Patterns of response to visual scenes are linked to the low-level properties of the image

CrossMark

David M. Watson, Tom Hartley, Timothy J. Andrews *

Department of Psychology and York Neuroimaging Centre, University of York, York YO10 5DD, United Kingdom

ABSTRACT

Scene-selective regions in the brain play an important role in the way that we navigate through our visual environment. However, the principles that govern the organization of these regions are not fully understood. For example, it is not clear whether patterns of response in scene-selective regions are linked to high-level semantic category or to low-level spatial structure in scenes. To address this issue, we used multivariate pattern analysis with fMRI to compare patterns of response to different categories of scenes. Although we found distinct patterns of neural response to each category of scene, the magnitude of the within-category similarity varied across different scenes. To determine whether this variation in the categorical response to scenes could reflect variation in the low-level image properties, we measured the similarity of images from each category of scene. Although we found that the low-level properties of images from each category were more similar to each other than to other categories of scenes, we also found that the magnitude of the within-category similarity varied across different scenes. Finally, we compared variation in the neural response to different categories of scenes with corresponding variation in the low-level image properties. We found a strong positive correlation between the similarity in the patterns of neural response to different scenes and the similarity in the image properties. Together, these results suggest that categorical patterns of response to scenes are linked to the low-level properties of the images.

© 2014 Elsevier Inc. All rights reserved.

Introduction

The ability to perceive and recognize different visual scenes is essential for spatial navigation in the world. Although real-world scenes can be incredibly complex and heterogeneous, human observers are able to reliably recognize and categorize images of scenes even when the images are shown briefly (Greene and Oliva, 2009; Joubert et al., 2007; Potter, 1975). These studies have been taken to suggest that the initial perception of natural images is based on the global, visual properties – the gist – of the scene (Greene and Oliva, 2009; Oliva and Torralba, 2001).

Neuroimaging studies have found a number of regions of the human brain that respond selectively to visual scenes. Damage to these regions often leads to impairments that are specific to scene perception and spatial navigation (Aguirre and D'Esposito, 1999; Mendez and Cherrier, 2003). The parahippocampal place area (PPA) is a region of the posterior parahippocampal gyrus that displays preferential activity to images of scenes over and above images of objects and faces (Aguirre and D'Esposito, 1997; Epstein and Kanwisher, 1998). Other place selective regions include the retrosplenial complex (RSC) located immediately superior to the PPA and the transverse occipital sulcus (TOS) or occipital

place area (OPA) on the lateral surface of the occipital lobe (Dilks et al., 2013; Epstein, 2008).

The spatial layout of different categories of scenes can vary quite considerably (Torralba and Oliva, 2003). Although neuroimaging studies using univariate analyses have reported comparable levels of response to scenes as diverse as natural landscapes, cityscapes and indoor scenes in scene-selective regions (Aguirre and D'Esposito, 1997; Epstein and Kanwisher, 1998), more recent studies using multivariate analyses have found distinct patterns of response in these regions to different categories of scene (Walther et al., 2009, 2011). Interestingly, these patterns of neural response have also been shown to correlate with patterns of behavioral response, but not with the low-level image properties of the images (Walther et al., 2009). This suggests that there is a dissociation between the perceptual categorization of scenes and their underlying image properties. However, this conclusion has been challenged by other studies that have suggested that the patterns of response in scene-selective regions are better explained by the spatial layout of the scene rather than by semantic category (Kravitz et al., 2011; Park et al., 2011). Although these studies are not explicit about how the image properties of the scene are linked to the patterns of neural response, work in computer vision indicates that semantically-distinct scene categories can be identified on the basis of their characteristic low-level image statistics. For example, the GIST descriptor can be used to accurately classify different scene

categories and derive spatial properties such as openness (Torralba and Oliva, 2003).

Our aim was to determine whether categorical patterns of brain activity within scene-selective regions are linked to the low-level properties of the images from each category of scene. To address this issue, we measured the pattern of response to different categories of scenes using fMRI. Next, we asked how similar the low-level properties of images from each category were to each other. Finally, we asked whether differences in the categorical response to different visual scenes might be due to variation in low-level image properties. Our prediction was that, if low-level visual properties are linked to categorical patterns of response in these regions, then scene categories with similar image statistics should elicit correspondingly similar patterns of brain activity.

