NeuroImage 56 (2011) 728–735

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Neural representation of observed actions in the parietal and premotor cortex

Kenji Ogawa [a,*], Toshio Inui [a,b]

[a] ERATO Asada project, Japan Science and Technology Agency, Kyoto, Japan
[b] Graduate School of Informatics, Kyoto University, Kyoto, Japan

ARTICLE INFO

*Article history:*
Received 30 November 2009
Revised 4 October 2010
Accepted 13 October 2010
Available online 23 October 2010

*Keywords:*
fMRI
Multi-voxel pattern analysis
Action recognition
Posterior parietal cortex
Premotor cortex

ABSTRACT

We investigated the neural representation of observed actions in the human parietal and premotor cortex, which comprise the action observation network or the mirror neuron system for action recognition. Participants observed object-directed hand actions, in which action as well as other properties were independently manipulated: action (grasp or touch), object (cup or bottle), perspective (1st or 3rd person), hand (right or left), and image size (large or small). We then used multi-voxel pattern analysis to determine whether each feature could be correctly decoded from regional activities. The early visual area showed significant above-chance classification accuracy, particularly high in perspective, hand, and size, consistent with pixel-wise dissimilarity of stimuli. In contrast, the highest decoding accuracy for action was observed in the anterior intraparietal sulcus (aIPS) and the ventral premotor cortex (PMv). Moreover, the decoder for action could be correctly generalized for images with high dissimilarity in the parietal and premotor region, but not in the visual area. Our study indicates that the parietal and premotor regions encode observed actions independent of retinal variations, which may subserve our capacity for invariant action recognition of others.


## Introduction

Recognition of others' actions is fundamental to our social interaction. Observing object-directed action involves multiple levels of descriptions ranging from low-level visual features, effectors in use, and means of manipulation, to targeted objects. We can recognize an observed action despite variations in viewpoint and distance as well as in an effector, and this requires invariant neural representations of actions in our brains. Compared with the view- or size-invariant object recognition subserved by the ventral visual stream in the occipito–temporal cortex (Riesenhuber and Poggio, 2002), little is known about neural mechanisms responsible for invariant action recognition with various visual features.

Previous studies have indicated that the human mirror neuron system (MNS) or action observation network, comprised of the posterior parietal cortex (PPC) and the ventral premotor region (PMv), has a role in the understanding of other's actions, based on findings of mirror neurons in monkeys (for review, see Rizzolatti et al., 2001; Iacoboni and Dapretto, 2006). A large number of papers with humans reported the parietal and premotor activations for observing or imitating actions of others using positron emission tomography (PET) (Grafton et al., 1996; Grezes et al., 1998), functional magnetic resonance imaging (fMRI) (Iacoboni et al., 1999; Buccino et al., 2001; Tanaka and Inui, 2002; Johnson-Frey et al., 2003), or magnetoencephalography (MEG) (Nishitani and Hari, 2000). However, the neural representation of observed actions in these regions is still unclear. One limitation of a conventional fMRI is the use of a subtraction paradigm for comparing different experimental conditions (e.g. observations of object-directed action versus those of object and hand separately), where the difference in various features is included. This makes it difficult to assess which levels of description in action understanding are encoded in which specific regions. Uncovering the neural representations involved in action recognition requires an understanding of the specificity for observed actions. This has been the approach taken in neurophysiological studies of mirror neurons tested for action selectivity (Gallese et al., 1996).

One possible way to achieve finer-scale spatial selectivity with fMRI is through the use of fMRI-adaptation or a repetition suppression paradigm (Grill-Spector and Malach, 2001). However, the largely unknown neurophysiological underpinnings and time-scales of adaptation, as well as attentional confounds accompanied with novelty or mismatch effects in repetitive presentation, make results difficult to interpret (Sawamura et al., 2006; Bartels et al., 2008; Logothetis, 2008). Recently, multi-voxel pattern analysis (MVPA) was proposed as a method for revealing neuronal selectivity beyond spatial resolution of fMRI voxels in a more direct way. MVPA differs from conventional analysis in that it evaluates brain activity regarding whether certain regions contain specific information about observed stimuli. It operates through analysis of differences in spatial pattern of activated voxels in the region, rather than overall differences in amplitudes of fMRI signals among different conditions (Haynes and

* Corresponding author. Asada Synergistic Intelligence Project, ERATO, Japan Science and Technology Agency, 89-3 Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8224, Japan. Fax: +81 75 706 7149.
*E-mail address:* ogawa@cog.ist.i.kyoto-u.ac.jp (K. Ogawa).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.10.043

Rees, 2005; Kamitani and Tong, 2005; Kriegeskorte et al., 2006; Norman et al., 2006). This enables investigation of neuronal selectivity as well as invariance to observed features, while controlling for potential confounders inherent in the conventional paradigms.

The present study used MVPA to reveal the neural representation of observed actions within the parietal and premotor cortex. Participants observed object-directed hand actions, which were independently manipulated in five different properties as follows: action, object, perspective, hand, and size of image. Neural representation of observed actions was investigated by analyzing whether these features could be correctly decoded from multi-voxel pattern activity of specific regions of interests (ROIs). This manipulation over multiple dimensions enabled us to reveal distinct neural representations encoded in different regions, as well as to dissociate low-level visual similarity of input images with action selectivity. We predicted that the early visual areas would show highest decoding accuracy, congruent with the retinal dissimilarity of observed images reflecting its retinotopic coordinates. In contrast, the parietal and premotor cortex regions were predicted to encode observed actions that would be rather robust to retinotopic variations, compared with the early visual area.

## Materials and methods

### *Subjects*

Subjects comprised nine volunteers (5 males and 4 females) with a mean age of 29.0 years (range, 23–33 years). No volunteer exhibited excessive head movement (over 2 mm) during scanning. All except one subject were right-handed, as assessed using the improved version of the Edinburgh Handedness Inventory (Oldfield, 1971). Written informed consent was obtained from all subjects in accordance with the Declaration of Helsinki. The experimental protocol received approval from the ethics committee of the Advanced Telecommunication Research Institute, Japan.

