty healthy, right-handed volunteers (10 females; average age: 22.25 years; range: 20–24 years) with no

history of neurological or psychiatric disorders took part in the study. All participants had normal or corrected-to-normal vision and provided written informed consent before the experiment. The study was approved by the Institutional Review Board (IRB) of Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University. Data of two participants were discarded from further analysis due to excessive head movement during scanning.

Experimental Stimuli

The facial stimuli used in this study were taken from the Amsterdam Dynamic Facial Expression Set [Van der Schalk et al., 2011]. Video clips of 12 different individuals (6 male and 6 female) displaying expressions of the six basic emotions (anger, disgust, fear, joy, sadness, and surprise) were chosen as the dynamic stimuli (see Fig. 1A). All videos were cropped to 1,520 ms to retain the transition from a neutral expression to the expression apex, and the final still images depicting the apex expressions were used as the static stimuli. To explore the specific role of eye-related information in facial expression decoding, we masked the eye region in the dynamic facial videos by blurring to create the obscured stimuli (with Adobe Premiere) [Engell and McCarthy, 2014; Stock et al., 2011; Tamietto et al., 2009]. Exemplar obscured stimuli are shown in Figure 1B. All videos and images were converted into grayscale.

Decoding Experiment

There were three conditions in our decoding experiment: (1) static expressions, (2) dynamic expressions, and (3) dynamic expressions with obscured eye-region. Each condition included all six basic expressions: anger, disgust, fear, joy, sadness, and surprise. The experiment used a block design, with four runs. In each run, the participants viewed 18 blocks (3 conditions \times 6 emotions) in a pseudo-random order. Figure 1C shows the schematic representation of the employed paradigm. Each run began with a fixation cross (10 s), which was followed by a stimulus block (the same condition and expression) and then a 4-s button press task. Successive stimulus blocks were separated by 10 s fixation cross. In each stimulus block, 12 expression stimuli were presented (each for 1,520 ms) with an inter-stimulus interval (ISI) of 480 ms. During the button press task, the participants were asked to choose between six basic emotions by pressing a button. The total duration of the decoding experiment was 45.6 min, with each run lasting 11.4 min. The stimuli were presented using E-Prime 2.0 Professional (Psychology Software Tools, Pittsburgh, PA, USA).

After scanning, participants were required to complete a supplementary behavioral experiment in which they made speeded categorization of emotional category for each face stimulus in the decoding experiment and rated the

emotional intensity on a 1–9 scale. Each stimulus was presented once in a random order, with the same duration as in the fMRI experiment.

fMRI Data Acquisition

fMRI data were collected at Yantai Affiliated Hospital of Binzhou Medical University using a 3.0-T Siemens scanner with an eight-channel head coil. Foam pads and earplugs were used to reduce head motion and scanner noise. T2*-weighted images were acquired using a gradient echo-planar imaging (EPI) sequence (TR = 2,000 ms, TE = 30 ms, voxel size = $3.1 \times 3.1 \times 4.0$ mm³, matrix size = 64×64 , slices = 33, slice thickness = 4 mm, slice gap = 0.6 mm). T1-weighted anatomical images were acquired with a three-dimensional magnetization-prepared rapid-acquisition gradient echo (3D MPRAGE) sequence (TR = 1,900 ms, TE = 2.52 ms, TI = 1,100 ms, voxel size = $1 \times 1 \times 1$ mm³, matrix size = 256×256). The participants viewed the stimuli through the high-resolution stereo 3D glasses of the VisuaStim Digital MRI Compatible fMRI system.

Data Preprocessing

Functional images were preprocessed with SPM8 software (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). For each run, the first 5 volumes were discarded to minimize the effects of magnetic saturation. The remaining images were corrected for slice timing and head motion, normalized in the standard Montreal Neurological Institute (MNI) space using anatomical image unified segmentation, subsampled at an isotropic voxel size of 3 mm. The normalized data were smoothed with a 4-mm full-width at half-maximum Gaussian kernel. In subsequent analyses, the smoothed and unsmoothed data were used separately in the univariate and pattern classification analyses.

Localization of Face-Selective and Motion-Sensitive Areas

For the purpose of decoding, we identified face-selective and motion-sensitive areas using a separate localizer run. The total duration of the localizer was 7.8 min. During this scan, the participants viewed video clips or static images of objects and faces presented in separate blocks. Each condition appeared three times in a pseudo-random order, resulting in 12 blocks total. In each block, 14 stimuli (12 novel and 2 repeated) were presented, each of which was displayed for 1,520 ms, with an ISI of 480 ms. The participants performed a “one-back” task by pressing the button whenever two identical presentations appeared consecutively. There was some overlap between the face stimuli used in the localizer and decoding experiment runs. The object stimuli were chosen from materials used in a previous study [Fox et al., 2009a]. All selected objects displayed types of motion that did not create large transitions in

