

Dynamic and Static Facial Expressions Decoded from Motion-Sensitive Areas in the Macaque Monkey

Nicholas Furl,^{1*} Fadila Hadj-Bouziane,^{2*} Ning Liu,² Bruno B. Averbeck,^{1†} and Leslie G. Ungerleider^{2†}

Laboratories of ¹Neuropsychology and ²Brain and Cognition, NIMH/NIH, Bethesda, Maryland 20892

Humans adeptly use visual motion to recognize socially relevant facial information. The macaque provides a model visual system for studying neural coding of expression movements, as its superior temporal sulcus (STS) possesses brain areas selective for faces and areas sensitive to visual motion. We used functional magnetic resonance imaging and facial stimuli to localize motion-sensitive areas [motion in faces (Mf) areas], which responded more to dynamic faces compared with static faces, and face-selective areas, which responded selectively to faces compared with objects and places. Using multivariate analysis, we found that information about both dynamic and static facial expressions could be robustly decoded from Mf areas. By contrast, face-selective areas exhibited relatively less facial expression information. Classifiers trained with expressions from one motion type (dynamic or static) showed poor generalization to the other motion type, suggesting that Mf areas employ separate and nonconfusable neural codes for dynamic and static presentations of the same expressions. We also show that some of the motion sensitivity elicited by facial stimuli was not specific to faces but could also be elicited by moving dots, particularly in fundus of the superior temporal and middle superior temporal polysensory/lower superior temporal areas, confirming their already well established low-level motion sensitivity. A different pattern was found in anterior STS, which responded more to dynamic than to static faces but was not sensitive to dot motion. Overall, we show that emotional expressions are mostly represented outside of face-selective cortex, in areas sensitive to motion. These regions may play a fundamental role in enhancing recognition of facial expression despite the complex stimulus changes associated with motion.

Introduction

Humans and other primates depend on facial expressions for social interaction. However, their visual systems must cope with a difficult computational challenge: they must extract information about facial expressions despite complex naturalistic movements. Nevertheless, abundant evidence shows that motion enhances recognition of facial identity and expression (Knight and Johnston, 1997; Lander et al., 1999; Wehrle et al., 2000; O'Toole et al., 2002; Knappmeyer et al., 2003; Ambadar et al., 2005; Roark et al., 2006; Lander and Davies, 2007; Trautmann et al., 2009). We investigated the neural computations which might support this feat, using the macaque superior temporal sulcus (STS) as a model system. In the macaque, electrophysiological and func-

tional magnetic resonance imaging (fMRI) studies have localized candidate areas that could encode facial movements. Much attention has focused on face-selective areas (“patches”), which respond more to faces than to nonface objects (Tsao et al., 2006) and encode many facial attributes (Freiwald et al., 2009). However, it is uncertain to what extent face-selective representations incorporate information about facial expressions (Hadj-Bouziane et al., 2008) or facial movements.

Other areas in the macaque STS might also participate in the representation of facial expression movements. These areas include those sensitive to “low-level” motion (e.g., moving dots, gratings, lines, etc.), such as the well characterized middle temporal (MT/V5) area (Dubner and Zeki, 1971), the medial superior temporal (MST) area (Desimone and Ungerleider, 1986), and the fundus of the superior temporal (FST) area, which can all be detected as discrete areas using fMRI (Vanduffel et al., 2001). Beyond these regions, in middle STS, there are also neurons sensitive to low-level and biological motion (Bruce et al., 1981; Vangeneugden et al., 2011) as well as to static presentations of implied biological motion (Barraclough et al., 2006). These neurons likely populate the discrete functional areas identified using fMRI, including the middle superior temporal polysensory (STPm) area and the lower superior temporal (LST) area (Nelissen et al., 2006). More recently, fMRI revealed a rostral region in the fundus of STS, sensitive to dynamic grasping actions (Nelissen et al., 2011). This area has not been previously reported using low-level motion stimuli and it is unknown whether it is also sensitive to facial motion or whether it encodes facial expressions.

The macaque STS also contains neurons selective for individual facial expressions (Hasselmo et al., 1989), although their re-

Received April 24, 2012; revised Aug. 3, 2012; accepted Sept. 6, 2012.

Author contributions: F.H.-B. and L.U. designed research; F.H.-B. and N.L. performed research; N.F. and F.H.-B. analyzed data; N.F., B.A., and L.U. wrote the paper.

This work was supported by funding from the National Institute of Mental Health Intramural Research Program to N.F., B.A., F.H.-B., N.L., and L.G.U. and funding from the United Kingdom Economic and Social Research Council (RES-062-23-2925) to N.F. We are grateful to our colleagues at the National Institute of Mental Health: Fern Baldwin and Lucas Glover for their help in training the animals; Jennifer Frihauf and Kathleen Hansen for their help with stimulus preparation; Ziad Saad and Maria Barsky for their help with data analysis; Frank Ye, Charles Zhu, Neal Phelps, and Wenming Luh for their assistance during scanning, and George Dold, David Ide, and Tom Talbot for technical assistance. Editorial assistance for the article was provided by the National Institutes of Health Fellows Editorial Board.

*N.F. and F.H.-B. contributed equally and are co-first authors, and †B.A. and L.G.U. contributed equally and are co-senior authors.

This article is freely available online through the *JNeurosci* Open Choice option.

Correspondence should be addressed to Nicholas Furl at his present address: Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK. E-mail: nick.furl@mrc-cbu.cam.ac.uk.

DOI:10.1523/JNEUROSCI.1992-12.2012

Copyright © 2012 the authors 0270-6474/12/3215952-11\$15.00/0

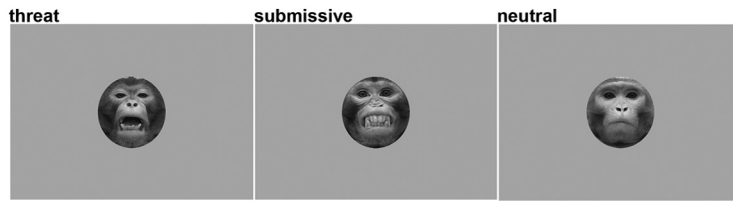


Figure 1. Facial expressions. Sample static images from the visual motion in faces runs showing threat, submissive, and neutral expressions for the three monkey identities. These images were also frames from the dynamic expression videos.

relationship to motion-sensitive and face-selective areas is unknown. Nevertheless, these neurons could give rise to distributed fMRI response patterns detectable using multivariate decoding analysis. We used fMRI to localize areas responsive to visual motion using facial stimuli [motion in faces (Mf) areas] and then decoded expression information from their fMRI response patterns. We also localized face-selective areas, which responded more to faces than to places and objects, and therein quantified expression information. We further tested whether responses in Mf and face-selective areas were sensitive to low-level, dot motion. We hypothesized that, because motion-related cues can facilitate expression recognition, areas sensitive to motion in faces and dots would transmit measurable quantities of expression information.

Materials and Methods

Subjects and training. Three male macaque monkeys were used (*Macaca mulatta*, 6–8 kg). All procedures were in accordance with the Guide for the Care and Use of Laboratory Animals, were approved by the NIMH Animal Care and Use Committee and conformed to all NIH guidelines. Each animal was implanted with a plastic head post under anesthesia and aseptic conditions. After recovery, monkeys were trained to sit in a sphinx position in a plastic restraint barrel (Applied Prototype) with their heads fixed, facing a screen on which visual stimuli were presented. During MR scanning, gaze location was monitored using an infrared pupil tracking system (ISCAN).

Stimuli and task. Stimuli were presented using Presentation (Neurobehavioral Systems, www.neurobs.com), and displayed via an LCD projector (Sharp NoteVision 3 in the 3 T scanner or Avotec Silent Vision SV-6011-2 in the 4.7 T scanner) onto a front-projection screen positioned within the magnet bore. In the 4.7 T scanner, this screen was viewed via a mirror. Throughout all scanning runs, stimuli were overlaid with a 0.2° centrally located fixation spot, on which the monkeys were required to fixate to receive a liquid reward. In the reward schedule, the frequency of reward increased as the duration of fixation increased, with reward delivery occurring at any possible time during a trial.

Two of the monkeys (1 and 2) participated in all three types of scanning runs, all of which implemented block designs to localize respectively: (1) face-selective areas; (2) areas sensitive to motion using facial stimuli (Mf areas) and; (3) areas sensitive to motion using dot stimuli [motion in dot (Md) areas]. Monkey 3 participated in the first two types of scanning runs listed above. Across all these runs, the order of the blocks was counterbalanced.