### *Task procedures*

Experimental stimuli, controlled by a personal computer external to the MRI scanner, were presented on a liquid crystal display and projected onto a custom-made viewing screen. Subjects lay supine in the scanner with the head immobilized and viewed the screen via a mirror. In the experiment, grayscale photos of object-directed hand actions were presented to participants (Fig. 1). We independently manipulated the following five properties of the observed images: ACTION (grasp or touch), OBJECT to be manipulated (cup or bottle), PERSPECTIVE (1st or 3rd person perspective), HAND (right or left), and SIZE of image (large; 7.5 deg. in visual angle or small; 5.0 deg.), which resulted in a total of different 32 ($=2^5$) types of stimuli (Fig. 2). The left-hand images were flipped versions of the right-handed ones. These were presented in separate 12-s blocks per run, in which 10 images of the same properties were presented every 1200 ms, with 800 ms for image presentation and 400 ms for a blank display. Every block was interleaved with a 12-s rest period of a fixation display. Each subject underwent 4 sessions, each consisting of 32 trials (=blocks), corresponding to 32 types of stimuli. The order of trials was pseudo-randomized across sessions and subjects. This resulted in a total of 128 ($=32\times4$) trials per subject, with 64 blocks for each property (e.g., regarding ACTION, there were 64 trials for 'grasp' and 64 for 'touch'). To ensure the subject's attention during scanning, we introduced subtle variations into the images, in which the hand's shape or orientation was slightly changed across images. Participants performed a one-back task, simultaneously pressing buttons with index fingers of both hands whenever the same stimulus appeared twice in a row, which occurred once or twice per block.

### *MRI acquisition*

A 3-T Siemens Trio scanner (Erlangen, Germany) with a 12-channel head coil was used to perform both T2-weighted anatomical imaging and T2*-weighted echo planar imaging (EPI). A total of 324

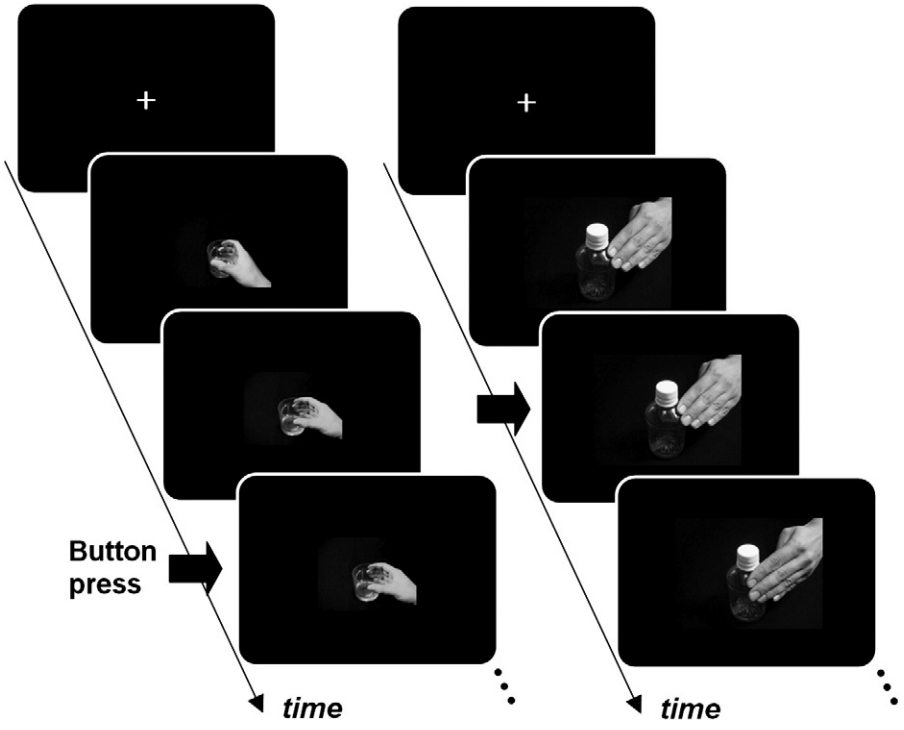
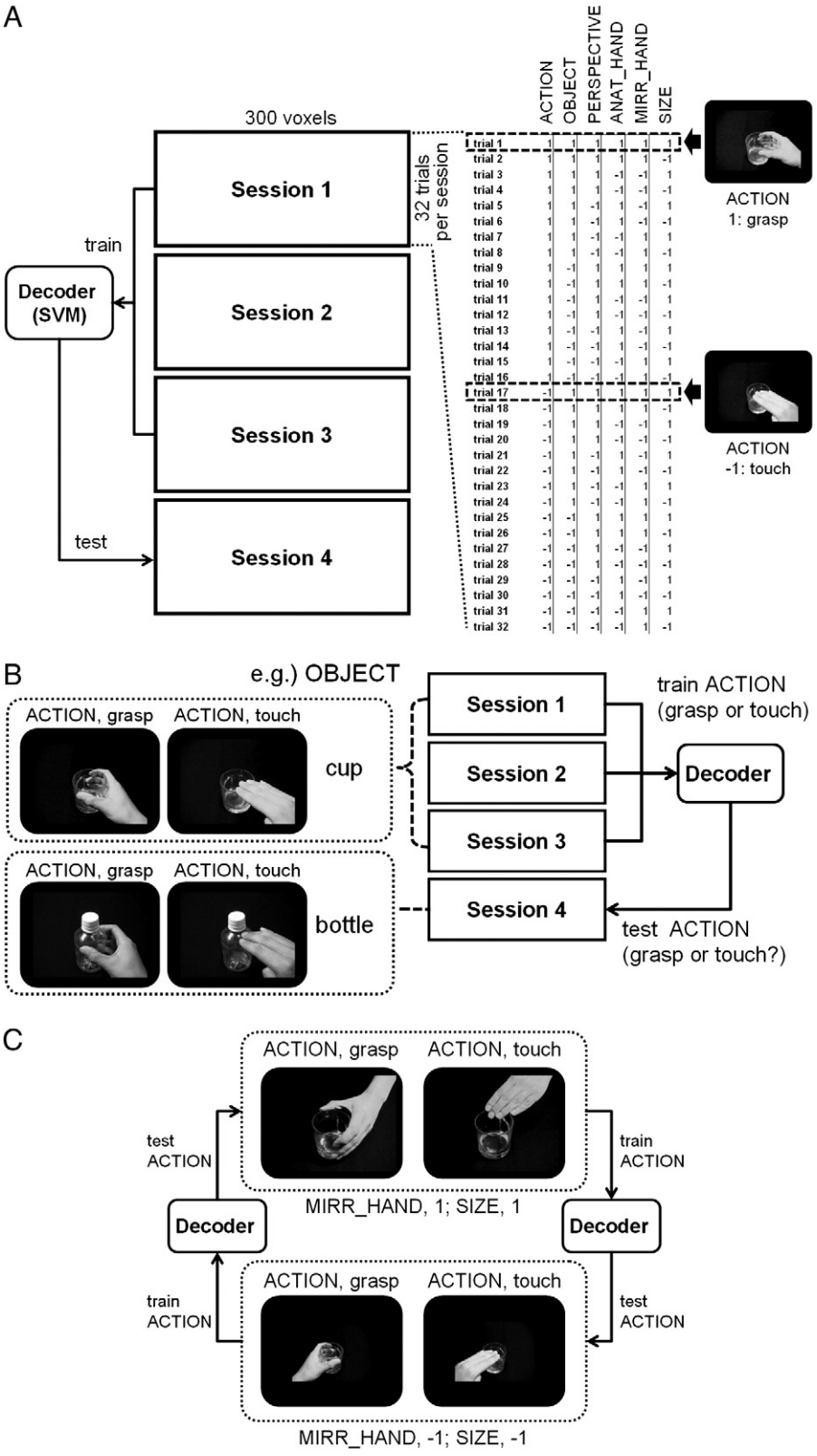