Methods

Participants

Twenty participants took part in Experiment 1 (9 males, mean age: 24.5) and 20 participants took part in Experiment 2 (9 males, mean age: 25.2). All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision.

Stimuli

All images were taken from the LabelMe scene database (http://cvcl. mit.edu/database.htm); (Oliva and Torralba, 2001) and presented in greyscale at a resolution of 256 × 256 pixels. All further image processing was performed in MATLAB v7.10 (http://www.mathworks.co.uk/). Fourier-scrambled images were created by randomizing the phase of each 2-dimensional frequency in the original image while keeping the power of the components constant. For each experiment, the luminance histogram of images across all conditions was equated using the SHINE toolbox (Willenbockel et al., 2010).

Experimental design

In Experiment 1 and Experiment 2, participants viewed images from 5 stimulus conditions. Fig. 1 shows examples of images taken from the stimulus conditions used in both experiments. The stimulus conditions in Experiment 1 included: (1) cityscapes, (2) indoor scenes, (3) natural landscapes, (4) mixed (interleaved images from conditions 1–3) and (5) scrambled (Fourier scrambled versions of the mixed condition). The stimulus conditions in Experiment 2 included: (1) coast, (2) forest, (3) mountains, (4) mixed (interleaved images from conditions 1–3) and (5) scrambled (Fourier scrambled versions of the mixed condition). In each experiment, images from each condition were presented in a block design with 9 images in each block. Each image was presented for 850 ms followed by a 150 ms black screen. Each stimulus block was separated by a 9 s period in which a fixation cross was superimposed on a gray screen that was equal in mean luminance to the scene images. Each condition was repeated 8 times in a counterbalanced block design, giving a total of 40 blocks. To maintain attention throughout the scan session, participants performed a one-back task in which one image from each block was repeated.

Imaging parameters

All scanning was conducted at the York Neuroimaging Centre (YNiC) using a GE 3 Tesla HDx Excite MRI scanner. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage, radio-frequency coil tuned to 127.7 MHz. Data were collected from 240 volumes each comprising 38 contigual axial slices via a gradient-echo EPI sequence (TR = 3 s, TE = 32 ms, FOV = 28.8 × 28.8 cm, matrix size = 128 × 128, voxel dimensions = 2.25 × 2.25 mm, slice
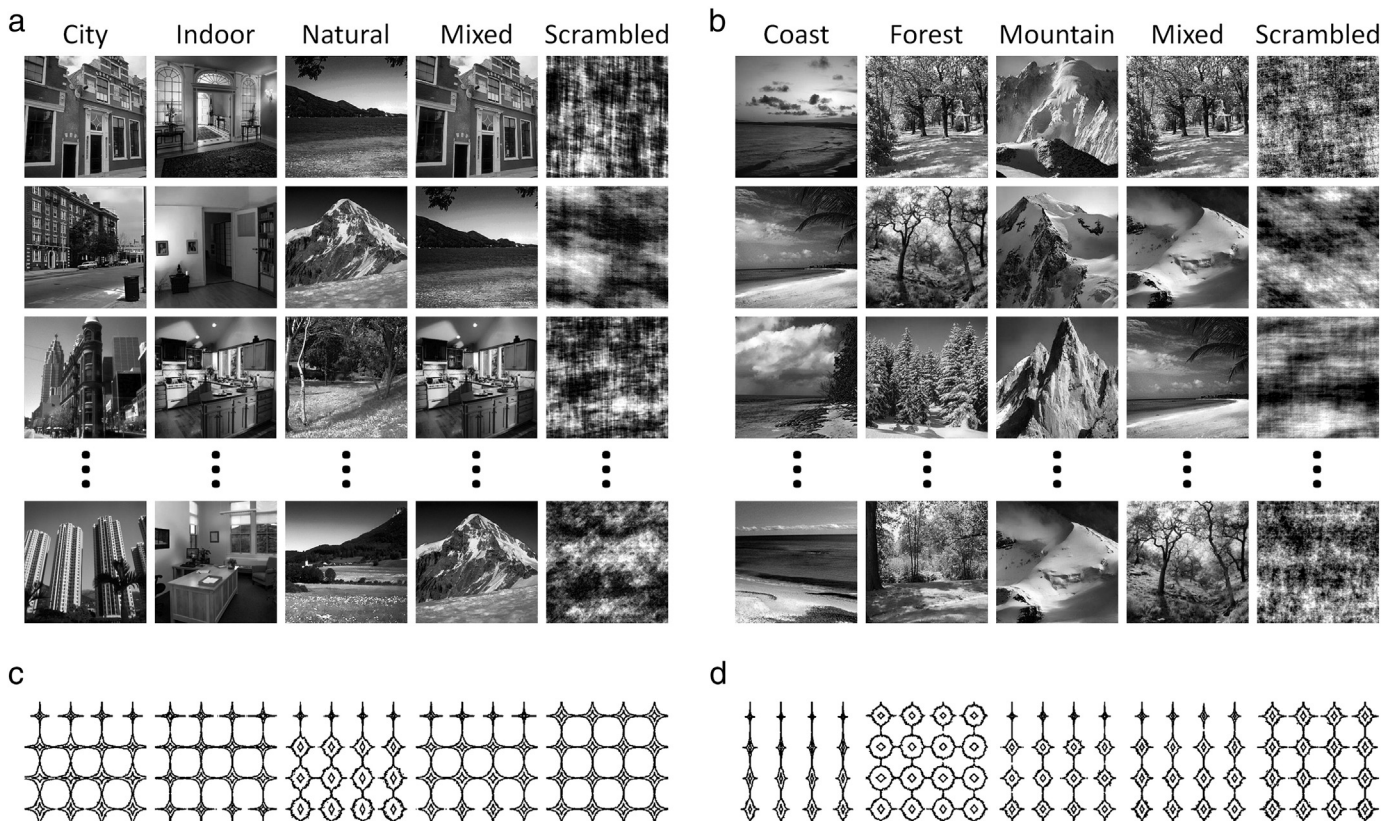