In each face-selectivity run (Monkey 1: 20 runs; Monkey 2: 18 runs; Monkey 3: 24 runs), there were three blocks respectively devoted to macaque faces, nonface objects, or places. All stimuli were grayscale static photographs and were familiar to the three monkeys. Within each block of faces, the faces reflected 17 possible facial identities, presented in a random order. All expressions were neutral and all faces were frontal view. All blocks in the face-selectivity runs lasted 40 s, during which 20 images (11° wide) were presented for 2 s each. Each block was followed with a period of 20 s blank (gray background).

In each Mf run (Monkey 1: 34 runs; Monkey 2: 18 runs; Monkey 3: 39 runs), there were six different blocks, devoted to frontally viewed dynamic or static presentations, depicting one of three expressions (Fig. 1). Threat expressions were defined as aggressive, open-mouthed postures

with directed gaze. Submissive expressions were fearful or appeasing gestures including mixtures of lip smacks and fear grins. We also included neutral expression blocks. All face stimuli were embedded in a gray oval mask and included three macaque identities that were familiar to the monkeys (Fig. 1). In each 36 s block, 18 presentations from one of the six categories appeared for 2 s each. Each block was followed with a period of 20 s blank.

Monkeys 1 and 2 participated in the Md runs, which implemented a similar counterbalanced block design. In each run (Monkey 1: 20 runs; Monkey 2: 32 runs), each of the four blocks was devoted to one condition: static random dots, translating random dots, static optic flow (radiating dots), and expanding/contracting optic flow. Each block consisted of 20 2 s stimulus presentations with 20 s of fixation following each block. The stimuli used were white random dots (diameter, 0.2°), forming a circular aperture (diameter, 8°) in fully coherent motion (speed, 2°/s) on a black background. To maintain a constant dot density, each dot that left the aperture reentered from the other side at a random location.

Scanning. Before each scan session, the exogenous contrast agent monocrystalline iron oxide nanocolloid (MION) was injected into the saphenous vein (10–12 mg/kg) to increase the contrast-to-noise ratio and to optimize the localization of fMRI signals. Face-selectivity and Mf runs were collected using a 3 tesla General Electric MRI scanner and an 8-loop surface coil (RAPID Biomedical). Functional data for these runs were obtained using a gradient echo sequence (EPI) and SENSE (factor of 2), TR = 2 s, TE = 17.9 ms, flip angle = 90°, field of view (FOV) = 100 mm, matrix = 64 × 64 voxels, slice thickness = 1.5 mm, 27 coronal slices (no gap). The slice package included most of the temporal lobe beginning in posterior STS (just anterior to area MT/MST) and extending anteriorly, covering TE and TEO, the amygdala, and most of the frontal lobe. In the coordinate space of the Saleem and Logothetis (2007) stereotaxic atlas, this coverage spanned from $-y = -1$ to $+43$. Monkeys 1 and 2 also participated in Md runs, acquired on a 4.7 tesla Bruker MRI scanner [EPI, TR = 2 s, TE = 13 ms, flip angle = 90°, FOV = 96 × 44 mm, matrix = 64 × 32 voxels, slice thickness = 2 mm, 25 coronal slices (no gap)]. This slice package included the whole brain. In separate sessions, we also acquired at 4.7 T, high-resolution anatomical scans from each monkey under anesthesia (3D MPRAGE, TR = 2.5 s, TE = 4.35 ms, flip angle = 8°, matrix = 384 × 384 voxels, voxel size = 0.35 mm isotropic).

Preprocessing and first-level general linear model analyses. We analyzed MRI data using MATLAB (The MathWorks), SPM8 (Wellcome Trust Centre for Neuroimaging, London; <http://www.fil.ion.ucl.ac.uk/spm/>), CARET (Van Essen et al., 2001), and AFNI (Cox, 1996). All types of runs were motion-corrected using six-parameter rigid body realignment. For purposes of localizing the functional areas, the fMRI data were smoothed to 2.0 mm³ full-width half maximum.

We performed separate “first-level” fixed effects general linear models (GLMs) for face-selectivity, Mf and Md runs. For all GLMs, each block was treated as a separate regressor with a single event, which was then convolved with a canonical MION function (Leite et al., 2002). We computed contrasts of interest in each monkey using these first-level regressors. For face-selectivity runs, we identified face-selective areas by comparing all face blocks versus all nonface blocks. For Mf runs, we identified Mf areas by comparing all dynamic face blocks versus all static face blocks (See Results for a motivation of this contrast). For the Md runs, we identified Md areas by comparing all moving dots blocks versus all static dots blocks.

Visualization and localization of functional areas. Using AFNI, we then computed a spatial normalization (Saad et al., 2009) of the functional data for each monkey to a population-average MRI-based atlas collection for the Rhesus macaque (McLaren et al., 2009). This MRI-based template image was previously normalized to the Saleem and Logothetis (2007) stereotaxic atlas. Thus, our normalization procedure allowed us to project our statistical results into this standardized coordinate space, and thereby derive coordinates in a common space for the peak activation locations of the different functional areas identified in each animal. We

report in the Results the range of anterior–posterior (AP) y -axis coordinates for the peak effects in each contrast. We also projected our statistical results onto a rendered and inflated version of a single macaque cortical surface (F99, packaged with CARET), which was normalized to the standardized Saleem and Logothetis stereotaxic space.

Region of interest definition for decoding. For purposes of decoding, we defined regions of interest (ROIs) using the coordinate in each monkey's native space corresponding to the peak effect in every face-selective and Mf area. For Mf ROIs, we identified these coordinates using eight runs from Monkeys 1 and 3 and four runs from Monkey 2, and decoding was performed on the remaining runs. This ensured independence between voxel selection and subsequent decoding. We selected voxels to submit to decoding within 4 mm radius spheres around the coordinate of the peak effects (see Fig. 2 for examples). All ROIs contained mutually exclusive sets of voxels. To extract data used for decoding, we repeated the first-level GLMs using unsmoothed Mf run data in each monkey's native space. We could then decode the MION-deconvolved response estimates to each individual block.

Univariate analyses of ROI mean responses. We first compared the mean response profiles in the ROIs (averaged over voxels) that were selected for decoding: face-selective and Mf ROIs. This entailed testing whether the two ROI types differed in their (1) selectivity to faces; (2) responses to facial dynamics and expressions, and (3) responses to low-level motion (dot) stimuli. We therefore submitted the ROI data to three ANOVAs (see Results, Comparison of mean responses in face-selective and Mf ROIs). For all three of these ANOVAs, we included the ROI (nested in the face-selective and Mf ROI types) and monkey as nuisance random-effects factors. This procedure allowed us to test for our fixed effects of interest (described below), while also statistically controlling for nuisance variability among monkeys and the individual ROIs composing each ROI type.

The first of these ANOVAs was applied to the face-selectivity run data and used two fixed-effects factors: face category (face or non-face) and ROI type (face-selective or Mf ROI). The second of these ANOVAs was applied to all the Mf run data that was not used for definition of the Mf ROIs (see above) and was therefore an independent dataset. It used three fixed-effects factors: motion (dynamic vs static faces), expression (threat, submissive, or neutral expressions), and ROI type (face-selective vs Mf ROIs). The third ANOVA was applied to the Mf run data and compared the sensitivity of face-selective versus Mf ROIs to translation and optic flow motion in dots. Seven face-selective and 12 Mf ROIs, derived from Monkeys 1 and 2 were used. This ANOVA used three fixed-effects factors: motion (dynamic vs static dots), motion type (translation vs optic flow), and ROI type (face-selective vs Mf ROIs).

For all three ANOVAs, we were primarily interested in interactions between ROI type and the other fixed-effect factors. Whenever one of these three ANOVAs revealed such an interaction, we then further characterized the pattern of effects within each ROI type. This was done by computing similar ANOVAs separately for the two ROI types (see Results, Within face-selective and Mf ROI types).