**Fig. 1.** The experimental time-course with examples of sample stimuli used in the experiment. Left) Grasp the cup with the right hand in the 1st person perspective, displayed in small size. Right) Touch the bottle with the left hand in the 3rd person perspective, displayed in large size. Participants performed a simple one-back task, pressing buttons whenever the same image appeared twice in a row.

scans were acquired with a gradient echo EPI sequence. The first four scans were discarded to allow for T1 equilibration. Scanning parameters were as follows: repetition time (TR), 2400 ms; echo time (TE), 30 ms; flip angle (FA), 80°; field of view (FOV), 192×192 mm; matrix, 64×64; 40 axial slices; and slice thickness, 3 mm without gap. T1-weighted anatomical imaging with a MP-RAGE sequence was also performed with the following parameters: TR, 2250 ms; TE, 3.06 ms; FA, 9°; FOV, 256×256 mm; matrix, 256×256; 192 axial slices; and slice thickness, 1 mm without gap.

*Preprocessing and modeling of fMRI data*

Image preprocessing was performed using SPM5 software (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.



uk/spm). All functional images were realigned to adjust for motion-related artifacts. The realigned images were spatially normalized with the Montreal Neurological Institute (MNI) template, based on spatial transformations derived from co-registered T2-weighted anatomical images, being resampled into 3 mm-cube voxels with sinc interpolation. All images were spatially smoothed using a Gaussian kernel of $8\times8\times8$ mm full width at half-maximum, except for data used for the MVPA analysis.

Using the general linear model (GLM), the 32 blocks per session were modeled as separate 32 box-car regressors that were convolved with a canonical hemodynamic response function. The six movement parameters resulting from the realignment stage were modeled as confounding covariates. Before estimation, low-frequency noise was removed using a high-pass filter with a cut-off period of 128 s, and serial correlations among scans were estimated with an autoregressive model implemented in SPM5. This analysis yielded 32 independently estimated parameters (beta-values) per session, which were subsequently used as inputs for decoding analysis (Eger et al., 2008). Use of beta-values for decoding, rather than raw signal intensities, allowed us to include motion covariates to obtain better estimates (Mur et al., 2009). Although individual modeling of every trial led to increased noise and decreased fitting in GLM, this could ensure a sufficient number of examples for decoding: a total of 128 samples per subject (Pereira et al., 2009). We then conducted two types of analysis: MVPA and conventional univariate analysis of single voxel with SPM5.

### *Definition of regions of interest (ROIs)*

We defined six ROIs, as follows: the left (L) and right (R) anterior parts of the intraparietal sulcus (aIPS) located posterior to the postcentral sulcus (L.aIPS [−39, −45, 47]; R.aIPS [37, −47, 52] in MNI coordinates), and the bilateral ventral parts of the premotor cortex (PMv) located at the junction of precentral sulcus and the inferior frontal gyrus (L.PMv [−53, 5, 25]; R.PMv [55, 8, 20]). These four ROIs were defined at the mean coordinates from 26 previously published functional imaging studies of execution and observation of hand action (a complete list is supplied in Supplementary Table 1 and Supplementary Fig. 1). As control ROIs, we chose the white matter (WM, [0, 0, 0]) as well as the right early visual area in Brodmann Area (BA) 17/18 corresponding to the primary/secondary visual cortex (R.Vis, [18, −96, −6]) (Amunts et al., 2000) that showed the most positive response during the task block compared with rest period based on the univariate analysis ($p<0.001$, uncorrected). All ROIs were defined as 15-mm radius spheres centered at each coordinate. The left early visual area (L.Vis, [−18, −96, −6]) was also analyzed, and it generated a pattern of decoding accuracy that was the same as that found for the right hemisphere (Supplementary Table 2).

### *Univariate analysis of a single voxel*

For univariate analysis of a single voxel, we compared the task period collapsed over all conditions with that of the rest period. We also contrasted each pair of all properties to reveal brain activations specific to differences in each feature. Contrast images of each subject, generated using a fixed-effects model, were taken into the group analysis using a random-effects model of a one sample *t* test. Percent signal changes were also calculated as averaged responses across all voxels contained in the specified ROIs, using Marsbar software (http://marsbar.sourceforge.net/).