Fig. 1. Examples of images from each experimental condition in (a) Experiment 1 and (b) Experiment 2. Category average contour plots of Fourier power spectra within 4 × 4 windows are shown for (c) Experiment 1 and (d) Experiment 2.

thickness = 3 mm, flip angle = 90°). Visual stimuli were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant with images subtending approximately 9.5° of visual angle.

fMRI analysis

Univariate analysis of the fMRI data was performed with FEAT v 5.98 (http://www.fmrib.ox.ac.uk/fsl). All analyses were performed separately for each experiment in the manner described below. In all scans the initial 9 s of data was removed to reduce the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL) was applied followed by temporal high-pass filtering (Gaussian-weighted least-squares straight line fitting, $\sigma = 50$ s). Spatial smoothing (Gaussian) was applied at 6 mm (FWHM). Individual participant data were entered into a higher-level group analysis using a mixed-effects design (FLAME, http://www.fmrib.ox.ac.uk/fsl). Functional data were first registered to a high-resolution T1-anatomical image and then onto the standard MNI brain (ICBM152). A scene-selective region of interest was defined by the contrast of mixed > scrambled. The resulting group statistical maps were thresholded at $Z > 2.3$. The thresholded statistical maps were then combined across experiments to generate a single scene-selective region of interest (ROI) used for subsequent MVPA analyses across both experiments (Suppl. Fig. 1). We also generated a more restrictive ROI constrained to the scene-selective regions (parahippocampal place area (PPA), retrosplenial cortex (RSC) and the transverse occipital sulcus (TOS) or occipital place area (OPA)) that have been reported in previous fMRI studies (Epstein and Kanwisher, 1998; Grill-Spector, 2003; Maguire, 2001). This ROI was defined as follows: firstly, group mixed > scrambled statistical maps were averaged across the experiments. Next, seed points were defined at the peak voxels within this average statistical map for each region (PPA, RSC, TOS/OPA) in each hemisphere. The peak voxels of the ROIs had similar coordinates to those found in previous studies (Suppl. Table 1). For a given seed, a flood fill algorithm was used to identify a cluster of spatially contiguous voxels around that seed which exceeded a given threshold. This threshold was in turn iteratively adjusted till a cluster size of 500 voxels was achieved. This process was then repeated for each seed. Clusters for each region were combined across hemispheres to yield 3 ROIs each comprising 1000 voxels. Additionally, a single ROI combining all clusters across both hemispheres was defined. MNI coordinates of the seeds and corresponding thresholds are given in Table 1. All further analyses were restricted to these regions of interest.