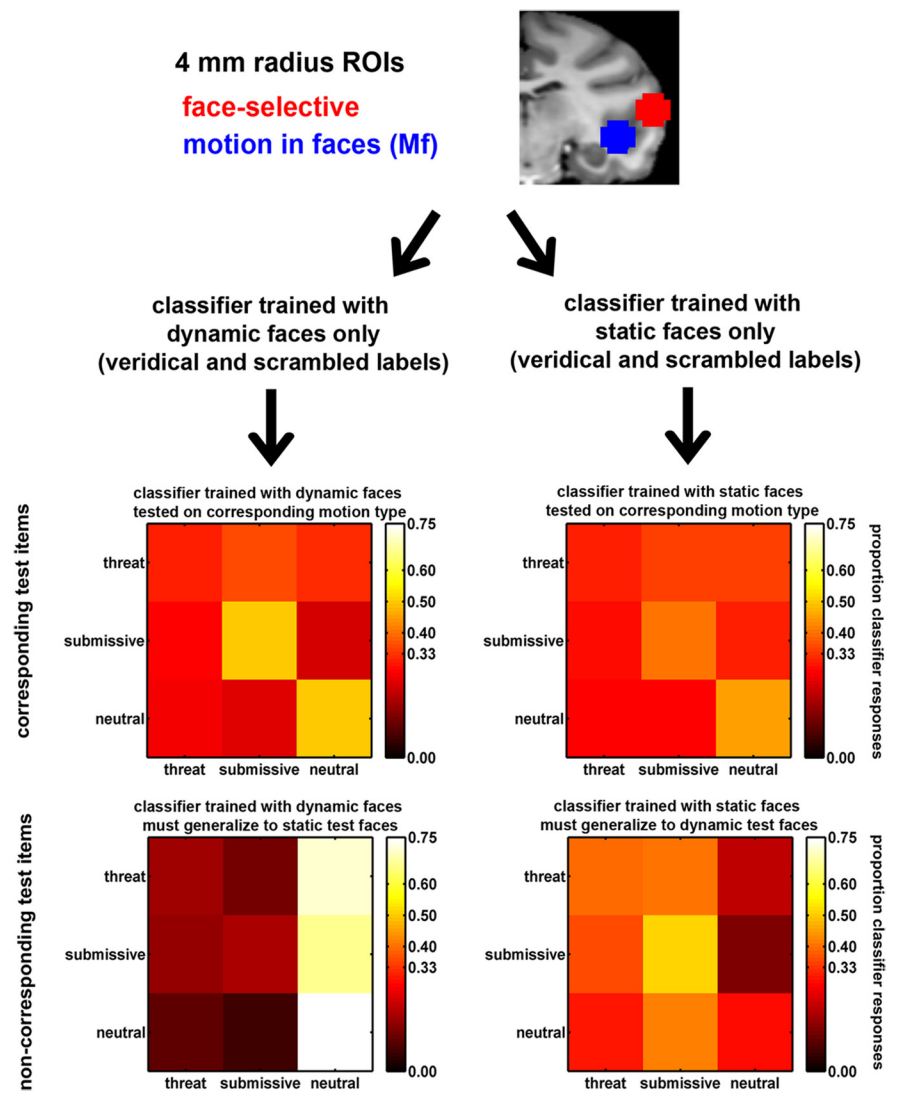


Figure 2. Decoding methods. Shown here are 4 mm radius spherical ROIs for the right middle face-selective (red) and visual Mf (blue) areas in Monkey 1. Voxel data from the Mf runs were used to train classifiers to perform “threat,” “submissive,” and “neutral” classifications. Training sets used either only dynamic (left) or static expressions (right) and used veridical or scrambled labels. We show example confusion matrices for the right posterior Mf ROI in Monkey 1. Here, expressions of test items are shown in rows and classifier responses to these items are shown in the columns. On-diagonal hit rates of 0.45–0.51 were well above chance (0.33) for dynamic and static submissive and neutral expressions when test motion corresponded to the training set (top row). For noncorresponding items (bottom row) classification is more unstable. Note that this is one example. Other ROIs showed much different patterns of accuracy and confusions. Our conclusions are based on summary measures over all ROIs.

Decoding strategy. Our decoding strategy (Fig. 2) was to train classifiers to perform a three-way expression classification of the Mf run data. Classifiers were trained with either only the dynamic expressions or only the static expressions. Furthermore, these classifiers were trained with either veridical or scrambled labels. Scrambled-labels classification provided an estimate of chance performance, which we compared against performance using veridical labels. Each classifier was tested with “corresponding” test items that had the same motion type (dynamic or static) as the classifier’s training set and separately with “noncorresponding” test items that had a different motion type than the training set. These noncorresponding items tested whether training with one motion type was useful for classifying the alternate motion type. We computed the average classification performance over the three expressions separately for classifiers trained with dynamic or static items, and veridical or scrambled labels and test items that were corresponding or noncorresponding. Below we provide more particular detail about these analysis steps.

Training and test data. We used linear discriminant analysis (Krzanowski, 1988) for multivariate decoding of the three expression

categories from data sampled within the 4 mm radius spherical ROIs. Performance of cross-validated classifiers is limited both by the true information in the data and by the number of trials available to train the classifier. Unless a sufficient number of trials are available to reveal the true information in the data, decoding performance can be degraded, especially using large numbers of voxels, because of the inclusion of noninformative voxels (Averbeck, 2009). To adjust for this, we repeated each classifier 200 times, each time training it with a different sample of 12 voxels randomly selected from each spherical ROI sphere (~70–80 voxels were in each sphere). We then report decoding based on the average over the ensuing distribution of performances. We found in practice that 12 voxels was sufficient for reliable classification while few enough to allow sampling of a large number of voxel samples from each ROI. All classifiers were trained with either the dynamic expression data or the static expression data. Last, all the aforementioned classifiers were also retrained 100 times more, but each time pseudo-randomly permuting the three expression labels. In summary, voxel samples from the face-selective and Mf spherical ROIs were used to train classifiers on either dynamic expressions or static expressions and using either veridical or scrambled expression labels (Fig. 2, top half).

All classifiers were tested using leave-one-run out cross-validation. The results from each classifier, when averaged over the left-out runs, yielded two 3×3 confusion matrices: one confusion matrix for items whose motion type corresponded to the training set and one where the motion type did not correspond to the training set. Figure 2 shows example confusion matrices for the right posterior Mf ROI in Monkey 1. These confusion matrices provided classifier response probabilities $p(R_j)$, stimulus category probabilities $p(S_i) = 1/3$, and joint probabilities $p(S_i, R_j)$. From these quantities we computed the partial information separately for corresponding and noncorresponding confusion matrices.

$$pMI_i = \sum_j \frac{p(S_i, R_j)}{p(S_i)} \log_2 \frac{p(S_i, R_j)}{p(S_i)p(R_j)}$$

This “conditional” or “partial” mutual information pMI_i , when averaged, gives the mutual information associated with the entire confusion matrix (Cover and Thomas, 1991).

$$MI = \sum_i p(S_i)pMI_i$$

In summary, pMI measures were computed separately for classifiers trained with dynamic or static expressions, and veridical or scrambled labels and test items that were corresponding or noncorresponding.

Significance testing. We used permutation testing to separately evaluate performance for each ROI when classifiers were trained using dynamic or static expressions and when they were tested with corresponding or noncorresponding test items. Figures 5A and 6A indicate in red those individual ROIs where the veridical performance exceeded performance for all 100 scrambled label permutations in more voxel samples than would be expected by chance at $p < 0.01$, according to the binomial distribution.

These permutation tests showed whether expression information could be decoded from individual ROIs. However, our primary hypothesis did not concern any individual ROI. Rather, we were interested in how face-selective and Mf ROIs systematically differed in the amount of expression information. We tested this hypothesis by using an ANOVA, where the dependent variable was the differences between performance using veridical and scrambled labels. We tested effects of fixed-effects factors, including ROI type (face-selective vs Mf ROI), motion (dynamic vs static faces), expression (threat, submissive, or neutral expressions), and correspondence (whether or not the motion type of test items corresponded to the motion type of the training set). Similar to our aforementioned univariate analyses of mean responses, the inclusion of ROI (nested in ROI type) and monkey as random effects allowed us to contrast face-selective and Mf ROIs, while also statistically controlling for irrelevant nuisance variability among monkeys and among the ROIs within each ROI type.

In summary, our approach allowed us to compute information-theoretic measures of expression coding in face-selective and Mf ROIs. These measures were derived separately for classifiers trained with dynamic or static expressions and test items that corresponded to the motion type of the training set or did not so correspond. We then compared this veridical performance against chance performance, as estimated by scrambled-labels classification. This approach allowed us to separately measure representations of dynamic and static expressions and to test whether their response patterns were distinct or confusable.

Results

Localization of areas

Face-selective areas

Using the face-selectivity run data, we first identified face-selective areas by contrasting fMRI responses to blocks of static neutral faces against the average response to places and nonface objects. Previous reports (Tsao et al., 2006) have shown one or more such patches to be clustered locally within two sites bilaterally, one in middle STS and one in anterior STS. Consistent with previous reports, we observed patches of face selectivity in both hemispheres at these two sites. In Monkeys 2 and 3, these areas were bilateral and were located in middle STS, around TEO (Fig. 3, red) and in anterior STS, in TE. We found the same areas in Monkey 1, although this monkey lacked the left anterior face-selective area (Fig. 1, top). This face selectivity was observed in the lateral aspect of the lower bank of the STS consistently across monkeys and areas. Occasionally, other macaques have shown face selectivity in the fundus of the STS (Ku et al., 2011). Despite our abundance of statistical power for detecting face selectivity in the lower lateral bank of the STS, our macaques did not show any suggestive numeric difference between faces and nonface stimuli within the fundus (see Fig. 4B for face selectivity within Mf areas, which are consistently situated deeper in the sulcus, see below), and this was consistent across our three monkeys. Peak face selectivity was located at AP y -axis coordinates (Saleem and Logothetis, 2007) in middle ($y = +7$) and anterior (y range between $+19$ and $+20$) locations in STS.