### *Multi-voxel pattern analysis (MVPA)*

The classification analysis of MVPA was performed with a linear support vector machine (SVM) implemented in LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/), with default parameters (a fixed regularization parameter C = 1). Parameter estimates (betas) of each trial of voxels within ROIs were used as inputs to the classifier. Because we were interested in differences in spatial distributions of activated voxels, rather than overall amplitudes of activity within ROIs, the beta-values were z-normalized within each trial. *T*-values of betas under task-period collapsed over conditions, compared with the rest period, were estimated using only training runs to avoid non-independence errors (Kriegeskorte et al., 2009). Voxels in each ROI were selected in the order of highest *t*-value of training runs, which was based on the univariate analysis, until the number of voxels reached 300 for each subject (Pereira et al., 2009).

We then evaluated selectivity to different properties (= dimensions) of the observed images, focusing on whether this information could be correctly decoded from the multi-voxel pattern activity of the region—a method called "information-based analysis" (Kriegeskorte et al., 2006) (Fig. 2). HAND was labeled in two ways: either as anatomical (e.g., right hand in 3rd person perspective corresponds to right hand in 1st person perspective; designated ANAT_HAND) or mirror (e.g., right hand in 3rd person perspective corresponds to left hand in 1st person perspective; designated MIRR_HAND) correspondence. The 1st and 3rd PERSPECTIVEs contained an equal number of right and left-hand presentations in ANAT_HAND and MIRR_HAND trials. The averaged decoding accuracy was estimated with a 4-fold "leave-one-out" cross-validation based on 4 sessions, in which 3 sessions were used as training and the one remaining session as test data. This was repeated 4 times, leaving out a different session each time. The statistical test was conducted to determine whether observed decoding accuracy was significantly higher than chance (50%) using a two-sided *t* test, with inter-subject difference treated as a random factor (degree of freedom [d.f] = 8). The *p*-values were corrected for multiple comparisons across the six decoding properties for each ROI, using the Holm–Bonferroni procedure (Holm, 1979).

**Fig. 2.** A schematic figure of multi-voxel pattern analysis (MVPA). (A) We independently manipulated the following five properties of the observed images: ACTION (grasp or touch), OBJECT to be manipulated (cup or bottle), PERSPECTIVE (1st or 3rd person perspective), HAND (right or left), and SIZE of image (large; 7.5 deg. in visual angle or small; 5.0 deg.), which resulted in a total of different 32 ($=2^5$) types of stimuli. Two different features in each dimension (= stimulus feature) are labeled as '1' or '−1' in the figure. Each subject underwent 4 sessions, each of which included 32 trials (= blocks), corresponding to 32 types of stimuli. For example, grasping (labeled as '1' in ACTION dimension) and touching (labeled as '−1') actions were presented in trial 1 and 17, respectively, while all other properties remained unchanged. HAND was labeled either as anatomical and mirror correspondence (designated ANAT_HAND and MIRR_HAND, respectively). The parameter estimates (betas) of 300 voxels were extracted from each ROI, and then used as inputs to the decoder (linear support vector machine: SVM). The averaged decoding accuracy was measured with 4-fold "leave-one-out" cross-validation based on 4 sessions, in which 3 sessions were used as training and the one remaining session as test data; this was repeated 4 times leaving out a different session each time. (B) Analyzing invariance of action representation by training and testing ACTION with different properties of other dimensions (OBJECT, PERSPECTIVE, ANAT_HAND, MIRR_HAND, and SIZE). For example, to test invariance of ACTION decoding in regard to an OBJECT feature, samples of one object (cup) and the other object (bottle) were used for training and testing, respectively (and vice versa), to decode ACTION. The same procedure was conducted for other dimensions (PERSPECTIVE, ANAT_HAND, MIRR_HAND, and SIZE). (C) Analyzing invariance of action representation by training and testing ACTION using data with different features for training and testing in both MIRR_HAND and SIZE dimension: the images with the left hand in the 3rd person perspective or with the right hand in the 1st person perspective (labeled as '1' in MIRR_HAND dimension) in large size ('1' in SIZE) were used as training the decoder for ACTION, and ones with left hand in 1st person perspective or with right hand in 3rd person perspective ('−1' in MIRR_HAND) in small size ('−1' in SIZE) were used to decode ACTION. The same operation was repeated for 4 combinations of two properties of MIRR_HAND ('1'/'−1') and SIZE ('1'/'−1').

*Dissimilarity analysis of stimulus images*

To investigate the effect of low-level visual similarities of input images on decoding results, we analyzed pixel-wise dissimilarity of the stimuli. Firstly, the original images containing 612×816 pixels were down-sampled to 204×272 pixels with a mosaic-filter of 3×3 pixels. We then used two methods: 1) Euclidian distances of luminance in all pixels were measured between each pair of stimuli with the same and different properties, which were then used to define a dissimilarity index (Grill-Spector et al., 1999). 2) The luminance of all pixels were used as inputs to the classifier of SVM, then the decoding accuracy of each property was estimated in the same way as the MVPA (Eger et al., 2008).

## Results

*Dissimilarity analysis of stimulus images*

The dissimilarity index of stimulus images, which indicates the degree of pixel-wise dissimilarity among images, was highest for MIRR_HAND (dissimilarity index of 6.9 in arbitrary units), followed by SIZE (3.7), PERSPECTIVE (2.9), ANAT_HAND (2.8), and OBJECT (1.2), and was lowest for ACTION (0.3). The classification analysis of input image revealed 100% decoding accuracy in HAND (mirror and anatomical), PERSPECTIVE, SIZE, and lower in OBJECT (95.3%) and in ACTON (80.5%).

*Single voxel univariate analysis*

Using conventional univariate analysis, we first analyzed greater activations during task blocks collapsed over all conditions compared with rest periods. Large activations were observed that encompassed the parietal and premotor cortex, as well as the occipito–temporal and medial motor areas, including supplementary and pre-supplementary motor areas ($p<0.001$, uncorrected for the voxel level) (Fig. 3A).