Parameter estimates from the univariate analysis were normalized by subtracting the response to the mixed condition. Pattern analyses were then performed using the PyMVPA toolbox http://www.pymvpa.org/; (Hanke et al., 2009). Fig. 2 illustrates the method for determining the reliability of these neural patterns within and across subjects. To determine the reliability of the data within individual participants, the parameter estimates for each scene condition were correlated across odd (1, 3, 5, 7) and even (2, 4, 6, 8) blocks across all voxels in the scene-selective region (Haxby et al., 2001). The individual participant (IP) analysis was complemented by a group analysis, to determine the

reliability of the pattern across participants. We used a leave-one-participant-out (LOPO) method (Poldrack et al., 2009; Shinkareva et al., 2008) in which the parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each scene condition in each voxel across the scene-selective region. This LOPO process was repeated such that every participant was left out of a group analysis once. For each LOPO iteration, the normalized patterns of response to each stimulus condition were correlated between the group and the participant that was left-out. This allowed us to determine whether there are reliable patterns of response that are consistent across individual participants. A Fisher's z-transformation was applied to the within-category and between-category correlations prior to further statistical analyses. For each category, the within-category and the average of the between-category correlations were calculated. These were entered into 3×2 repeated ANOVAs with the scene category (Experiment 1: city, indoor, natural; Experiment 2: coast, forest, mountain) and comparison (within, between) as the main factors. If neural response patterns to a given category can be distinguished from those to other categories, a significant main effect of comparison showing greater within- than between-category correlations would be expected. In order to obtain a measure of the decoding accuracy of our MVPA analyses, parameter estimates from the univariate analysis were also submitted to a k-nearest neighbor (kNN) classifier ($k = 1$) using correlation as the distance measure.

In addition to the ROI analyses listed above we also performed whole-brain searchlight analyses (Kriegeskorte et al., 2006). A spherical ROI of radius 6 mm was defined, and MVPA performed as described above. The average within- minus between-category correlation difference across categories was then assigned to the central voxel of the sphere, and the process repeated iterating the sphere over the whole-brain volume. A higher-level analysis using a mixed-effects design (FLAME) was used to determine whether the value at each voxel differed significantly from zero across individuals/LOPO-iterations. The resulting group statistical maps were thresholded at $Z > 2.3$ with a cluster-correction of $p < .05$ applied.

Image properties

Finally, we asked whether the patterns of neural response in Experiment 1 and 2 could be explained by the image statistics of the visual scenes. The image statistics of the scene images were computed using the GIST descriptor (http://people.csail.mit.edu/torralba/code/spatialenvelope/; Oliva and Torralba, 2001). First, each image is passed through a series of Gabor filters across 8 orientation and 4 spatial frequencies. This generates 32 filtered images. Next, each image is divided into a 4×4 grid giving 16 windows. The mean intensity is measured in each window. This generates a vector of 512 (32×16) values – the GIST descriptor – which represents the image in terms of the spatial frequencies and orientations present at different positions across the image. A schematic illustration of the calculation is given in Fig. 3. In order to determine the similarity between individual scenes and the average of each scene category, GIST descriptors were correlated between each image and the average descriptor derived for each scene condition. This cross-validation procedure was used to determine how similar each image was to the average of its own category and to the other categories. Similarity with the neural response was determined by correlating the average GIST correlations matrix with the average MVPA correlations matrix. In order to assess the significance of this relationship, a simple regression analysis was performed using the average GIST correlations matrix as the regressor, and the corresponding MVPA correlation matrices concatenated across individuals/LOPO iterations as the outcomes. If the GIST correlations matrix is able to explain a significant amount of the variance in the corresponding MVPA correlation matrices, the model regression coefficient (β) can be expected to be significantly greater than zero. All regressor and

Table 1
MNI mm coordinates and thresholds of standard place-selective (PPA, RSC, TOS/OPA) clusters.