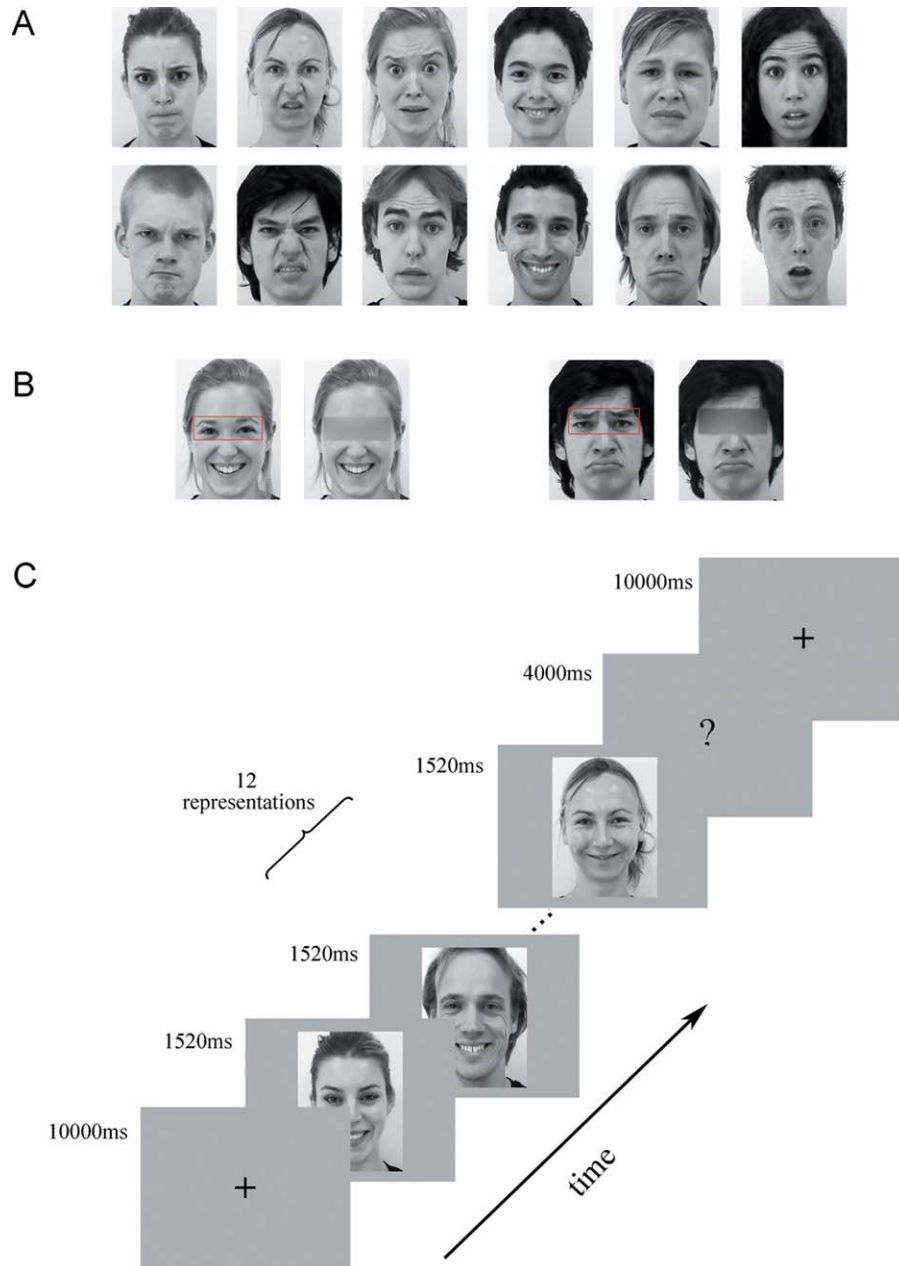


Figure 1.

Exemplar facial stimuli and schematic representation of the employed paradigm. (A) All facial stimuli were taken from the ADFES database. Twelve different individuals (6 males) displaying expressions of the 6 basic emotions (anger, disgust, fear, joy, sadness, and surprise) were chosen. (B) Exemplar obscured stimuli. (C) Schematic representation of the decoding

experiment. A cross was presented for 10 s before each block, and then 12 expression stimuli (all in same condition and expression) appeared. Finally, the participants completed a button task to indicate their discrimination of the expression they had seen in the previous block. [Color figure can be viewed at wileyonlinelibrary.com]

position so that the dynamic changes in the objects could be compared with those in the faces. The preprocessing procedures were similar to those used for the decoding experiment (see Data preprocessing). At the first level,

four effects of interest were modeled: dynamic face, static face, dynamic object, and static object. Six head motion parameters were regressed as covariates to correct for movement-related artifacts, and low-frequency drifts were

removed with a high-pass filter of 1/128 Hz. Face-selective and motion-sensitive areas were then functionally identified by contrasting the average response to dynamic and static faces versus the average response to dynamic and static objects and by contrasting dynamic versus static faces separately.

Univariate Analysis

For each participant, we first examined the mean response profiles to static, dynamic, and obscured eye dynamic facial expressions in each face-selective and motion-sensitive ROI. Averaged time courses across voxels in each ROI were extracted for each condition using the MarsBaR software package (<http://marsbar.sourceforge.net/>, Brett et al. [2002]). The estimated signal change data were then subjected to further statistical analysis.

Multivoxel Pattern Analysis

We used MVPA to examine whether facial expressions could be decoded from the neural activation patterns in each ROI. Our decoding strategy was to perform a six-way expression classification separately for the static, dynamic and obscured eye expression conditions. MVPA was conducted on the preprocessed data (unsmoothed) from the decoding experiment. Raw intensity values for all voxels within an ROI were extracted and normalized with a z-score transformation function. This procedure was used to remove any between-run differences due to baseline shifts and was applied for the full-run voxel time course (for each run separately). To account for hemodynamic lag, the time courses were shifted 2 TRs (4 s). With 12 trials in each block, across the 4 runs, there were $12 \times 4 = 48$ samples for each expression in each condition [Axelrod and Yovel, 2012; Song et al., 2011; Wolbers et al., 2011]. We employed a lib-SVM classifier with leave-one-run-out cross-validation, and then the results were averaged. Feature selection was executed using ANOVA, yielding a P value for each voxel that expressed the probability that a given voxel's activity varied significantly between conditions over the course of the experiment. The ANOVA analysis was performed only on the training data to avoid peeking, and any voxel that did not pass a threshold ($P < 0.05$) was not used for further classification analysis. To ensure that the classification scheme was valid, the same cross-validation schemes were carried out for 1,000 random shuffles of class labels, which generalized with chance performance. To evaluate decoding performance, the significance of the classification results was established as a group level one-sample t -test above the chance performance (0.1667) and was FDR corrected for the number of ROIs.

RESULTS

Localization of ROIs

By contrasting faces versus objects, we identified conventional face-selective areas, including the bilateral FFA, OFA, and a posterior part of the right STS, that responded more to faces than to non-face stimuli. We also identified motion-sensitive areas that exhibited sensitivity to facial movements, as reported in recent studies, by contrasting dynamic versus static facial expressions [Furl et al., 2012, 2013, 2015]. We found bilateral areas (V5f) sensitive to facial motion within human hMT+/V5 [Foley et al., 2011; Furl et al., 2013, 2015; Pitcher et al., 2011; Schultz et al., 2013; Schultz and Pilz, 2009]. Consistent with previous studies [Foley et al., 2011; Furl et al., 2013; Schultz et al., 2013], we also identified an area that exhibited motion sensitivity to faces in the right posterior STS. This motion-sensitive area pSTS2 partially overlapped with the identified face-selective area pSTS1, and further analysis was conducted on both regions to investigate their facial expression decoding performance. Figure 2 shows the results of group-level analysis of all ROIs ($P < 0.001$, uncorrected); the MNI coordinates, peak intensity, and contrasts used are described in Table I.