We then compared each pair in all features and found different activation, mostly in the early visual areas, that closely corresponded to visual field effects of the hand presented in the display. For example, observing the right HAND in the 1st person PERSPECTIVE led to greater activation in the upper left part of the occipital cortex, because the hand was presented in the lower right part in the display. Regarding ACTION, greater activity was observed in the precuneus for touch compared with grasp. For MIRR_HAND, we also found selective activations in the posterior parietal cortex, mainly in the anterior part of intraparietal sulcus (aIPS): observing the hand from 1st and 3rd person perspective led to greater activity in the contra- and ipsi-lateral aIPS, respectively (Supplementary Fig. 2). We also conducted ROI analysis to examine differences in ACTION. This allowed direct comparison of averaged responses for viewing two actions (grasp vs. touch) with the rest, across all voxels in ROIs. We found significantly increased activation in grasp compared with touch in R.Vis ($p<0.05$), whereas no significant differences were found among other ROIs, including the bilateral aIPS and PMv ($p>0.1$) (Fig. 3B).

*Multi-voxel pattern analysis (MVPA)*

First, we independently estimated decoding accuracy for each property of observed images (ACTION, OBJECT, PERSPECTIVE, ANAT_-HAND, MIRR_HAND, and SIZE), collapsed over other dimensions. The right visual area (R.Vis) showed significant above-chance classification accuracy in all dimensions, with particularly high accuracy in MIRR_HAND (86.5%), SIZE (84.1%), ANAT_HAND (75.0%), and PERSPECTIVE (75.0%), compared with ACTION (67.4%) and OBJECT (62.9%). In contrast, significant above-chance decoding accuracy was observed in the bilateral aIPS, showing the highest accuracy in ACTION (L, 68.8%; R, 75.7%), followed by MIRR_HAND (L, 62.9%; R, 62.6%) and OBJECT (L, 56.8%; R, 55.5%). No significant accuracy was noted for other properties. The bilateral PMv showed significant accuracy for ACTION (L, 61.6%; R, 59.7%) with no significant accuracy for other properties. The control ROI in the white matter (WM) showed no greater decoding accuracy significantly above chance for any property ($p>0.1$) (Fig. 4; Supplementary Table 2). To investigate the effect of number of voxels used in the analysis, we also compared decoding accuracies with 100, 200, and 300 voxels from each ROI. Although decoding accuracies generally tended to be better with increasing number of voxels, overall patterns of decoding accuracies within each ROI were mostly stable and robust (Supplementary Fig. 3).

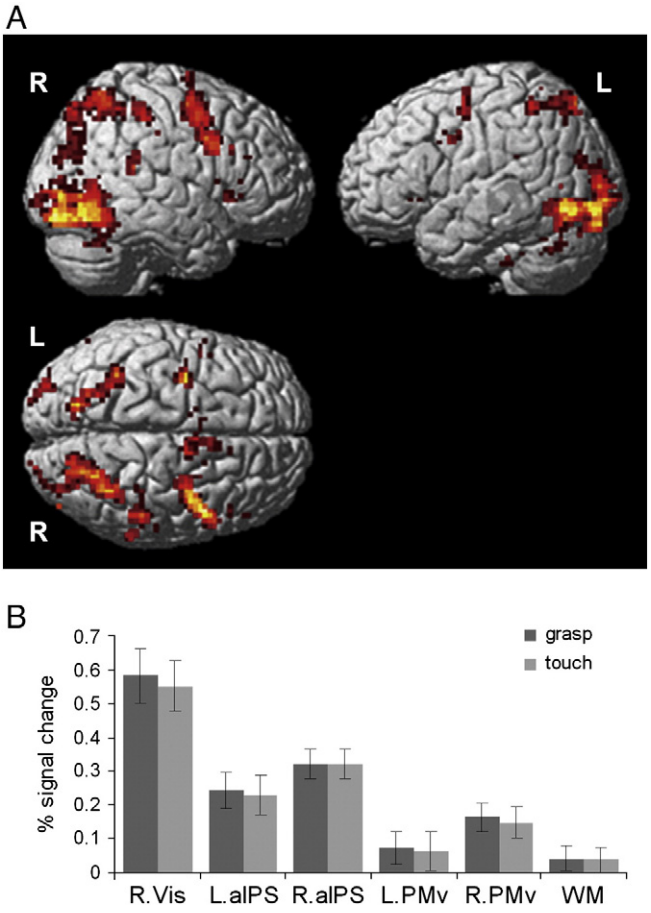



**Fig. 3.** Results of univariate analysis. (A) Regions showing greater activation during a task period, compared with a rest period, revealed with a conventional uni-voxel analysis ($p<0.001$ uncorrected for the voxel level; cluster threshold = 1 voxel). R, right; L, left. (B) Mean percent signal changes compared with a rest period regarding ACTION differences (grasp vs. touch) across all voxels in ROIs by the uni-voxel analysis. Error bars denote SEMs.

Next, to investigate the invariance or robustness of ACTION representation with respect to other properties (=dimensions), we used data with different features in all other dimensions (OBJECT, PERSPECTIVE, ANAT_HAND, MIRR_HAND, and SIZE) except ACTION, for decoding of ACTION. For example, to investigate invariance of ACTION representation in regard to an OBJECT property, samples of one object (cup) and the other object (bottle) were used for training and testing, respectively, to decode ACTION (see Fig. 2B for details). (In contrast, the previous analysis collapsed the OBJECT dimension, using all samples of two objects for both training and testing, to decode ACTION.) This operation was repeated 8 times based on different combinations of 4 fold structures (=sessions) and 2 properties of the other stimulus dimension (e.g., for OBJECT, there are two properties: 'cup' and 'bottle'), and then the averaged decoding accuracy was obtained. The same procedure was conducted for other