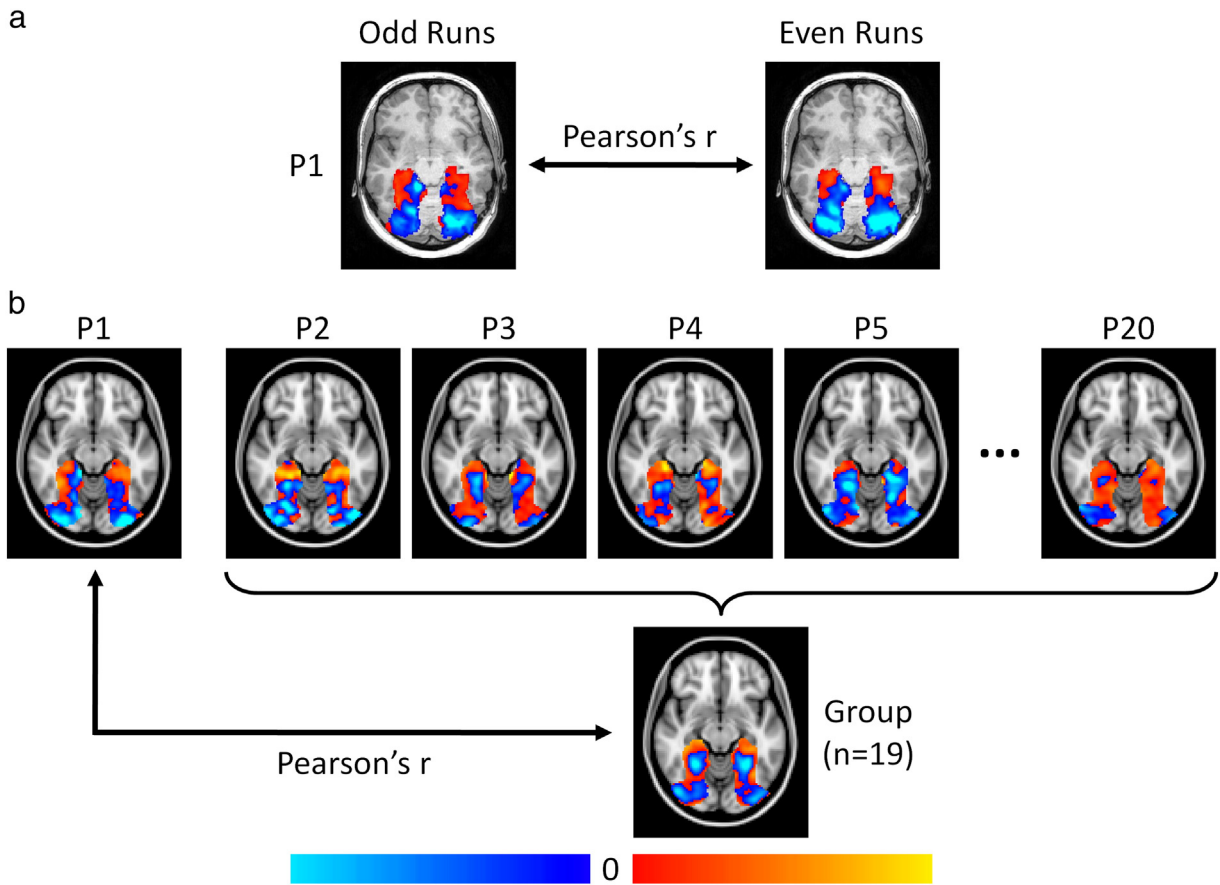| Region	Hemisphere	x	y	Z	Threshold (Z)
PPA	L	−26	−48	−14	4.23
	R	30	−42	−16	4.24
RSC	L	−16	−60	4	3.58
	R	18	−56	6	3.77
TOS/OPA	L	−42	−84	20	3.52
	R	32	−88	12	3.28