Decoding Static and Dynamic Facial Expressions in Face-Selective and Motion-Sensitive Areas

We examined both neural responses and classification accuracies for static and dynamic facial expressions. Figure 3A and B show the results in all 5 ROIs, and the inferential statistical results are summarized in Table II. In the univariate analysis, paired t -test (one-tailed) suggested that all the face-selective and motion-sensitive areas showed greater responses to dynamic relative to static facial expressions, and this increase in response for dynamic over static was more obvious in the pSTS1, pSTS2, and V5f. One further comparison for dynamic expressions showed that responses in the motion-sensitive pSTS2 were significantly higher than those in the face-selective pSTS1 ($t_{(17)} = 3.26$, $P = 0.0023$). In the decoding analysis, we computed the average classification results over the six expressions separately for classifiers that were trained and tested with either static or dynamic items. We found that classification accuracies for the dynamic facial expressions were significantly above chance level in all five ROIs (one-tailed one-sample t -test against chance performance of 0.1667), and paired t -test (one-tailed) showed that all of these regions exhibited significantly higher decoding accuracies for dynamic compared with static facial expressions.

In this part, to further investigate the role of the domain-general motion-sensitive areas in the decoding of facial expressions, we also identified motion-sensitive areas that were not necessarily face specific by contrasting dynamic versus static objects from the localizer. We identified the bilateral STS (right MNI: 60, -33, 12; left MNI: -51, -42, 12) and hMT+/V5 (right MNI: 54, -69, 0; left MNI: -45, -69, 6) and then calculated the neural responses and decoding performance for static and dynamic facial

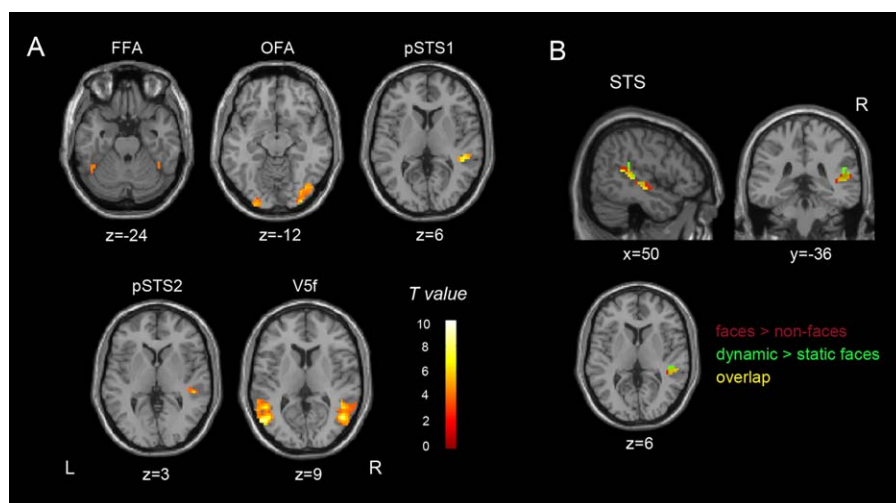


Figure 2.

Group-level ROI analysis. (A) ROI results ($P < 0.001$, uncorrected) from the localizer, showing voxels that were more responsive to faces versus objects and to facial motion. FFA, fusiform face area; OFA, occipital face area; pSTS1, face-selective area in the posterior superior temporal sulcus (pSTS); pSTS2, motion-sensitive area in the pSTS; V5f, motion-sensitive area within

human hMT+/V5. (B) Face selectivity and motion sensitivity to faces overlap in the pSTS ($P < 0.001$, uncorrected). Red represents face-selective voxels; green represents voxels sensitive to facial motion; yellow represents their overlap. [Color figure can be viewed at wileyonlinelibrary.com]

expressions based on these regions. These results are shown in Table III and Figure 4. Consistent with our previous findings, in these not necessarily face-specific motion-sensitive areas, classification accuracies for dynamic expressions were also significantly above chance, and both responses and classification accuracies were significantly higher for dynamic compared with static facial expressions.

In summary, we could robustly decode dynamic facial expressions in both face-selective and motion-sensitive

areas. In all ROIs, we found consistently higher responses and classification accuracies for dynamic than for static facial expressions.

Role of Eye-Related Information in Facial Expression Recognition

To further estimate the role of the eyes in facial expression decoding, obscured stimuli were created by blurring

TABLE I. MNI coordinates, peak intensity, and contrasts used for each ROI

Functional ROIs	MNI coordinates			Peak intensity	Contrast
	x	y	z		
FFA					
L	-45	-51	-24	4.57	Face(d + s) > object(d + s)
R	42	-45	-21	6.05	Face(d + s) > object(d + s)
OFA					
L	-27	-96	-12	5.16	Face(d + s) > object(d + s)
R	42	-75	-15	7.88	Face(d + s) > object(d + s)
pSTS1					
R	48	-36	6	7.58	Face(d + s) > object(d + s)
pSTS2					
R	51	-33	3	5.62	Face(d) > face(s)
V5f					
L	-54	-69	6	9.49	Face(d) > face(s)
R	48	-63	9	9.10	Face(d) > face(s)

FFA, fusiform face area; OFA, occipital face area; pSTS1, face-selective area in the posterior superior temporal sulcus (pSTS); pSTS2, motion-sensitive area in the pSTS; V5f, motion-sensitive area within human hMT+/V5; L, left; R, right; d, dynamic; s, static.

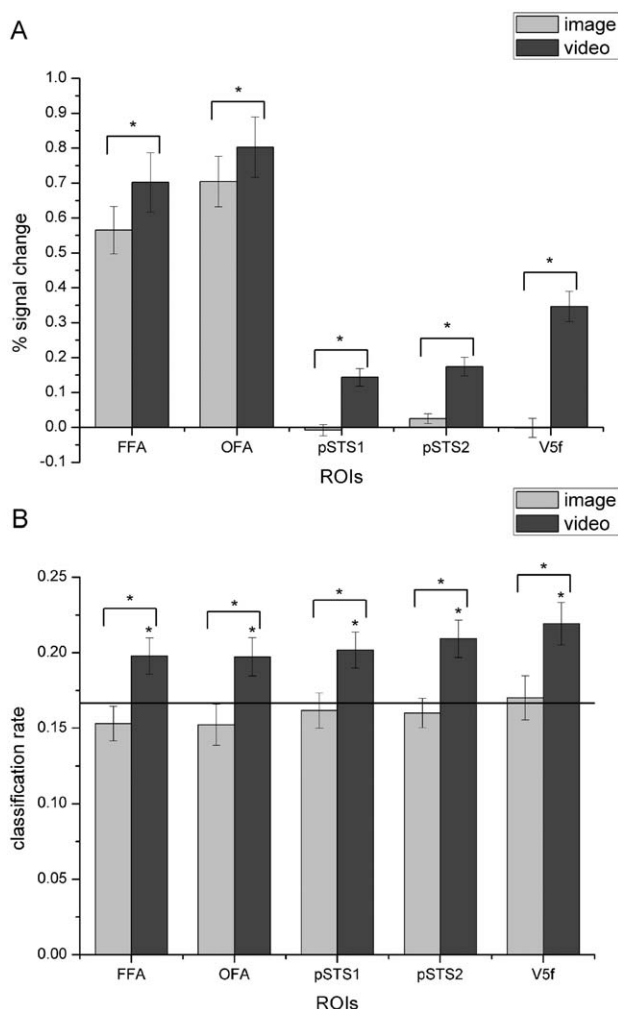


Figure 3.