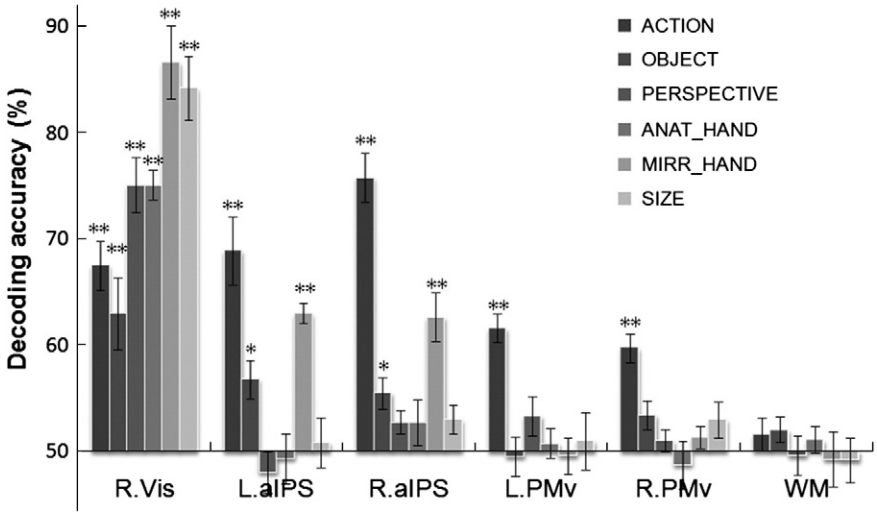



**Fig. 4.** Decoding accuracy for each feature of the observed images collapsed over other features (*, $p<0.05$; **, $p<0.01$ Holm–Bonferroni corrected for multiple comparisons). Error bars denote SEMs. The statistical results including *t*- and *p*-values were summarized in Supplementary Table 2.

dimensions (PERSPECTIVE, ANAT_HAND, MIRR_HAND, and SIZE). This process revealed significant above-chance classification accuracies for ACTION regarding all other dimensions in the bilateral aIPS and L.PMv, as well as the R.Vis. Significant decoding accuracies regarding ANAT_HAND, and SIZE were also observed in the R.PMv, with no significance in terms of other properties (Fig. 5; Supplementary Table 3).

As a last step, we aimed for maximum elimination of the effects of low-level similarity of stimulus images and further testing of invariance/robustness of ACTION representation. We performed another MVPA that decoded ACTION by combining two properties that had high dissimilarity indexes of images (MIRR_HAND and SIZE): the images with the left hand in the 3rd person perspective or with the right hand in the 1st person perspective (labeled as '1' in MIRR_HAND dimension), in large size ('1' in SIZE), were used to train the decoder for ACTION. The images with the left hand in the 1st person perspective or with the right hand in the 3rd person perspective ('−1' in MIRR_HAND), in small size ('−1' in SIZE), were used to decode ACTION (see Fig. 2C for details). This operation was repeated 16 times with different combinations of 4 fold structures (=sessions), and 4 combinations of two properties of MIRR_HAND ('1'/'−1') and SIZE ('1'/'−1'), and then averaged decoding accuracy was obtained. Non-significant decoding accuracy was revealed in R.Vis (52.8%, $p>0.1$), while significant decoding accuracies were observed in aIPS (L, 66.8%; R, 66.7%; $p<0.01$) as well as in PMv (L, 57.6%, $p<0.01$; R, 55.6%, $p<0.05$) (Fig. 5; Supplementary Table 3).

## Discussion

The current study investigated the neural representation of observed actions in the parietal and premotor areas comprising the human MNS or action observation network. Conventional univariate analysis showed greater activations primarily in the occipital, parietal, and premotor areas for action observation compared with a rest period, which is generally congruent with previous studies of action observation (for review, see Grezes and Decety, 2001). However, it failed to show the selectivity for action within the aIPS and PMv previously observed in fMRI studies of human MNS (for review, see Rizzolatti et al., 2001; Iacoboni and Dapretto, 2006). In contrast, MVPA revealed selectivity for action in all ROIs, except for the control ROI in the white matter. More importantly, each ROI showed different patterns of decoding accuracy: the early visual area showed

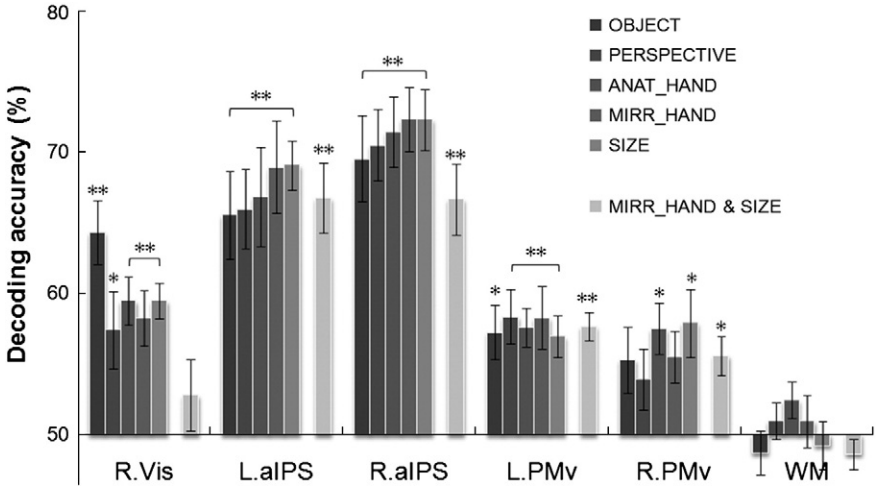



**Fig. 5.** Decoding accuracy for each feature of the observed images (*, $p<0.05$; **, $p<0.01$ Holm–Bonferroni corrected for multiple comparisons), using data with different features for training and testing of other dimensions (see Fig. 2B for details). The bar graph of 'MIRR_HAND & SIZE' represents decoding accuracy of ACTION using data with different features for training and testing in both MIRR_HAND and SIZE dimension (see Fig. 2C for details). Error bars denote SEMs. The statistical results including *t*- and *p*-values were summarized in Supplementary Table 3.

significant above-chance classification accuracy in all properties, with particularly high accuracy in perspective, hand (both anatomical and mirror), and size. This pattern of decoding accuracy was mostly consistent with the dissimilarity of the low-level stimulus images, which was measured with pixel-wise correlations as well as decoding accuracies using a stimulus image itself as input. In contrast, the parietal and premotor areas showed distinct patterns of decoding accuracy when compared with the early visual area: in the aIPS, the highest decoding accuracy was observed for action, followed by mirror-image of hand, then object. In the PMv, significant decoding accuracy was only observed for action. Thus, distinct neural representations of observed images were clearly indicated in the parietal and premotor regions compared with the early visual area. Our results obtained for the early visual area may reflect its retinotopic coordinates representing low-level visual information of input stimuli. In contrast, the parietal and premotor regions that comprise the MNS appear to encode action-dominant representations of observed images that are independent of retinal variations.