Mf areas

We next examined sensitivity to dynamic faces using the Mf run data. We defined our areas by comparing dynamic expressions against static expressions. This contrast can reflect a mix of different types of responses to visual motion. We expected that this contrast would elicit, for example, responses in brain areas sensitive to low-level (e.g., dot) motion, including FST, STPm, and LST. However, because this contrast used face stimuli, we also expected it to elicit additional motion sensitivity in voxels that might not be detected using a conventional low-level motion localizer, because they are sensitive to more complex forms of motion such as object motion or biological motion. For example, an area in anterior STS has been shown previously to be sensitive to dynamic hand grasps (Nelissen et al., 2011) but has not also been reported for low-level motion (Vanduffel et al., 2001; Nelissen et al., 2006). And, further, this contrast might reveal areas that are specific to facial motion, compared with other body parts, objects, low-level stimuli, etc. That is, this contrast can demonstrate all responses to visual motion in faces, whether those responses are specific to facial motion, biological motion, complex motion, low-level motion, etc. We assume that all these varieties of motion might produce responses useful to some degree for classifying facial expressions.

We chose not to define Mf areas as the contrast of dynamic faces versus dynamic objects. Because this contrast subtracts re-

sponses to objects from that of faces, it will reflect, in part, face selectivity. Our goal was to compare decoding in motion-sensitive areas against that in face-selective areas, and we did not wish to confound this comparison by defining motion sensitivity in a way that would also identify face selectivity. We also chose not to restrict our analysis only to areas that are selective to facial motion, compared with nonface motion, as identified by the interaction of face selectivity and motion sensitivity. We were especially interested in whether expression information could be decoded from areas that are not specialized for representing only facial attributes. Thus, we used a contrast we expected would reliably elicit motion areas known to be relatively domain-general, such as FST, STPm, and LST, for example.

Consistent with these expectations, we found three Mf areas in each hemisphere in each of the three monkeys (Fig. 3, yellow), which were consistently located in posterior, middle, and anterior STS. No areas were more activated by static faces compared with dynamic faces. All Mf areas showed clearly identifiable peak voxels, which were distinct and distant from the peaks of the face-selective areas identified above. Mf areas were situated more medially (deeper in the sulcus) than the peaks, corresponding to face-selective areas in every monkey (Fig. 3, black circles; Fig. 2, top). The posterior Mf area in all monkeys peaked in the fundus, near the lower bank of FST ($y = 0$). The middle Mf area peaked in IPa ($y = +4$ to $+6$). The anterior Mf area was also located deep within the sulcus, favoring the medial aspect of the upper bank, in IPa near TPO, according to the Saleem and Logothetis stereotaxic atlas ($y = +19$ to $+20$). The slice coverage in our Mf runs did not extend sufficiently posterior to fully evaluate MT/MST (Dubner and Zeki, 1971). The Mf areas we observed in posterior and middle STS were anatomically situated within locations previously identified as FST (Desimone and Ungerleider, 1986), STPm, and/or LST, all of which are well established as sensitive to low-level, nonface motion (Vanduffel et al., 2001; Nelissen et al., 2006). We therefore performed a separate conventional low-level localizer so that we could verify the low-level motion sensitivity of the Mf areas we found in FST, STPm, and LST, as well as to ascertain the low-level motion sensitivity of face-selective areas (Fig. 4).

Md areas

We used the Md run data to identify areas sensitive to dot motion by contrasting responses to all motion blocks with responses to all static blocks (Fig. 3, blue). Both monkeys showed bilateral Md areas in posterior STS encompassing MT and MST (denoted MT in Fig. 3, peaking at $y = -2$ to -3). A second cluster of Md voxels was located anterior to MT in bilateral FST (denoted FST in Fig. 3, peaking $y = +4$ to $+5$). Monkeys 1 and 2 also showed middle

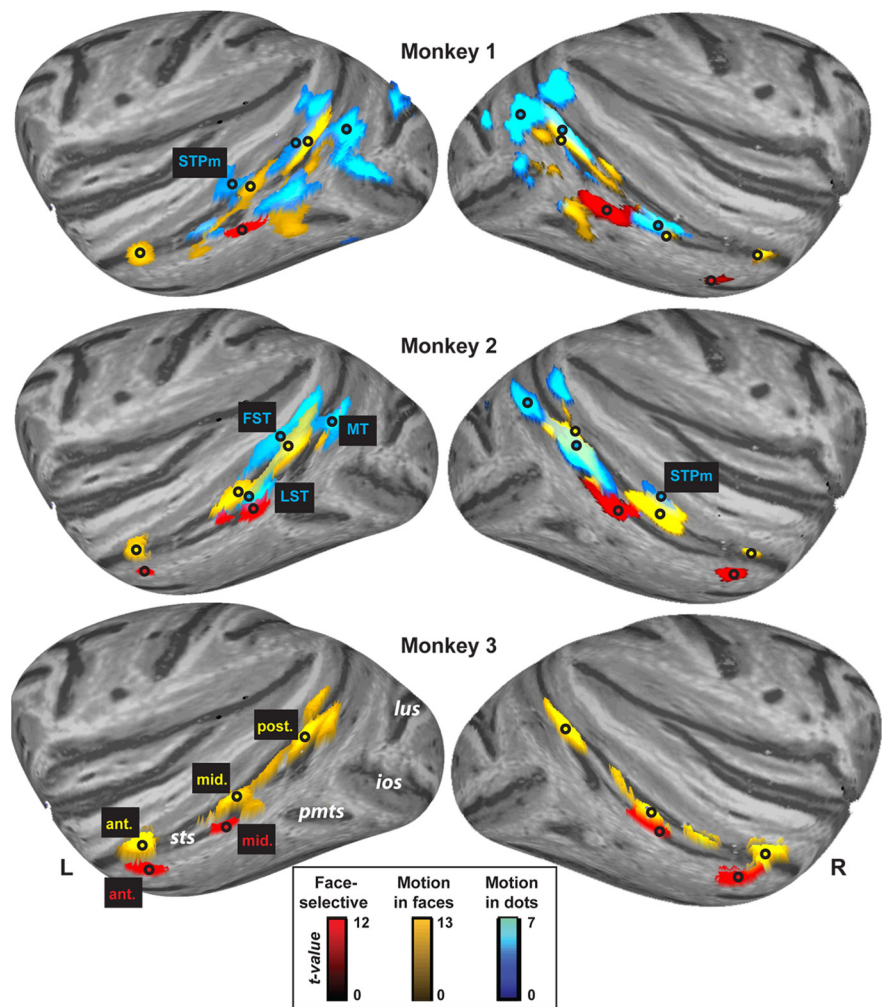


Figure 3. Functional areas in the superior temporal sulcus of both right and left hemispheres. Lateral inflated cortical surfaces showing significant results for dynamic versus static faces ($p < 1 \times 10^{-4}$ uncorrected) from the Motion in faces runs (yellow), faces versus nonfaces ($p < 1 \times 10^{-4}$) from the face-selectivity runs (red), and dynamic versus static dots ($p < 0.001$) from the Motion in dots runs (blue). Yellow text is used to label Motion in faces areas, red text for face-selective areas, and blue text for the Motion in dots areas. Black circles mark the location of peak effects. White text indicates relevant sulci. ant., Anterior STS area; ios, inferior occipital sulcus; lus, lunate sulcus; mid., middle STS area; pmts, posterior middle temporal sulcus; post., posterior STS area; sts, superior temporal sulcus; r, right hemisphere; l, left hemisphere.

STS Md areas (denoted STPm and LST in Fig. 3, peaking $y = +6$ to $+7$). These locations in MT, FST, STPm, and LST replicate numerous previous low-level motion studies (see Introduction). FST, STPm, and LST Md areas peaked near and overlapped with Mf areas, but there was no overlap with face-selective areas (Fig. 3, black circles and yellow and blue areas). Thus, we found that some Mf areas did not respond to faces specifically, but encompassed parts of cortex known to be domain general, sensitive to motion for a variety of stimulus categories. There was no evidence for any anterior Md areas, consistent with previous reports (Vanduffel et al., 2001; Nelissen et al., 2006).

Univariate analyses of ROI mean responses

Comparison of mean responses in face-selective and Mf ROIs

In this section, we report ROI analyses that verify statistically that the Mf ROIs were outside of face-selective cortex yet, nevertheless, could be sensitive to nonface, low-level motion. These analyses also test whether face-selective ROIs show sensitivity to visual motion elicited by faces and by dots.