a

b

Fig. 2. Schematic diagram of pattern analysis procedures. (a) Individual participant analyses correlated neural patterns across odd and even runs of the stimulus presentation. (b) Group analyses compared individual patterns of response with the group pattern of response derived from all participants except that individual (LOPO). In both analyses this process is then repeated across all participants/LOPO iterations for all conditions.
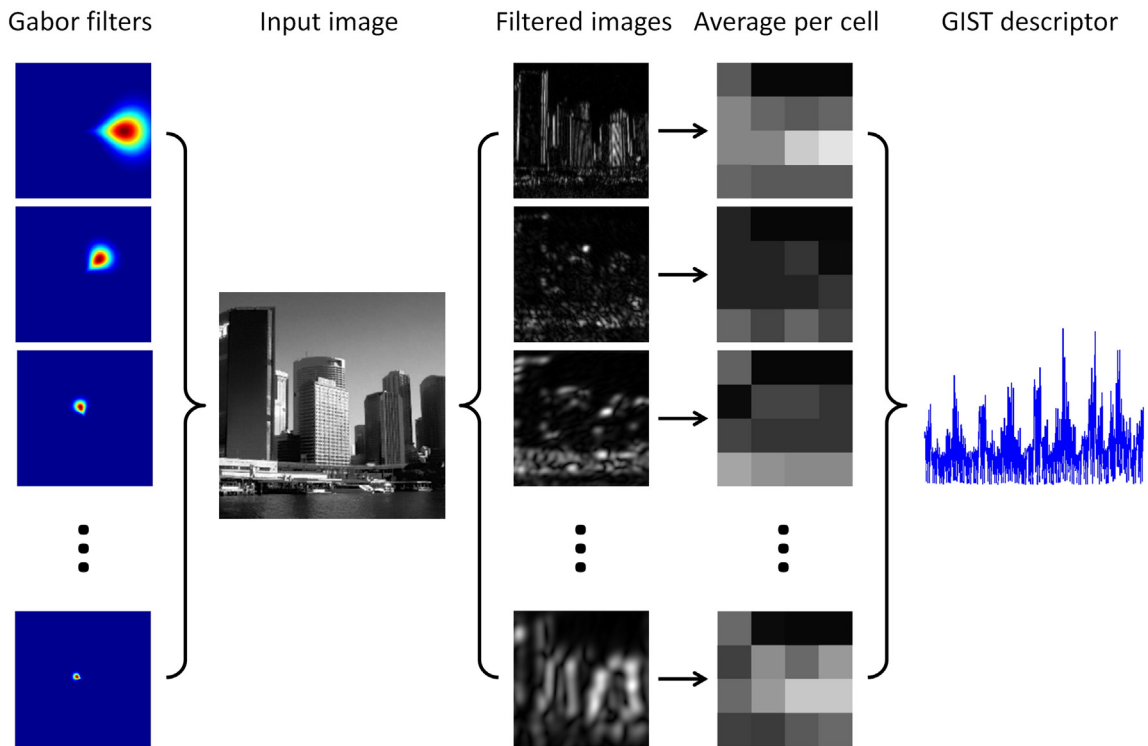

Fig. 3. Schematic illustration of the calculation of a GIST descriptor for an example image. A series of Gabor filters across 8 orientations and 4 spatial frequencies is applied to the image. Each of the resulting 32 filtered images is then windowed by a 4 × 4 grid and the pixel intensities within each grid cell averaged together. Each grid cell thus represents the degree to which that window of the image is preserved by a Gabor filter at a given orientation and spatial frequency. The final GIST descriptor is a vector of 512 values yielded by concatenating these 16 cells across the 32 filtered images.

outcome variables were *Z*-scored prior to the regression analysis, such that all regression coefficients are given in standardized units. The image statistics of the scene images were also computed using pixelwise correlation of luminance values (cf. Walther et al., 2009). This provided us with a more basic image-based measure with which to compare with GIST descriptor.

Results

Experiment 1

In the first experiment, we measured the patterns of response to different categories of visual scenes: city, indoor and natural. Fig. 4 shows the normalized group response to city, indoor, and natural categories across the scene-selective region. Responses above the mean are shown in red and responses below the mean are shown in blue. Each category of scene had a distinct pattern of response, which was similar in appearance across the two cerebral hemispheres. Similar patterns were evident in individual participants (Suppl. Fig. 2).

Correlation based MVPA methods (Haxby et al., 2001) were used to measure the reliability of the neural response to these different categories of scene within individual participants (IP). Fig. 5A shows a matrix of the correlations for the within- and between-category correlations. A 3×2 repeated measures ANOVA with Scene (city, indoor, natural) and Comparison (within, between) as the main factors showed a significant main effect of Comparison ($F(1,19) = 11.6$, $p = .003$), showing that within-category correlations were higher than between-category correlations. However, there was no significant interaction between Scene * Comparison ($F(2,38) = 1.7$, $p = .196$). A kNN classifier revealed that the decoding accuracy across categories was 46.7%, $p = .008$ (chance = 33%). A similar classification was evident when the ROI was restricted to all the standard scene-selective regions (combined PPA + RSC + TOS: 58.3%, $p < .001$). Suppl. Fig. 3A shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed accuracies significantly above chance in

PPA and TOS, but not RSC (PPA: 64.1%, $p < .001$; RSC: 42.5%, $p = .09$; TOS: 53.3%, $p = .006$).