Average percent signal changes (A) and classification rates (B) for static and dynamic facial expressions of the six basic emotions in each ROI. The black line indicates chance level, and all error bars indicate the SEM. * indicates statistical significance with one-sample or paired *t*-test, $P < 0.05$ (FDR corrected for number of ROIs). In all five ROIs, classification accuracies for dynamic facial expressions were significantly above chance level, and both the activation and classification performance for the dynamic expressions were significantly higher than those for the static facial expressions.

the eye region in the dynamic facial expressions. We examined both the mean fMRI responses and the decoding performance for complete and obscured stimuli. We hypothesized that, because the absence of eye-region information was the only variable that distinguished the complete and obscured expressions, if the participants did not focus on the eyes when recognizing facial expressions (i.e., eye-related information did not play an important role), the response patterns evoked by the obscured stimuli

would not be largely distinct from those evoked by the complete stimuli. A significant difference in response and expression decoding performance would instead suggest that eye-region information played an important role. In the univariate analysis, we compared the percent signal change for complete and obscured stimuli. In addition, in the decoding analysis, we calculated the classification accuracies for the complete and obscured stimuli separately and then compared the accuracies for each stimulus type with chance level and with each other.

The neural responses and classification rates for each ROI are shown in Figure 5A and B, and the inferential statistical results are summarized in Table IV. Paired *t*-test (one-tailed) revealed that compared with those for the complete expressions, the response signals for the obscured expressions were significantly lower in all five regions except V5f. Consistently, all five ROIs exhibited significantly lower classification accuracies for the obscured compared with the complete stimuli.

In summary, when the eye region information was obscured, we found significantly decreased neural responses in all regions except V5f and significantly lower decoding accuracies in all ROIs.

Postscanning Behavioral Results

After scanning, participants made speeded classification of emotional category and rated the emotional intensity for each face stimulus. To maintain continuity across the results, we included behavioral data for the 18 participants who were included in the previous decoding analysis. We compared classification accuracies, intensity ratings and reaction times between the static and dynamic expressions and between the complete and obscured expressions to examine potential differences. Figure 6 shows the behavioral results, and the statistical results of the comparisons are summarized in Table V. We found that the participants showed higher classification accuracies for dynamic compared with static expressions and that the intensity of the emotion they perceived significantly decreased when information about the eyes was missing. For reaction times, there were no significant differences. In the questionnaires conducted after the experiment, the participants also reported that it was much easier to recognize the expressions in the dynamic stimuli compared with the static and obscured stimuli.

In sum, the behavioral measures were consistent with the fMRI results, showing that facial motion cues facilitate expression recognition and that eye-related information plays an important role in the recognition of facial expressions.

DISCUSSION

The main purpose of this study was to explore the roles of face-selective and motion-sensitive areas in the

TABLE II. Statistical results of the analysis of response and decoding accuracy of static and dynamic facial expressions in each ROI

	%signal change		Decoding accuracy					
	Video > Image		Image > Chance		Video > Chance		Video > Image	
	$t_{(17)}$	P	$t_{(17)}$	P	$t_{(17)}$	P	$t_{(17)}$	P
FFA	4.853	<0.001*	-1.186	0.847	2.595	0.009*	2.396	0.014*
OFA	3.487	0.001*	-1.050	0.846	2.401	0.014*	2.603	0.009*
pSTS1	8.021	<0.001*	-0.430	0.664	2.954	0.004*	2.116	0.025*
pSTS2	7.342	<0.001*	-0.671	0.744	3.453	0.002*	2.819	0.006*
V5f	14.384	<0.001*	0.237	0.408	3.739	<0.001*	2.504	0.011*

One-sample and paired t -test of percent signal change and decoding accuracies for static and dynamic facial expressions in each ROI; one-tailed. *Significant P values (FDR corrected for number of ROIs). In all five ROIs, classification accuracies for dynamic facial expressions were significantly above chance; both the activation and classification performance were significantly higher for the dynamic expressions than for the static facial expressions.

recognition of facial expressions. To address this issue, we employed a block design experiment and conducted univariate analysis and MVPA. We also examined the impact of facial motion and eye-region information on facial expression decoding by employing multiple types of expression stimuli. Our results showed that motion-sensitive areas also responded to expressions and that the dynamic facial expressions could be successfully decoded in both face-selective and motion-sensitive areas. We found that the dynamic expressions, which are most commonly experienced in the natural environment, elicited higher responses and classification accuracies than static expressions in all ROIs. We also observed a significant decrease in both responses and decoding accuracies due to the absence of eye-related information.

Motion-Sensitive Areas Represent Expression Information in Addition to Conventional Face-Selective Areas

We examined the fMRI responses and the decoding performance for facial expressions in both face-selective and motion-sensitive areas. Our analysis suggests that

expression information is represented not only in face-selective but also in motion-sensitive areas.

Neuroimaging studies of facial expression perception have focused on face-selective areas, especially the core face network, which consists of the OFA for early face perception [Rotshtein et al., 2005] and for the processing of structural changes in faces [Fox et al., 2009b], the FFA for the processing of facial features and identity [Fox et al., 2009b; Ganel et al., 2005] and the pSTS for the processing of changeable features such as gaze or emotional expressions [Hoffman and Haxby, 2000]. Previous fMRI studies have showed that these face-selective areas are involved in facial expression processing and exhibit stronger responses to dynamic than static stimuli [Foley et al., 2011; Fox et al., 2009a; Lee et al., 2010; Pitcher et al., 2011; Sato et al., 2004; Schultz and Pilz, 2009]. Our results are consistent with these findings and in addition to the higher responses found in previous studies, further revealed significantly higher decoding accuracies for dynamic than static facial expressions through MVPA. Together, these findings confirm the importance of face-selective areas, which encode expression information, in facial expression recognition.