We first compared the face selectivity of the two ROI types using data from the face-selectivity runs. Face-selective ROIs,

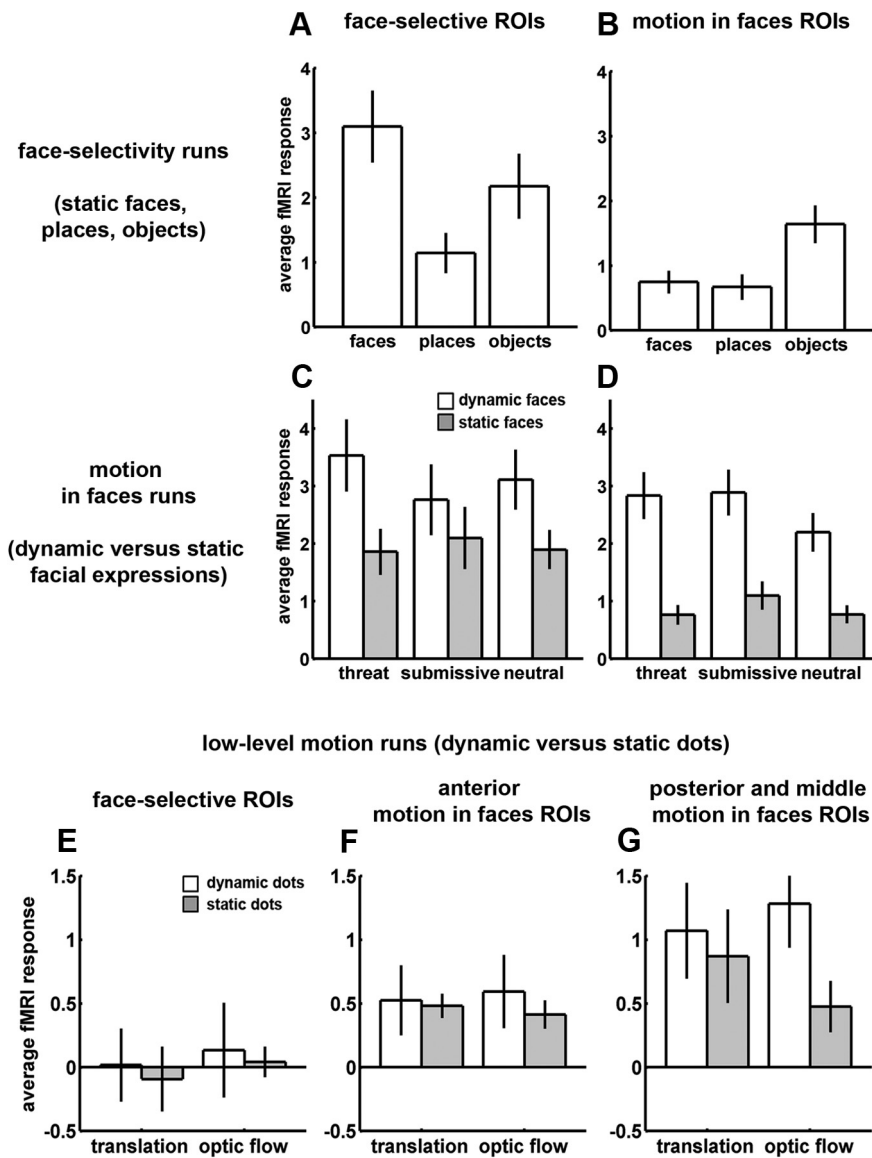


Figure 4. Response profiles of ROIs selected for decoding. Mean fMRI responses and SEs computed over face-selective (left column) and Mf ROIs (right column). All ROIs from both hemispheres are included. Rows, Responses from face-selectivity runs (A, B), Mf runs (C, D), and Md runs (E, F, G). Responses were multiplied by -1 to correct for the negative response deflection caused by MION (see Materials and Methods). ROI voxels were selected using an independent dataset. These response profiles were from the same unsmoothed data used for decoding (Figs. 2, 5, 6).

unlike Mf ROIs, showed robust responses to static faces (Fig. 4A) but smaller responses to places and objects (Fig. 4B). Statistically, this difference was demonstrated by a significant interaction of face category and ROI type ($F_{(1,73)} = 18.57, p < 0.0001$).

We also compared the responses of the two ROI types to motion in faces and to expressions using the Mf run data. Mf ROIs (Fig. 4D) showed a larger response difference between dynamic faces and static faces than did face-selective ROIs (Fig. 4C). Although both ROI types favored dynamic faces to some degree, this effect was larger for Mf ROIs as demonstrated by a significant interaction of ROI type and motion ($F_{(1,152)} = 3.8, p = 0.05$). There were no main effects of expression nor did expression interact with ROI type and/or motion.

Last, we compared the low-level, (dot) motion sensitivity of the two ROI types using the Md run data. Posterior and middle Mf ROIs showed a numerically larger response to moving dots (especially for optic flow, Fig. 4G) compared with anterior Mf

ROIs (Fig. 4F) and more so to the average of all face-selective ROIs (Fig. 4E). This pattern produced a significant main effect of motion $F_{(1,35)} = 4.79, p < 0.05$ (collapsed across all ROIs). Unfortunately, this ANOVA had less data, and hence less power, than the others, because there were insufficient observations available for testing differences among individual ROIs in only two monkeys. Moreover, not all Mf ROIs appeared to be sensitive to dot motion (anterior ROIs were not). Thus, there was not sufficient power to detect interactions with ROI type. However, ANOVAs performed separately for each ROI type yielded more conclusive results, as discussed in the next section.

Within face-selective or Mf ROI types

In addition to our aforementioned ANOVAs, which directly compared the two ROI types, we also used similar ANOVAs, but which were restricted to analysis of the individual ROI types. Before describing in detail the effects that we found, we mention that none of them showed any interactions with our ROI factor. This means all effects reported were statistically consistent across ROIs and, moreover, suggests no evidence for any effects that were lateralized to one hemisphere.

We used the Mf data to examine responses to motion and expressions in only the face-selective ROIs. Overall, we found greater responses to dynamic than static faces in face-selective ROIs, but this was inconsistent across expressions. Only threat and neutral expressions showed greater responses to dynamic faces (Fig. 4C), reflected by a significant interaction between expression and motion ($F_{(2,40)} = 4.42, p = 0.05$). When we tested for effects of motion for each expression individually, we found that threat ($F_{(1,15)} = 7.97, p = 0.01$) and neutral ($F_{(1,15)} = 10.46, p = 0.006$) but not submissive ($p = 0.25$) expressions showed greater responses to dynamic than static expressions.

We next used the Mf run data to examine responses to motion and expressions in only the Mf ROIs. Overall, these ROIs showed robust motion sensitivity across all expressions with relatively small responses to static faces (Fig. 4D). While there was no significant expression \times motion interaction, there was a significant main effect of motion ($F_{(1,70)} = 72.96, p = 0.0004$). When analysis was restricted to individual expressions, every expression separately showed significant effects of motion (all $p < 0.0001$).

We also examined the Mf run data to test whether either face-selective ROIs or Mf ROIs showed any mean differences between expressions. Face-selective ROIs did not show any significant effects of expression, either when the two motion conditions were collapsed together, or when static expressions and dynamic expressions were tested individually. Mf areas, on the other hand, showed response differences for dynamic expressions. When col-

lapping over motion conditions, Mf ROIs produced a significant main effect of expression ($F_{(2,70)} = 12.08, p = 0.002$). Significant pairwise differences among expressions emerged when dynamic expressions were considered alone, with Mf responses to neutral expressions differing from both threat ($F_{(1,27)} = 4.46, p = 0.04$) and submissive ($F_{(1,27)} = 5.37, p = 0.03$) expressions. This is not surprising, as there is likely to be less motion in neutral expression videos. Static expressions showed no significant pairwise differences.

Finally, for the Md runs, face-selective ROIs (Fig. 4E) did not show an effect of dot motion ($p = 0.61$) while Mf ROIs did ($F_{(1,23)} = 11.58, p < 0.02$). Numerically, posterior and middle Mf areas showed more sensitivity to moving dots than anterior Mf ROIs (Fig. 4F,G), although there was not sufficient data to show any effects involving ROI. Neither ROI type showed any significant main effect or interactions involving motion type (translation or optic flow). In summary, despite our reduced statistical power for the Md runs, we nevertheless obtained positive evidence favoring low-level (dot) motion sensitivity in Mf ROIs, but no such evidence for face-selective ROIs.