The PMv showed the significant decoding accuracy for action. Neurophysiological studies with monkeys indicated that the PMv neurons encode object-directed motor repertories (Rizzolatti et al., 1988). Some of these neurons, called mirror neurons, also showed selectivity to observed actions (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). Since the current study measured activity related only to the observation and not the execution of actions, our results might reflect the activity of both mirror neurons and visual neurons in the PMv (Dinstein et al., 2008a). However, our results may fit well with the idea that action-selective activity of the PMv mirror neuron was invariant regarding effectors of action recognition and responded to actions made with tools as well as with the hand and mouth (Ferrari et al., 2005) or to action-related sound (Kohler et al., 2002) This indicates that effector-independent high-level action representation in the PMv may comprise action vocabulary that subserves invariant action recognition.

The aIPS showed the highest accuracy for action, followed by mirror-image correspondence of hand then object. The aIPS is considered to be a human homologue of AIP in monkeys, whose neurons show selectivity to objects to be grasped as well as to finger configurations (Sakata et al., 1995; Murata et al., 2000). The fMRI studies using adaptation or repetition suppression also revealed selectivity for action types, as well as objects during execution or observation, in the aIPS (Shmuelof and Zohary, 2005; Hamilton and Grafton, 2006; Dinstein et al., 2007; Kroliczak et al., 2008). Because our experimental design did not dissociate right/left visual fields with hand in the stimuli, the observed hand mirror selectivity in the aIPS was confounded with visual field effects. However, the study by Shmuelof and Zohary (2008), which did dissociate the effect of hand with visual field, showed the increased activity for mirror-image of hand in the aIPS and confirmed the hand-selectivity in this region. In agreement with current results, neurophysiological study of monkeys has recently revealed that visuo-tactile bimodal neurons in the IPS, selective for body parts of self, are also activated by vision of others' bodies mostly by way of a mirror-image (Ishida et al., 2010). In contrast, action property (grasp or touch) was not confounded with the visual field effects, because they are presented in the same hemifield. Thus, the visual field effects cannot explain the decoding accuracy pattern for action in the parietal and premotor cortices.

Using data with different features for training and testing of other dimensions except action, we also tested the invariance/robustness of action representation to retinal variations in the parietal and premotor cortex, to decode action (see Fig. 2B for details). Consequently, we observed significantly above-chance decoding accuracy for action for all dimensions in the bilateral aIPS and left PMv, as well as in the early visual area. We considered that the significant decoding accuracy in the early visual area might be due to slight but still significant differences in low-level (pixel-wise) dissimilarity of the stimuli. This, in turn, might lead to a successful decoding of action, while the parietal and premotor cortices encode invariant action representation of observed images. To test this prediction, we then performed another MVPA that would decode action by combining two different features having high dissimilarity indexes of images (MIRR_HAND and SIZE), with the aim of maximally eliminating the effects of low-level image similarity (see Fig. 2C for details). We found a clear distinction in decoding accuracy patterns between the parietal/premotor and early visual area: non-significant decoding accuracy for action in the early visual area, while significant decoding accuracies of action were observed in aIPS as well as in PMv, in agreement with our prediction. This result indicates that action representation for observed images in the parietal and premotor areas is rather robust to variability in input stimuli or retinal images, compared with those in the early visual area. Our current results indicate roles of parietal and premotor regions in representing action-dominant information from observed images which is rather invariant to retinal variations.

Recent study by Dinstein et al. (2008b) showed that three types of observed gestures (rock, paper, and scissors) were successfully decoded from aIPS activity using MVPA. Our result is consistent with their study, as it indicated that observed actions were represented in the parietal cortex. The difference between the current study and this previous work is that Dinstein et al. manipulated only the action property, whereas we independently manipulated different features of images to dissociate low-level visual similarity of stimuli with action selectivity. We were then able to show action selectivity in the aIPS as well as in the PMv, which gave further support to the existence of neural representation of observed actions in these regions. Another difference is that the Dinstein et al.'s study failed to show observed action selectivity in the PMv. This could have been due to use of intransitive gestures, rather than the transitive actions that had been originally used in mirror neuron studies of monkeys.

One caveat of MVPA is that lack of significance in a higher decoding accuracy does not ensure that neurons in local region are totally unselective to manipulated features. Although their neural underpinning is still unclear, decoding accuracy of MVPA is considered to be dependent on a sampling bias caused by spatial patterns of distinct neuronal populations and/or accompanying vasculature units, and spatial resolution of fMRI (Bartels et al., 2008; Gardner, 2010). A part of action-selective mirror neurons in the PMv are also reported to be selective to mirror-images of the observed hand (Gallese et al., 1996), which was not revealed in the PMv, but in the aIPS, by the present study. Given a strong mutual connection between the aIPS and PMv (Matelli et al., 1986; Rizzolatti and Luppino, 2001), neural representation may be continuous, rather than discrete, in the parieto-premotor pathway that subserves action recognition.

It should be noted that our decoding results are mostly limited to the action feature, as other properties (such as perspective and hand) tend to be largely confounded with the visual similarity of the observed images. This is reflected in the decoding accuracy in the early visual area. Regarding a separate pathway of visual recognition for action and for perception (Goodale and Milner, 1992), it should be also noted that the object selectivity in the current study differs from object recognition without action possibly subserved by the ventral visual stream (Riesenhuber and Poggio, 2002). The parietal cortex activity observed in our study may therefore be related to extraction of action-related information regarding the objects to be manipulated (Fagg and Arbib, 1998).