Moreover, our study suggests a role for motion-sensitive areas in the recognition of facial expressions. A number of recent studies have found that motion-sensitive areas

TABLE III. Statistical results of the analysis of response and decoding accuracy of static and dynamic facial expressions in object-defined motion-sensitive areas

	%Signal change		Decoding accuracy					
	Video > Image		Image > Chance		Video > Chance		Video > Image	
	$t_{(17)}$	P	$t_{(17)}$	P	$t_{(17)}$	P	$t_{(17)}$	P
STS	6.457	<0.001*	0.385	0.353	3.589	0.001*	2.123	0.024*
hMT+ /V5	13.345	<0.001*	0.347	0.366	3.393	0.002*	2.300	0.017*

One-sample and paired t -test of percent signal change and decoding accuracies for static and dynamic facial expressions in object-defined ROIs; one-tailed. *Significant P values (FDR corrected for number of ROIs). In both object-defined motion-sensitive ROIs, classification accuracies for dynamic expressions were significantly above chance, and both activation and classification performance were significantly higher for dynamic compared with static facial expressions.

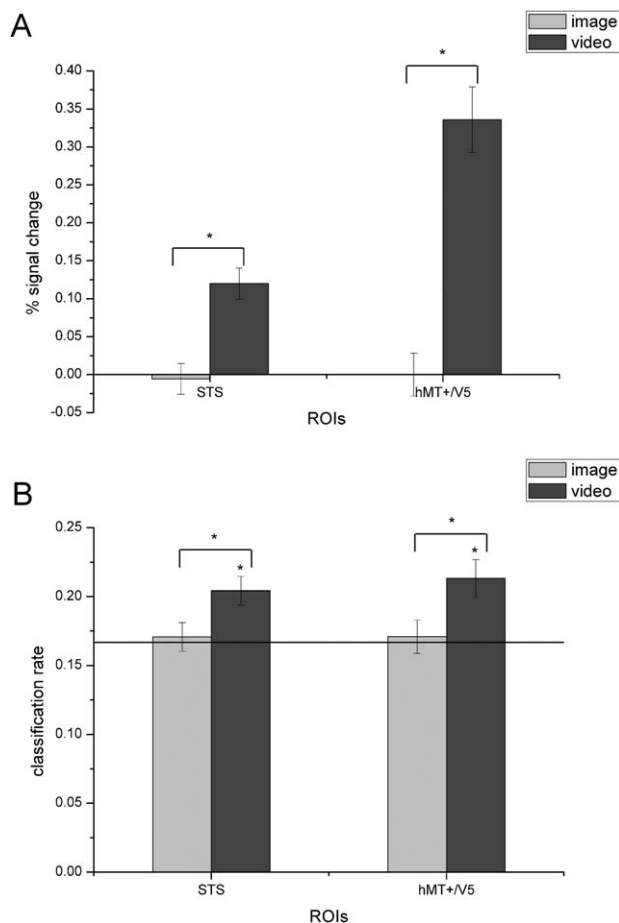


Figure 4.

Average percent signal changes (A) and classification rates (B) for static and dynamic facial expressions in object-defined motion-sensitive ROIs. The black line indicates chance level, and all error bars indicate the SEM. * indicates statistical significance with one-sample or paired *t*-test, $P < 0.05$ (FDR corrected for number of ROIs). In both object-defined ROIs, classification accuracies for dynamic facial expressions were significantly above chance level, and both the activation and classification performance were significantly higher for the dynamic compared with the static facial expressions.

within pSTS and human motion complex showed strong responses to dynamic facial stimuli [Foley et al., 2011; Furl et al., 2013; Furl et al., 2015; Pitcher et al., 2011; Schultz et al., 2013; Schultz and Pilz, 2009]. By employing contrast identical to that used in previous studies [Furl et al., 2012, 2013, 2015], we identified these motion-sensitive areas that responded to facial motion. Our results were not only in line with previous findings but also further revealed the successful decoding of dynamic facial expressions in these motion-sensitive areas. Furthermore, the results for the object-defined motion-sensitive areas, which were not necessarily face specific, provided additional support for our

analysis and revealed the expression decoding performance of relatively domain-general motion-sensitive areas. Taken together, our results suggest that expression information need not be represented exclusively by face-selective areas that encode facial attributes and reflect specialization for face selectivity; relatively domain-general motion-sensitive areas, which respond to various visual motions and are not necessarily face specific, may also effectively contribute to human facial expression recognition. Our findings are consistent with observations in macaque [Furl et al., 2012] that suggest that macaque motion-sensitive areas also represent expression information. Moreover, we also detected overlapping face-selective

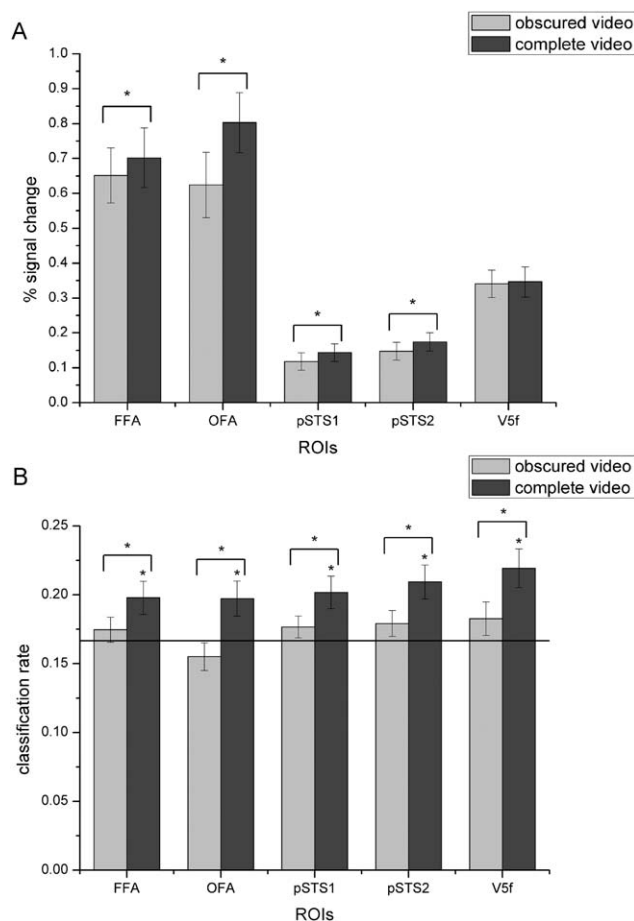


Figure 5.