Together, these results show that face-selective and Mf ROIs exhibited notable differences in their mean response patterns. The former showed a strong response to static faces compared with nonface objects while Mf ROIs were not selective for faces. Both ROI types showed larger responses to dynamic than to static faces, although this effect was larger and more consistent for Mf ROIs. Face-selective ROI showed no sensitivity to dot motion while Mf ROIs did. These findings are important for characterizing the functional separability of face-selective and Mf areas. Face-selective areas are sensitive to visual motion in faces and might appear encompassed by Mf areas at liberal significance thresholds (Fig. 3). However, the spherical Mf voxels we selected from around the peak contrast (see above) cannot be construed as face-selective at any reasonable significance level (Fig. 4B) and so constitute an area clearly outside of face-selective cortex.

Decoding performance

For each ROI, we trained linear discriminant classifiers to perform three-way expression classifications on the Mf run data using threat, submissive, and neutral labels. Separate classifiers were trained on the dynamic and static expressions but all six motion and expression combinations were tested. Figure 2 shows confusion matrices resulting from analysis of the right posterior Mf area in Monkey 1. For test items whose motion type corresponded to that of the training set, this ROI showed accurate classification of both submissive and neutral expressions, whether they were dynamic or static. The hit rate, shown on the diagonal, was between 0.45 and 0.51 for these two expressions,

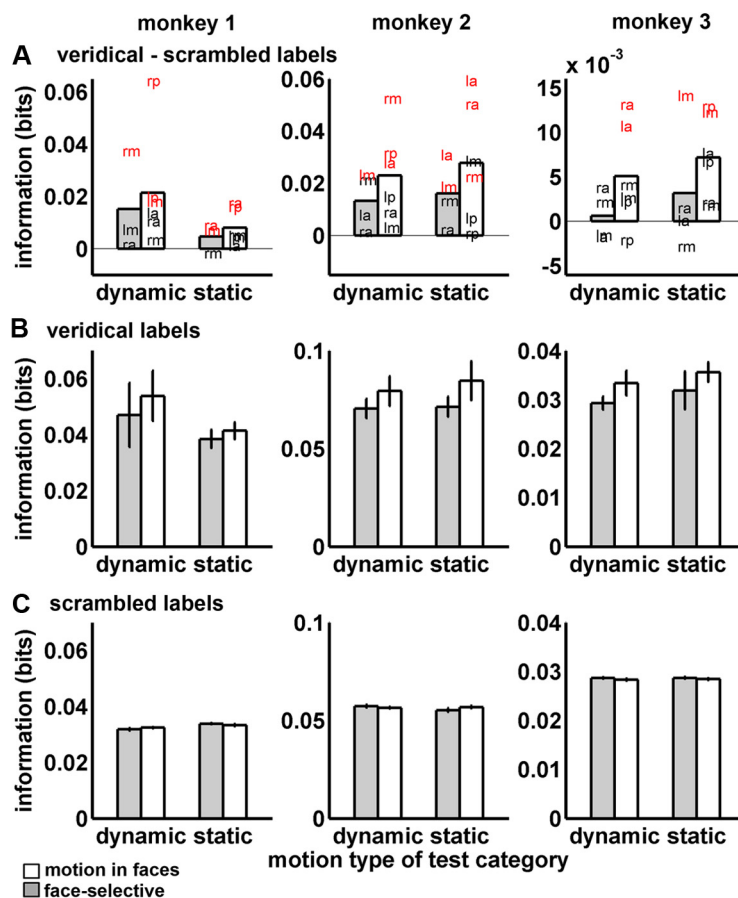


Figure 5. Expression decoding from corresponding items. Bars show means and SDs of the mutual information (bits) with face-selective ROIs in white and Mf areas in gray. The three monkeys are shown in the columns. In all graphs, training and test motion types corresponded. Thus, bars denoted “dynamic” show performance where training and test expressions were both dynamic and bars denoted “static” show performance where training and test items were both static. Graphs in **A** plot the difference between veridical and scrambled labels decoding. Letter positions indicate the mean performance (bits) for each ROI. lp, Left posterior; rp, right posterior; lm, left middle; rm, right middle; la, left anterior; ra, right anterior. Red letters denote ROIs where more voxel samples than were expected by chance at $p < 0.01$ outperformed distributions of scrambled labels classifications. Graphs in **B** plot the veridical performance and graphs in **C** plot the scrambled performance. In **B** and **C**, error bars show SEs over ROIs.

well above 0.33 chance performance. When the motion of the test items did not match the training set, classification was much more unstable. We note, however, that the ROIs showed heterogeneous patterns of confusions, with other ROIs showing a wide variety of classification patterns, making single ROIs such as shown in Figure 2 difficult to interpret. Conclusions should best be drawn as described below using summary performance measures using all the ROI data.

To summarize the findings, expressions could be decoded successfully from voxel patterns in our ROIs when the training and test items corresponded (Fig. 5A, differences between veridical and scrambled labels; Fig. 5B, veridical labels). There was, on average, more expression information in Mf ROIs than face-selective ROIs for both dynamic and static expressions, and this pattern replicated across all three monkeys. Figure 6 shows performance when a classifier trained with one motion type (dynamic or static) was challenged to decode test items of the alternate motion type (Fig. 6A, differences between veridical and scrambled labels; Fig. 6B, veridical labels). Overall, the ROIs showed inconsistent generalization between motion types. Next, we support this overall pattern of results statistically.

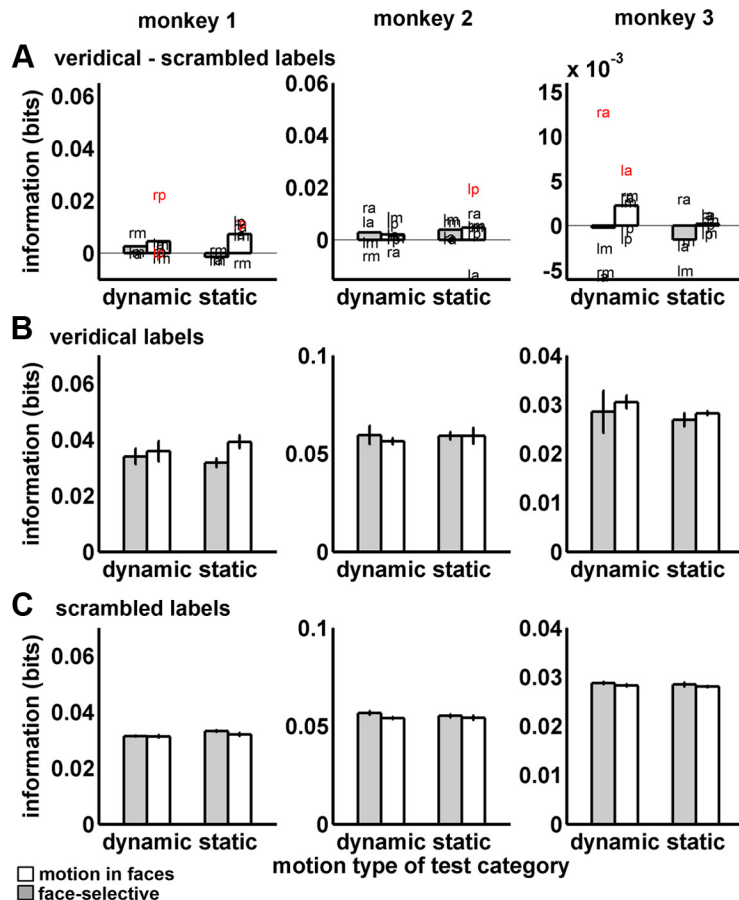


Figure 6. Expression decoding from noncorresponding items. Bars show means and SDs of the mutual information (bits) with face-selective ROIs in white and visual motion in faces areas in gray. The three monkeys are shown in the columns. In all graphs, training and test motion types did not correspond and therefore the ability of each classifier to generalize across motion types was tested. Bars denoted “dynamic” show performance where test items were dynamic but training items were static. Bars denoted “static” show performance where test items were static but training items were dynamic. Graphs in **A** plot the difference between veridical and scrambled labels decoding. Letter positions indicate the mean performance (bits) for each ROI. lp, left posterior; rp, right posterior; lm, left middle; rm, right middle; la, left anterior; ra, right anterior. Red letters denote ROIs where more voxel samples than were expected by chance at $p < 0.01$ outperformed distributions of scrambled labels classifications. Graphs in **B** plot the veridical performance and graphs in **C** plot the scrambled performance. In **B** and **C**, error bars show SEs over ROIs.

Comparing face-selective and Mf ROIs

When all factors were included in ANOVA (see Materials and Methods), our primary finding was a significant three-way interaction between ROI type, correspondence, and monkey ($F_{(2,326)} = 4.56, p = 0.01$). This interaction arose because the difference between ROI types (Mf vs face-selective ROIs) was observed for all three monkeys when training and test motion corresponded (Fig. 5). Meanwhile, there was a less consistent pattern across monkeys in the noncorrespondence condition, with some Mf ROIs showing inconsistent generalization across monkeys (Fig. 6). As this higher-order interaction qualifies interpretation of the lower-order effects in the ANOVA, we will describe this effect in more detail using ANOVAs applied separately to the two correspondence conditions.