We then determined the extent to which these patterns were consistent across participants using the LOPO method (see Methods). Fig. 5B shows the correlation matrix using the LOPO method. There was a significant main effect of Comparison ($F(1,19) = 90.8$, $p < .001$), which was due to higher within-category compared to between-category correlations. There was also a significant Scene * Comparison interaction ($F(2,38) = 3.9$, $p = .028$). This interaction was due to larger differences in within- versus between-category comparisons for the indoor and natural conditions compared to the city condition (city: $p = .004$, indoor: $p < .001$, natural: $p < .001$). A kNN classifier revealed a decoding accuracy across categories of 72.5%, $p < .001$ (chance = 33%). A similar classification was evident when the ROI was restricted to the standard scene-selective regions (combined PPA + RSC + TOS: 59.2%, $p < .001$); Suppl. Fig. 3B shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed a similar pattern of results (PPA: 59.1%, $p < .001$; RSC: 50.8%, $p = .003$; TOS: 54.1%, $p = .002$).

To address the spatial scale of the patterns we repeated the LOPO and IP analyses with no spatial smoothing. Consistent with a coarser scale representation, we found a similar pattern of results (Suppl. Fig. 4). We then repeated the LOPO and IP analyses using a whole-brain searchlight paradigm. Consistent with the previous analysis, we found that the majority of significant spheres clustered around the scene selective cortices defined by the ROI (Suppl. Fig. 5).

Next, we used the GIST descriptor to measure the statistics of each image used in the fMRI experiment. Fig. 5C shows the within- and between-category correlations in image properties for different categories of visual scenes. We found higher within-category than between-category correlations (city: $p < .001$, indoor: $p < .001$, natural: $p < .001$). To determine whether there was a relationship between image properties of the stimuli and patterns of brain activity, the GIST correlations for each combination of scene were then correlated with the corresponding neural correlations for both the IP and LOPO.
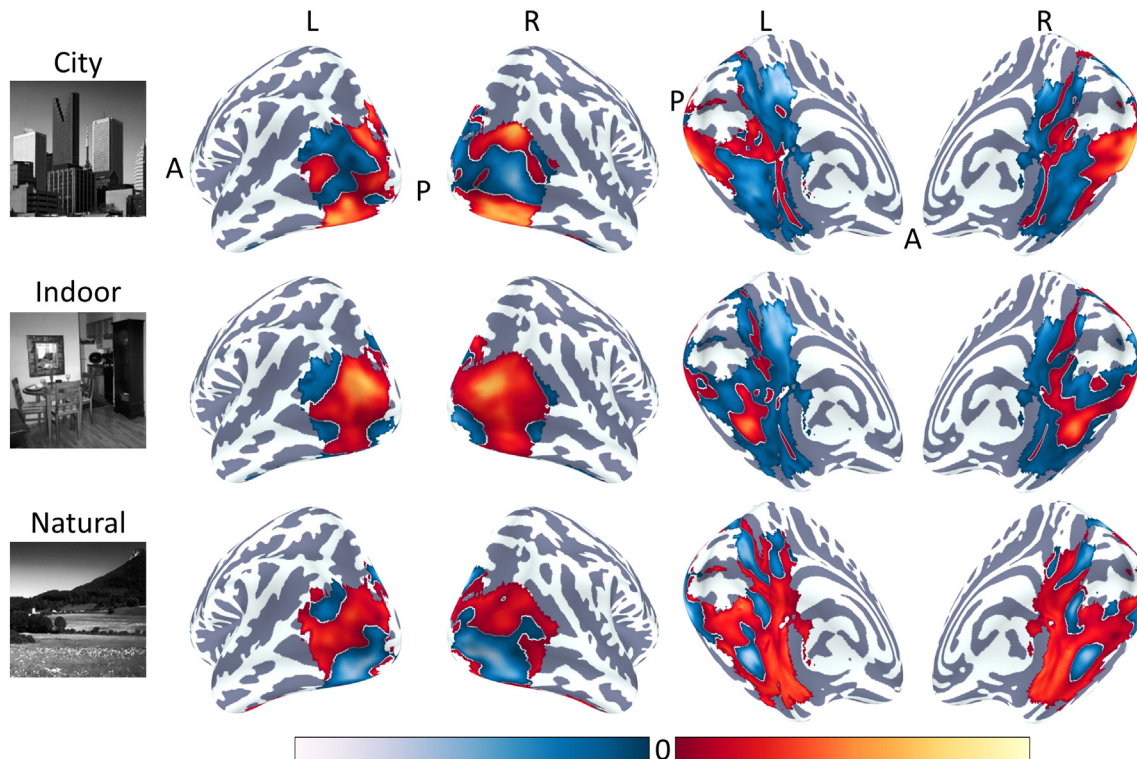