Average percent signal changes (A) and classification rates (B) for the obscured compared with the complete expressions in all ROIs. The black line indicates chance level, and all error bars indicate the SEM. * indicates statistical significance with one-sample or paired *t*-test, $P < 0.05$ (FDR corrected for number of ROIs). Activations for the obscured expressions were significantly lower than those for the complete expressions in all ROIs except V5f, and in all ROIs, the classification accuracies for the obscured expressions were significantly reduced.

TABLE IV. Statistical results of the analysis of response and decoding accuracy of complete and eyes-obscured expressions in each ROI

	%Signal change		Decoding accuracy					
	Complete > Obscured		Complete > Chance		Obscured > Chance		Complete > Obscured	
	$t_{(17)}$	P	$t_{(17)}$	P	$t_{(17)}$	P	$t_{(17)}$	P
FFA	2.215	0.020*	2.595	0.009*	0.880	0.196	2.647	0.008*
OFA	4.643	<0.001*	2.401	0.014*	-1.158	0.869	2.800	0.006*
pSTS1	2.236	0.019*	2.954	0.004*	1.567	0.068	2.277	0.018*
pSTS2	2.188	0.021*	3.453	0.002*	1.668	0.057	2.114	0.025*
V5f	0.403	0.346	3.739	<0.001*	1.719	0.052	2.113	0.025*

One-sample and paired t -test of percent signal change and decoding accuracies for complete and eye-region obscured expressions in each ROI; one-tailed. *Significant P values (FDR corrected for number of ROIs). Activations for the obscured expressions were significantly lower than those for the complete expressions in all ROIs except V5f, and the classification accuracies for the obscured expressions were significantly reduced in all ROIs.

and motion-sensitive voxels in the pSTS [Furl et al., 2013], revealing that pSTS could represent both facial attributes and facial motion information in the processing of facial expressions. Together with the successful decoding of expression information, our study further corroborates the key role of the pSTS in facial expression recognition.

Facial Motion Facilitates the Decoding of Facial Expressions

We compared neural responses and decoding performance between static and dynamic facial expressions to explore the impact of facial motion in the recognition of facial expressions. We found consistent enhancement of activation and decoding performance for dynamic compared with static facial expressions in all ROIs. In social communication, facial expressions are dynamic; moving depictions of facial expression are more ecologically valid than their static counterparts [Arsalidou et al., 2011; Fox et al., 2009a; Sato et al., 2004; Schultz and Pilz, 2009; Trautmann et al., 2009]. Behavioral studies have also found an advantage in the recognition of dynamic versus static facial expressions [Ambadar et al., 2005; Cunningham and Wallraven, 2009; Wehrle et al., 2000], and fMRI studies have demonstrated enhanced activation patterns for dynamic versus static facial stimuli [Arsalidou et al., 2011; Fox et al., 2009a; Lee et al., 2010; Pitcher et al., 2011; Sato et al., 2004; Schultz et al., 2013; Schultz and Pilz, 2009; Trautmann et al., 2009]. Our results were completely consistent with these findings, revealing increased responses to dynamic relative to static facial expressions due to facial motion. Moreover, it has been suggested that dynamic facial expressions might provide some form of information that is available only over time, which could facilitate the facial expression recognition [Cunningham and Wallraven, 2009]. The results of our decoding analysis revealed the important role of facial motion in the decoding of facial expressions and thus provide additional support for this

idea. We found successful decoding of dynamic facial expressions and significantly higher decoding accuracies for dynamic compared with static expressions. Combined with previous findings, our results suggest that motion-related cues transmit expression information that facilitates facial expression decoding.

Eye-Region Information Plays an Important Role in Facial Expression Recognition

A number of previous behavior, eye movement and electrophysiological studies have indicated that the eyes play an important role in face perception [Jack et al., 2009, 2012; Kret et al., 2013; Nemrodov et al., 2014; Yuki et al., 2007]. Moreover, Asian observers were found to prefer eye-related expressive information, as they persistently fixated the eye region and strongly focused on the eyes when categorizing others' facial expressions [Jack et al., 2009, 2012; Yuki et al., 2007]. We obtained compatible findings through our analysis. Compared with complete expressions, when the eye region was obscured, univariate analysis showed significantly decreased neural responses in all regions except V5f, and MVPA revealed significantly reduced decoding accuracies in all regions. These results suggest that the response patterns evoked by the obscured stimuli were largely distinct from those evoked by the complete facial expressions. Neural response significantly decreased and the available information for facial expression decoding contained therein significantly reduced due to the absence of eye-related information. Although the decrease in response signals for obscured stimuli was not significant in V5f, possibly because this region is located within the human motion complex that is sensitive to visual motion [Furl et al., 2013, 2015; Schultz et al., 2013], significantly reduced decoding accuracies were observed, suggesting that the absence of eye-region information caused a significant decrease of information that available for facial expression decoding even when the motion