For classifier performance when training and test items came from the same motion type (corresponding items), better performance was found for Mf ROIs on average than for face-selective ROIs ($F_{(1,152)} = 9.29, p = 0.01$). This effect came about entirely from differences in veridical performance, with no differences shown for scrambled labels decoding (Fig. 5B,C). This advantage was consistent across dynamic and static test items and across expressions. ROI type did not significantly interact with the motion (dynamic or static) of the test items ($p = 0.33$) and/or with expression ($p = 0.58$).

Corresponding test items also showed a main effect of expression ($F_{(2,152)} = 6.41, p = 0.002$), driven by some reduction in performance for dynamic and static threat items, which was consistent across the three monkeys. There was no significant motion \times expression interaction ($p = 0.54$). By contrast, for classifier performance when training and test items came from different motion types (noncorresponding items), there were no significant main effects or interactions at $p < 0.05$, including no significant main effect of ROI type ($p = 0.19$).

In summary, when the motion type of training items corresponded to the motion type of test items, our classifiers showed better facial expression decoding from Mf ROIs than from face-selective ROIs. This was true for both dynamic and static expressions. However, when classifiers were trained with either dynamic or static expressions and were then challenged to generalize their training to the alternate motion type, then decoding performance was reduced. Thus, the classifiers did not often confuse response patterns to dynamic expressions with those to the static expressions.

Discussion

We examined visual coding of facial expressions in the macaque STS. We found that: (1) Mf areas in posterior and middle STS responded more to dynamic than static facial expressions and were sensitive to nonface (dot) motion but were not selective to faces; (2) a Mf area in anterior STS responded more to dynamic than static expressions but showed less sensitivity to dot motion and was not face-selective; (3) face-selective areas showed inconsistent differences between dynamic

and static faces but no sensitivity to dot motion; (4) facial expressions were more robustly decoded from Mf areas than from face-selective areas; and (5) Mf areas encoded dynamic and static versions of the same expressions using distinct response patterns.

Facial expressions are decoded outside face-selective cortex

The response amplitudes in Mf areas signaled the presence of visual motion. However, Mf area responses did not simply detect motion, as their multivariate patterns could be used to decode static facial expressions. Indeed, despite the weak fMRI response to static faces compared with dynamic faces, decoding for static expressions was still substantial, even equal to that for dynamic expressions. Decoding of static expressions may arise in part from neurons sensitive to static expression images, recorded in STS (Perrett et al., 1984). Indeed, Hasselmo et al. (1989) anticipated our findings, hypothesizing that expression-sensitive cells and motion-sensitive cells may be colocalized in the STS. Furthermore, motion implied by static photographs modulates responses in human motion-sensitive areas MT+/V5 and MST (Kourtzi and Kanwisher, 2000; Senior et al., 2000). Psychophysical experiments in humans also suggest that the brain predicts or

extrapolates motion trajectories of static presentations of implied motion. Static images can induce memory biases along trajectories predicted by implied motion (Freyd, 1983; Senior et al., 2000) and static facial expressions show similar predictive perceptual effects (Furl et al., 2010). This evidence can be explained if static images are coded in terms of implied motion. Static images of expressions are elements of larger sequences of facial positions and motion-sensitive areas may code static expressions in terms of the movement sequences from which they derive.

Face-selective areas manifested less expression information than Mf areas. These face-selective areas encode many attributes of faces and face-selectivity is often interpreted as reflecting domain-specific specialization for faces (Freiwald et al., 2009). Nevertheless, facial expressions need not be represented exclusively by domain-specific modules, which are dedicated to only to facial attributes. Indeed, Mf areas showed no evidence for such face-selectivity (Figs. 3, 4B). Moreover, we chose our dynamic versus static faces contrast to localize responses to all the visual motion in faces, not just motion specific to faces, compared with nonfaces. Thereby, we elicited Mf responses (Fig. 3) subsuming areas well established to be domain-general, including FST, STPm, and LST. These areas showed sensitivity to nonface, low-level motion both in our data (Fig. 2G) and in previous electrophysiological (Desimone and Ungerleider, 1986; Jellema and Perrett, 2003) and fMRI (Vanduffel et al., 2001; Nelissen et al., 2006) experiments. Thus, posterior and middle Mf areas manifested facial expression information, even though their response patterns were sensitive to nonface motion, while areas thought to be specialized for faces showed less facial expression information.

Although speculative, there is a greater possibility of domain specificity in the anterior Mf area we found. Neither our Md localizer (Fig. 3) nor previous fMRI studies of low-level motion have yet revealed this anterior area. Nevertheless, this area was not face-selective and extant evidence suggests it may not be sensitive only to facial motion. Electrophysiological recordings from anterior STS show sensitivity to biological motion (Oram and Perrett, 1996). fMRI data show dynamic hand actions elicit similar anterior STS sensitivity (Nelissen et al., 2011). Thus, the anterior STS area is perhaps selective for biological actions or complex motion.

Dynamic and static expressions exhibited different response patterns

Although Mf areas could code both dynamic and static expressions in terms of motion cues, the presence of motion also introduced differences in their response patterns. We trained discriminant classifiers separately with only dynamic or static expressions and showed accurate decoding from Mf areas when the motion types of training and test items corresponded. However, when classifiers trained on one motion type were challenged with test items from a different motion type, performance was degraded, showing that response patterns to dynamic and static versions of an expression are not sufficiently similar to produce many confusions. Even though both dynamic and static expressions are coded in Mf areas, their response patterns are largely distinct. This finding has important implications for numerous previous studies, which used static expressions under the questionable assumption that the brain represents static photographs similarly as naturalistic expressions. Interestingly, generalization across motion types was sometimes nonzero, suggesting that some subelements of the response pattern might also be shared between motion types. This topic could be further explored at the single neuron level.

Implications for face processing models in primates?

Facial expression representations therefore appear segregated from representations of other facial attributes. A similar organization may exist in the human, where the STS appears to be functionally distinct from a more ventral temporal lobe pathway, thought to be specialized for representing facial identities (Haxby et al., 2000) and where identities have been successfully decoded (Kriegeskorte et al., 2007; Natu et al., 2010; Nestor et al., 2011). The human STS, instead, may be part of a more dorsal pathway, implicated in expression representation, among other changeable facial attributes (Haxby et al., 2000). This pathway includes the posterior STS, which is sensitive to facial expression in static faces (Engell and Haxby, 2007). Some theories further assert a role for motion representations in the human STS (O'Toole et al., 2002; Calder and Young, 2005). Posterior STS is situated near the low-level motion-sensitive area hMT+ (O'Toole et al., 2002). Numerous fMRI studies have shown sensitivity to visual motion in faces in the human STS, with limited or absent findings in the ventral pathway (Thompson et al., 2007; Fox et al., 2009; Schultz and Pilz, 2009; Trautmann et al., 2009; Pitcher et al., 2011; Foley et al., 2012; Schultz et al., 2012). One study reports decoding of dynamic expressions from human STS (Said et al., 2010), while other studies suggest that this region may integrate form and motion information during face perception (Puce et al., 2003). Although the posterior STS is sometimes face-selective, motion-sensitivity in posterior STS is not specific to faces (Thompson et al., 2007). Nonface biological motion representation in the posterior STS has been widely studied (Giese and Poggio, 2003) and right hemisphere temporal lobe lesions anterior to MT+/V5 show impaired biological motion perception (Vaina and Gross, 2004).

Our data raise the possibility of a similar functional distinction in the macaque between motion-sensitive areas, which can represent expressions, and face-selective areas, which can represent other facial attributes, such as identity. However, there are fundamental differences between humans' and macaques' temporal lobe organization that complicate direct inferences about homology. Nevertheless, it is encouraging that the structure of object-related temporal lobe response patterns in human and macaque are highly similar (Kriegeskorte et al., 2008). Despite inevitable species-related differences, it is possible that both the macaque and the human may possess a distinct pathway that is responsive to visual motion and codes movements such as expressions.

This hypothetical homology between macaque and human STS areas is only one new research avenue suggested by our findings. Indeed, our findings offer a new perspective on facial expression coding that raises several new questions for research. Expression and identity decoding have never been directly compared in either the human or the macaque, although similar methods have been successfully applied to the human auditory system (Formisano et al., 2008). One goal would be to eventually discover the neural mechanisms by which motion enhances face recognition (Wehrle et al., 2000; O'Toole et al., 2002; Ambadar et al., 2005; Trautmann et al., 2009), using invasive procedures such as fMRI-guided electrophysiological recordings that allow direct measurement of population coding. Achieving this goal could motivate technological advancement, in the form of face recognition algorithms that benefit from motion information in video.