Lastly, significant above-chance decoding accuracy for action was shown in the parietal cortex (aIPS) and premotor cortex (PMv), as well as in the early visual areas, even though we used only visual stimuli. Previous decoding studies mostly targeted the primary visual areas to decode observed images (Haynes and Rees, 2005; Kamitani and Tong, 2005; Miyawaki et al., 2008). Our study therefore suggests the merit of using activation patterns of the higher brain regions, rather than the primary visual areas, to decode invariant information

(e.g., action selectivity in the current study) that is distinct from low-level visual similarity of input images.

In summary, this study used MVPA to investigate the neural representation of observed actions in the parietal and premotor areas comprising the human MNS—or the action observation network. Our results revealed the neural representation of observed actions was encoded in the parietal and premotor cortex and was independent of retinal variations. This may subserve the human capacity for invariant actions recognition of others.

## Acknowledgments

We thank N. Takemura, Y. Shimada, T. Kochiyama, and H. Mizuhara for their technical assistance and helpful comments.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2010.10.043.

## References

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., Zilles, K., 2000. Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable? Neuroimage 11, 66–84.
Bartels, A., Logothetis, N.K., Moutoussis, K., 2008. fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. Trends Neurosci. 31, 444–453.
Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur. J. Neurosci. 13, 400–404.
di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. Exp. Brain Res. 91, 176–180.
Dinstein, I., Hasson, U., Rubin, N., Heeger, D.J., 2007. Brain areas selective for both observed and executed movements. J. Neurophysiol. 98, 1415–1427.
Dinstein, I., Thomas, C., Behrmann, M., Heeger, D.J., 2008a. A mirror up to nature. Curr. Biol. 18, R13–R18.
Dinstein, I., Gardner, J.L., Jazayeri, M., Heeger, D.J., 2008b. Executed and observed movements have different distributed representations in human aIPS. J. Neurosci. 28, 11231–11239.
Eger, E., Ashburner, J., Haynes, J.D., Dolan, R.J., Rees, G., 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. J. Cogn. Neurosci. 20, 356–370.
Fagg, A.H., Arbib, M.A., 1998. Modeling parietal-premotor interactions in primate control of grasping. Neural Netw. 11, 1277–1303.
Ferrari, P.F., Rozzi, S., Fogassi, L., 2005. Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. J. Cogn. Neurosci. 17, 212–226.
Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119 (Pt 2), 593–609.
Gardner, J.L., 2010. Is cortical vasculature functionally organized? Neuroimage 49, 1953–1956.
Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends Neurosci. 15, 20–25.
Grafton, S.T., Arbib, M.A., Fadiga, L., Rizzolatti, G., 1996. Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. Exp. Brain Res. 112, 103–111.
Grezes, J., Decety, J., 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. Hum. Brain Mapp. 12, 1–19.
Grezes, J., Costes, N., Decety, J., 1998. Top down effect of strategy on the perception of human biological motion: a PET investigation. Cogn. Neuropsychol. 15, 553–582.
Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. (Amst.) 107, 293–321.
Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24, 187–203.
Hamilton, A.F., Grafton, S.T., 2006. Goal representation in human anterior intraparietal sulcus. J. Neurosci. 26, 1133–1137.
Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70.
Iacoboni, M., Dapretto, M., 2006. The mirror neuron system and the consequences of its dysfunction. Nat. Rev. Neurosci. 7, 942–951.
Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. Science 286, 2526–2528.
Ishida, H., Nakajima, K., Inase, M., Murata, A., 2010. Shared mapping of own and others' bodies in visuo-tactile bimodal area of the monkey parietal cortex. J. Cogn. Neurosci. 22, 83–96.
Johnson-Frey, S.H., Maloof, F.R., Newman-Norlund, R., Farrer, C., Inati, S., Grafton, S.T., 2003. Actions or hand-object interactions? Human inferior frontal cortex and action observation. Neuron 39, 1053–1058.
Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.
Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.
Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl Acad. Sci. USA 103, 3863–3868.
Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540.
Kroliczak, G., McAdam, T.D., Quinlan, D.J., Culham, J.C., 2008. The human dorsal stream adapts to real actions and 3D shape processing: a functional magnetic resonance imaging study. J. Neurophysiol. 100, 2627–2639.
Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.
Matelli, M., Camarda, R., Glickstein, M., Rizzolatti, G., 1986. Afferent and efferent projections of the inferior area 6 in the macaque monkey. J. Comp. Neurol. 251, 281–298.
Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.A., Morito, Y., Tanabe, H.C., Sadato, N., Kamitani, Y., 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60, 915–929.
Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI–an introductory guide. Soc. Cogn. Affect. Neurosci. 4, 101–109.
Murata, A., Gallese, V., Luppino, G., Kaseda, M., Sakata, H., 2000. Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. J. Neurophysiol. 83, 2580–2601.
Nishitani, N., Hari, R., 2000. Temporal dynamics of cortical representation for action. Proc. Natl Acad. Sci. USA 97, 913–918.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45, S199–S209.
Riesenhuber, M., Poggio, T., 2002. Neural mechanisms of object recognition. Curr. Opin. Neurobiol. 12, 162–168.
Rizzolatti, G., Luppino, G., 2001. The cortical motor system. Neuron 31, 889–901.
Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., Matelli, M., 1988. Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. Exp. Brain Res. 71, 491–507.
Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Brain Res. Cogn. Brain Res. 3, 131–141.
Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. 2, 661–670.
Sakata, H., Taira, M., Murata, A., Mine, S., 1995. Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. Cereb. Cortex 5, 429–438.
Sawamura, H., Orban, G.A., Vogels, R., 2006. Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. Neuron 49, 307–318.
Shmuelof, L., Zohary, E., 2005. Dissociation between ventral and dorsal fMRI activation during object and action recognition. Neuron 47, 457–470.
Shmuelof, L., Zohary, E., 2008. Mirror-image representation of action in the anterior parietal cortex. Nat. Neurosci. 11, 1267–1269.
Tanaka, S., Inui, T., 2002. Cortical involvement for action imitation of hand/arm postures versus finger configurations: an fMRI study. NeuroReport 13, 1599–1602.