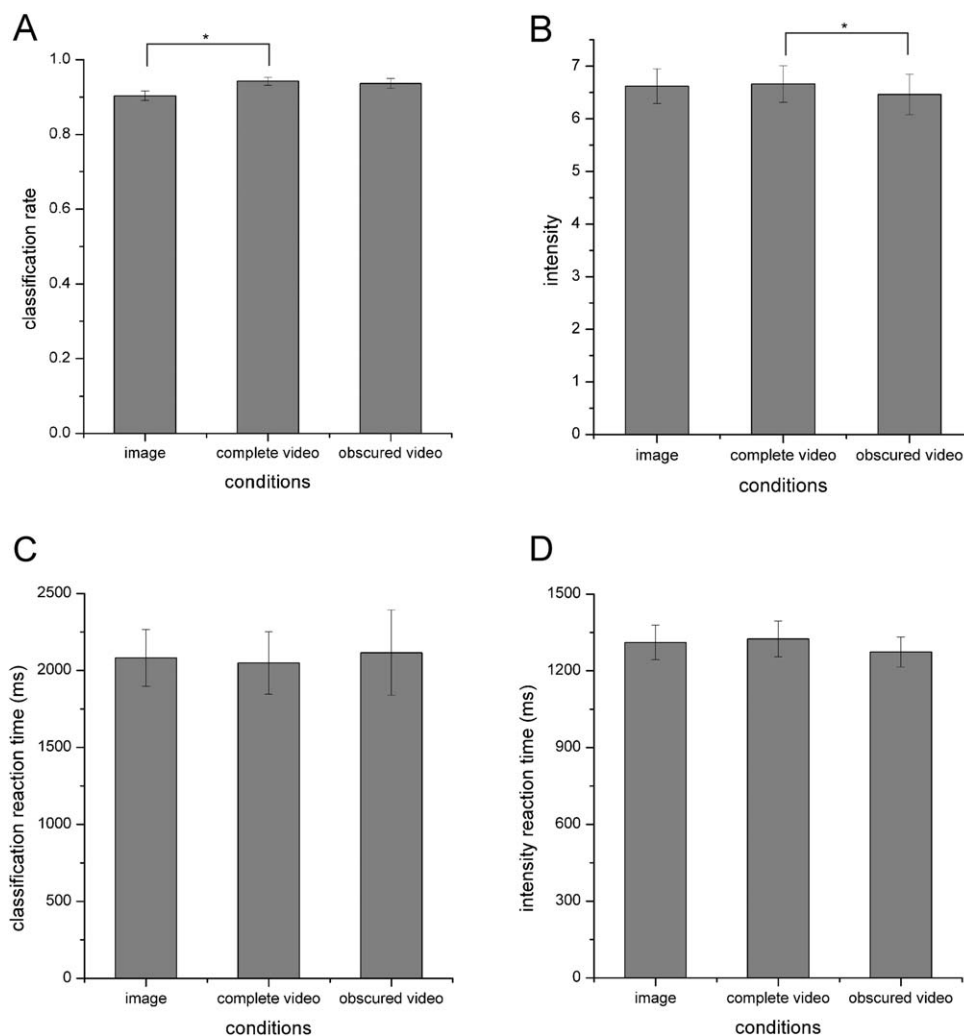


Figure 6.

Postscanning behavioral performance. (A) classification rates, (B) perceived emotional intensity, (C) reaction times for expression classification, and (D) reaction times for emotional intensity ratings. All error bars indicate the SEM. * indicates statistical significance with paired t -test, $P < 0.05$.

information contained in the obscured stimuli might not have significantly differed from that contained in the complete stimuli. To summarize, our results suggest that the eye region, which conveys expression information preferred by Asians, plays an important role in facial expression recognition.

In our current study, due to the number of conditions and expressions under investigation and taking into account scanning time and the participants' comfort, there were 48 samples for each expression in each condition. The inclusion of additional samples for each condition could further improve the implementation of the classification scheme and boost the accuracy; this possibility will be further examined in future studies. In addition, future studies with more samples for each expression could be

TABLE V. Postscanning behavioral results

	Static–Dynamic		Complete–Obscured	
	$t_{(17)}$	P	$t_{(17)}$	P
Accuracy	-3.265	0.002*	0.524	0.304
Intensity	-0.508	0.309	3.339	0.002*
Accuracy.RT	0.339	0.370	-0.717	0.165
Intensity.RT	-0.267	0.604	1.282	0.891

Paired t -test of classification accuracies, emotional intensity ratings, and corresponding reaction times between static and dynamic expressions and between complete and obscured expressions; one-tailed. *Significant P values.

used to further explore the patterns of activation in response to the individual emotional expressions and to investigate whether expressions could be reconstructed from specific patterns of activation within brain regions involved in facial expression recognition. Studies of individual expression reconstruction may help to further analyze the contribution of each ROI to the successful discrimination of different facial expressions.

CONCLUSION

In summary, we show that expression information is represented not only in face-selective but also in motion-sensitive areas. Our results substantiate the importance of face-selective areas in facial expression recognition and further suggest that relatively domain-general motion-sensitive areas, which are not specialized for representing only facial attributes, may also contribute to everyday expression recognition. Moreover, we show that motion-related cues transmit measurable quantities of expression information that can facilitate facial expression decoding. Consistent with the demonstrated preference of eye-related expressive information in Asians, we also suggest that eye-related information plays an important role in the recognition of facial expressions. Our study extends past research of facial expression recognition and may lead to better understanding of how human beings achieve quick and easy recognition of others' facial expressions in their everyday lives.

ACKNOWLEDGMENTS

We would like to thank Prof. Fox (University of British Columbia) for supplying the dynamic object stimuli used in the article: Defining the Face Processing Network: Optimization of the Functional Localizer in fMRI, *Hum Brain Mapp*, 2009.

CONFLICT OF INTEREST

We declare that we have no actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations that can inappropriately influence our work.

REFERENCES

- Adolphs R (2002): Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behav Cogn Neurosci Rev* 1:21–62.
- Allison T, Puce A, McCarthy G (2000): Social perception from visual cues: Role of the STS region. *Trends Cogn Sci* 4:267–278.
- Ambadar Z, Schooler JW, Cohn JF (2005): Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychol Sci* 16:403–410.
- Andrews TJ, Ewbank MP (2004): Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage* 23:905–913.
- Arsalidou M, Morris D, Taylor MJ (2011): Converging evidence for the advantage of dynamic facial expressions. *Brain Topogr* 24:149–163.
- Axelrod V, Yovel G (2012): Hierarchical processing of face view point in human visual cortex. *J Neurosci* 32:2442–2452.
- Brett M, Anton J, Valabregue R, Poline J (2002): Region of interest analysis using an SPM toolbox. *Neuroimage* 16:1140–1141.
- Cunningham DW, Wallraven C (2009): Dynamic information for the recognition of conversational expressions. *J Vis* 9:1–17.
- Engell AD, McCarthy G (2014): Face, eye, and body selective responses in fusiform gyrus and adjacent cortex: An intracranial EEG study. *Front Hum Neurosci* 8:642.
- Foley E, Rippon G, Thai NJ, Longe O, Senior C (2011): Dynamic facial expressions evoke distinct activation in the face perception network: A connectivity analysis study. *J Cogn Neurosci* 24:507–520.
- Fox CJ, Iaria G, Barton JJS (2009a): Defining the face processing network: Optimization of the functional localizer in fMRI. *Hum Brain Mapp* 30:1637–1651.
- Fox CJ, Moon SY, Iaria G, Barton JJS (2009b): The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage* 44:569–580.
- Furl N, Hadj-Bouziane F, Liu N, Averbeck BB, Ungerleider LG (2012): Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J Neurosci* 32:15952–15962.
- Furl N, Henson RN, Friston KJ, Calder AJ (2013): Top-down control of visual responses to fear by the amygdala. *J Neurosci* 33:17435–17443.
- Furl N, Henson RN, Friston KJ, Calder AJ (2015): Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cereb Cortex* 25:2876–2882.
- Ganel T, Valyear KF, Goshen-Gottstein Y, Goodale MA (2005): The involvement of the “fusiform face area” in processing facial expression. *Neuropsychologia* 43:1645–1654.
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000): The fusiform “face area” is part of a network that processes faces at the individual level. *J Cogn Neurosci* 12:495–504.
- Gobbini MI, Haxby JV (2007): Neural systems for recognition of familiar faces. *Neuropsychologia* 45:32–41.
- Gobbini MI, Gentili C, Ricciardi E, Bellucci C, Salvini P, Laschi C, Guazzelli M, Pietrini P (2011): Distinct neural systems involved in agency and animacy detection. *J Cogn Neurosci* 23:1911–1920.
- Grill-Spector K, Knouf N, Kanwisher N (2004): The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci* 7:555–562.
- Gur RC, Schroeder L, Turner T, McGrath C, Chan RM, Turetsky BI, Alsop D, Maldjian J, Gur RE (2002): Brain activation during facial emotion processing. *Neuroimage* 16:651–662.
- Harry B, Williams MA, Davis C, Kim J (2013): Emotional expressions evoke a differential response in the fusiform face area. *Front Hum Neurosci* 7:692.
- Harwood NK, Hall LJ, Shinkfield AJ (1999): Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *Am J Ment Retard* 104:270–278.
- Haxby JV, Hoffman EA, Gobbini MI (2000): The distributed human neural system for face perception. *Trends Cogn Sci* 4:223–233.
- Hoffman EA, Haxby JV (2000): Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci* 3:80–84.