We performed a comprehensive analysis of the neural coding of facial expressions in the macaque and our results introduce a new perspective on visual coding of actions in human and monkey, which raises several new hypotheses. We emphasize a role for

motion-sensitive areas in visual coding of facial expressions, even when they are static. We propose both similarities and differences with which dynamic and static expressions are coded. Together, our data suggest a role for domain-general motion-sensitive areas that are outside of face-selective areas. These dissociations suggest a complex functional specialization in the temporal lobe. Our results may lead to a better understanding of how face recognition can be enhanced by motion, despite the complex stimulus changes associated with motion.

References

- 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. [CrossRef Medline](#)
- Averbeck BB (2009) Noise correlations and information encoding and decoding. In: *Coherent behavior in neuronal networks* (Rubin J, Josic K, Matias M, Romo R, eds). New York: Springer.
- Barraclough NE, Xiao D, Oram MW, Perrett DI (2006) The sensitivity of primate STS neurons to walking sequences and to the degree of articulation in static images. *Prog Brain Res* 154:135–148. [CrossRef Medline](#)
- Bruce C, Desimone R, Gross CG (1981) Visual properties in a polysensory area in the superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384. [Medline](#)
- Calder AJ, Young AW (2005) Understanding recognition of facial identity and facial expression. *Nat Rev Neurosci* 6:641–651. [CrossRef Medline](#)
- Cover TM, Thomas JA (1991) *Elements of information theory*, 2nd ed. New York: Wiley-Interscience.
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173. [CrossRef Medline](#)
- Desimone R, Ungerleider LG (1986) Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J Comp Neurol* 248:164–189. [CrossRef Medline](#)
- Dubner R, Zeki SM (1971) Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res* 35:528–532. [CrossRef Medline](#)
- Engell AD, Haxby JV (2007) Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia* 45:3234–3241. [CrossRef](#)
- Foley E, Rippon G, Thai NJ, Longe O, Senior C (2012) Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. *J Cogn Neurosci* 24:507–520. [CrossRef Medline](#)
- Formisano E, De Martino F, Bonte M, Goebel R (2008) “Who” is saying “what?” Brain-based decoding of human voice and speech. *Science* 322:970–973. [CrossRef Medline](#)
- Fox CJ, Iaria G, Barton JJ (2009) Defining the face processing network: optimization of the functional localizer in fMRI. *Hum Brain Mapp* 30:1637–1651. [CrossRef Medline](#)
- Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187–1196. [CrossRef Medline](#)
- Freyd JJ (1983) The mental representation of movement when static stimuli are viewed. *Percept Psychophys* 33:575–581. [CrossRef Medline](#)
- Furl N, van Rijsbergen NJ, Kiebel SJ, Friston KJ, Treves A, Dolan RJ (2010) Modulation of perception and brain activity by predictable trajectories of facial expressions. *Cereb Cortex* 20:694–703. [CrossRef Medline](#)
- Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4:179–192. [CrossRef Medline](#)
- Hadj-Bouziane F, Bell AH, Knusten TA, Ungerleider LG, Tootell RB (2008) Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. *Proc Natl Acad Sci U S A* 105:5591–5596. [CrossRef Medline](#)
- Hasselmo ME, Rolls ET, Baylis GC (1989) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res* 32:203–218. [CrossRef Medline](#)
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223–233. [CrossRef Medline](#)
- Jellema T, Perrett DI (2003) Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia* 41:1728–1737. [CrossRef Medline](#)
- Knappmeyer B, Thornton IM, Bülthoff HH (2003) The use of facial motion and facial form during the processing of identity. *Vision Res* 43:1921–1936. [CrossRef Medline](#)
- Knight B, Johnston A (1997) The role of movement in face recognition. *Vis Cogn* 4:265–273. [CrossRef](#)
- Kourtzi Z, Kanwisher N (2000) Activation in human MT/MST by static images with implied motion. *J Cogn Neurosci* 12:48–55. [CrossRef Medline](#)
- Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci U S A* 104:20600–20605. [CrossRef Medline](#)
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141. [CrossRef Medline](#)
- Krzanowski WJ (1988) *Principles of multivariate analysis: a user’s perspective*. New York: Oxford UP.
- Ku SP, Tolia AS, Logothetis NK, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70:352–362. [CrossRef Medline](#)
- Lander K, Davies R (2007) Exploring the role of characteristic motion when learning new faces. *Q J Exp Psychol (Colchester)* 60:519–526. [CrossRef](#)
- Lander K, Christie F, Bruce V (1999) The role of movement in the recognition of famous faces. *Mem Cognit* 27:974–985. [CrossRef Medline](#)
- Leite FP, Tsao D, Vanduffel W, Fize D, Sasaki Y, Wald LL, Dale AM, Kwong KK, Orban GA, Rosen BR, Tottell RB, Mandeville JB (2002) Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. *Neuroimage* 16:283–294. [CrossRef Medline](#)
- McLaren DG, Kosmatka KJ, Oakes TR, Kroenke CD, Kohama SG, Matochik JA, Ingram DK, Johnson SC (2009) A population-average MRI-based atlas collection of the rhesus macaque. *Neuroimage* 45:52–59. [CrossRef Medline](#)
- Natu VS, Jiang F, Narvekar A, Keshvari S, Blanz V, O’Toole AJ (2010) Dissociable neural patterns of facial identity across changes in viewpoint. *J Cogn Neurosci* 22:1570–1582. [CrossRef](#)
- Nelissen K, Vanduffel W, Orban GA (2006) Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. *J Neurosci* 26:5929–5947. [CrossRef Medline](#)
- Nelissen K, Borra E, Gerbella M, Rozzi S, Luppino G, Vanduffel W, Rizzolatti G, Orban GA (2011) Action observation circuits in the macaque monkey cortex. *J Neurosci* 31:3743–3756. [CrossRef Medline](#)
- Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci U S A* 108:9998–10003. [CrossRef Medline](#)
- Oram MW, Perrett DI (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol* 76:109–129. [Medline](#)
- O’Toole AJ, Roark DA, Abdi H (2002) Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn Sci* 6:261–266. [CrossRef Medline](#)
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1984) Neurons responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Hum Neurobiol* 3:197–208. [Medline](#)
- 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. [CrossRef Medline](#)
- Puce A, Syngnetiotis A, Thompson JC, Abbott DF, Wheaton KJ, Castiello U (2003) The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *Neuroimage* 19:861–869. [CrossRef Medline](#)
- Roark DA, O’Toole AJ, Abdi H, Barrett SE (2006) Learning the moves: the effect of familiarity and facial motion on person recognition across large changes in viewing format. *Perception* 35:761–773. [CrossRef Medline](#)
- Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW (2009) A new method for improving functional-to-structural alignment using local Person correlation. *Neuroimage* 44:839–848. [CrossRef Medline](#)
- Said CP, Moore CD, Engell AD, Todorov A, Haxby JV (2010) Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis* 10:11. [CrossRef Medline](#)
- Saleem K, Logothetis NK (2007) A combined MRI and histology atlas of the rhesus monkey brain. Amsterdam: Academic.
- Schultz J, Pilz KS (2009) Natural facial motion enhances cortical responses to faces. *Exp Brain Res* 194:465–475. [CrossRef Medline](#)
- Schultz J, Brockhaus M, Bülthoff HH, Pilz KS (2012) What the human brain

- likes about facial motion. *Cereb Cortex*. Advance online publication. Retrieved . doi:10.1093/cercor/bhs106. [CrossRef Medline](#)
- Senior C, Barnes J, Giampietro V, Simmons A, Bullmore ET, Brammer M, David AS (2000) The functional neuroanatomy of implicit-motion perception or representational momentum. *Curr Biol* 10:16–22. [CrossRef Medline](#)
- Thompson JC, Hardee JE, Panayiotou A, Crewther D, Puce A (2007) Common and distinct brain activation to viewing dynamic sequences of face and hand movements. *Neuroimage* 37:966–973. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674. [CrossRef Medline](#)
- Vaina LM, Gross CG (2004) Perceptual deficits in patients with impaired recognition of biological motion after temporal lobe lesions. *Proc Natl Acad Sci U S A* 101:16947–16951. [CrossRef Medline](#)
- Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. *J Am Med Inform Assoc* 8:443–459. [CrossRef Medline](#)
- Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA (2001) Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron* 32:565–577. [CrossRef Medline](#)
- Vangeneugden J, De Mazière PA, Van Hulle MM, Jaeggli T, Van Gool L, Vogels R (2011) Distinct mechanisms for coding of visual actions in macaque temporal cortex. *J Neurosci* 31:385–401. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)