- Ishai A, Schmidt CF, Boesiger P (2005): Face perception is mediated by a distributed cortical network. *Brain Res Bull* 67:87–93.
- Jack RE, Blais C, Scheepers C, Schyns PG, Caldara R (2009): Cultural confusions show that facial expressions are not universal. *Curr Biol* 19:1543–1548.
- Jack RE, Caldara R, Schyns PG (2012): Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *J Exp Psychol Gen* 141:19–25.
- Johnston P, Mayes A, Hughes M, Young AW (2013): Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex* 49:2462–2472.
- Kawasaki H, Tsuchiya N, Kovach CK, Nourski KV, Oya H, Howard MA, Adolphs R (2012): Processing of facial emotion in the human fusiform gyrus. *J Cogn Neurosci* 24:1358–1370.
- Kret ME, Stekelenburg JJ, Roelofs K, de Gelder B (2013): Perception of face and body expressions using electromyography, pupillometry and gaze measures. *Front Psychol* 4:28.
- Lee LC, Andrews TJ, Johnson SJ, Woods W, Gouws A, Green GGR, Young AW (2010): Neural responses to rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG. *Neuropsychologia* 48:477–490.
- Mahmoudi A, Takerkart S, Regragui F, Boussaoud D, Brovelli A (2012): Multivoxel pattern analysis for FMRI data: A review. *Comput Math Methods Med* 2012:1817–1847.
- Monroe JF, Griffin M, Pinkham A, Loughhead J, Gur RC, Roberts TP, Christopher Edgar J (2013): The fusiform response to faces: Explicit versus implicit processing of emotion. *Hum Brain Mapp* 34:1–11.
- Murphy FC, Nimmo-Smith I, Lawrence AD (2003): Functional neuroanatomy of emotions: A meta-analysis. *Cogn Affect Behav Neurosci* 3:207–233.
- Nemrodov D, Anderson T, Preston FF, Itier RJ (2014): Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *Neuroimage* 97:81–94.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006): Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
- Phan KL, Wager T, Taylor SF, Liberzon I (2002): Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16:331–348.
- Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011): Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* 56:2356–2363.
- Rotshtein P, Henson RNA, Treves A, Driver J, Dolan RJ (2005): Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci* 8:107–113.
- Said CP, Moore CD, Engell AD, Todorov A, Haxby JV (2010): Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis* 10:71–76.
- Sato W, Kochiyama T, Yoshikawa S, Naito E, Matsumura M (2004): Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Brain Res Cogn Brain Res* 20:81–91.
- Schultz J, Pilz KS (2009): Natural facial motion enhances cortical responses to faces. *Exp Brain Res* 194:465–475.
- Schultz J, Brockhaus M, Bühlhoff HH, Pilz KS (2013): What the human brain likes about facial motion. *Cereb Cortex* 23:1167–1178.
- Song S, Zhan Z, Long Z, Zhang J, Yao L (2011): Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PLoS One* 6:e17191.
- Stock JVD, Tamietto M, Sorger B, Pichon S, Grézes J, de Gelder B (2011): Cortico-subcortical visual, somatosensory, and motor activations for perceiving dynamic whole-body emotional expressions with and without striate cortex (V1). *Proc Natl Acad Sci USA* 108:16188–16193.
- Tamietto M, Castell L, Vighetti S, Perozzo P, Geminiani G, Weiskrantz L, de Gelder B (2009): Unseen facial and bodily expressions trigger fast emotional reactions. *Proc Natl Acad Sci USA* 106:17661–17666.
- Trautmann SA, Fehr T, Herrmann M (2009): Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res* 1284:100–115.
- Van der Schalk J, Hawk ST, Fischer AH, Doosje B (2011): Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion* 11:907–920.
- Wegrzyn M, Riehle M, Labudda K, Woermann F, Baumgartner F, Pollmann S, Bien CG, Kissler J (2015): Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* 69:131–140.
- Wehrle T, Kaiser S, Schmidt S, Scherer KR (2000): Studying the dynamics of emotional expression using synthesized facial muscle movements. *J Pers Soc Psychol* 78:105–119.
- Williams MA, Dang S, Kanwisher NG (2007): Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685–686.
- Winston JS, Henson RNA, Fine-Goulden MR, Dolan RJ (2004): fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J Neurophysiol* 92:1830–1839.
- Wolbers T, Zahorik P, Giudice NA (2011): Decoding the direction of auditory motion in blind humans. *Neuroimage* 56:681–687.
- Yovel G, Kanwisher N (2004): Face perception: Domain specific, not process specific. *Neuron* 44:889–898.
- Yuki M, Maddux WW, Masuda T (2007): Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J Exp Soc Psychol* 43:303–311.