Fig. 4. Experiment 1: Group patterns of response to city, indoor, and natural conditions on lateral (leftmost panels) and ventro-medial surfaces (rightmost panels). Patterns are restricted to regions defined by the response of mixed scenes > scrambled scenes. Red and blue colors indicate normalized values above and below the mean respectively.
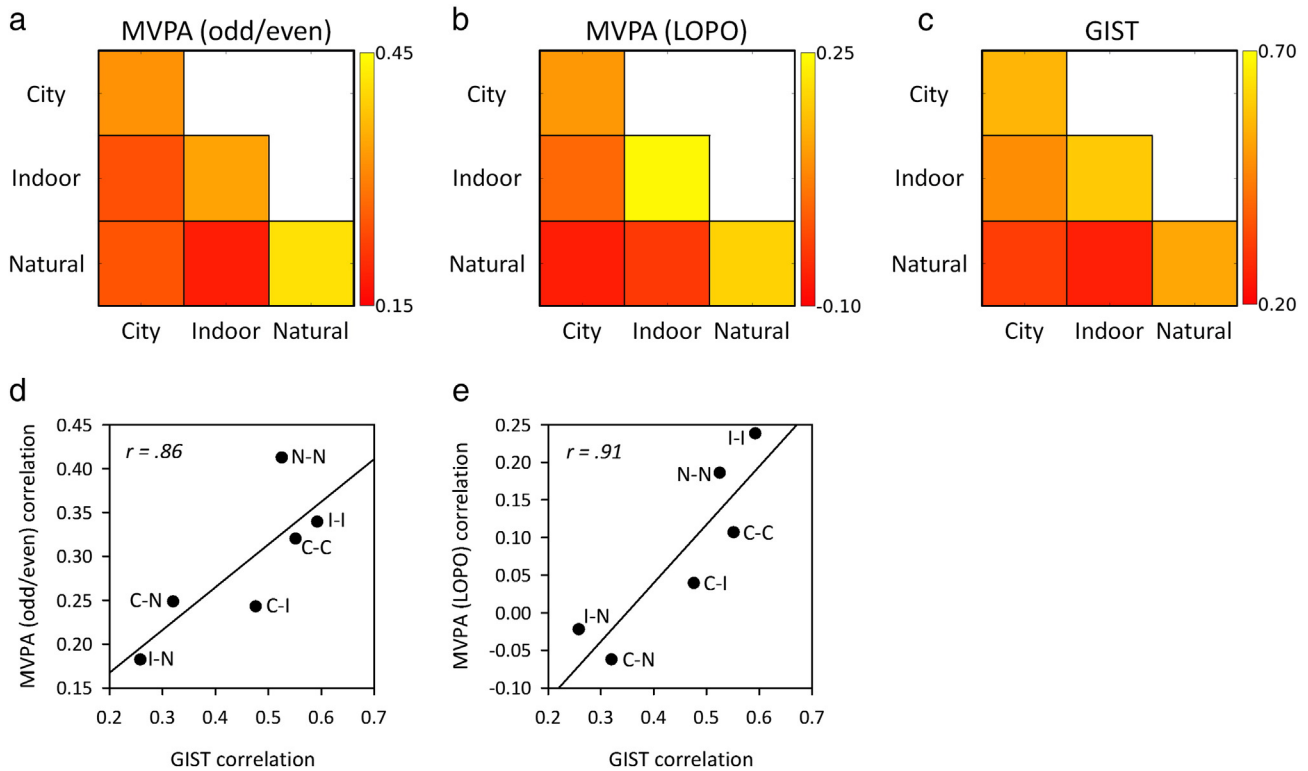
Fig. 5. Experiment 1: Relationship between fMRI response and low-level image properties. Within- and between-category correlations for city, indoor, and natural conditions as determined by the individual participant (a) and LOPO (b) MVPA analyses, and by the GIST image descriptor (c). Scatter-plots (d and e) showing strong positive correlations of the correlation matrices in (a) and (b) with (c) respectively.

Figs. 5D–E show the relationship between the similarity in image properties and the similarity in the pattern of response across different scenes. Strong positive correlations were evident for both the IP ($r = .86$) and LOPO analyses ($r = .91$). The significance of this relationship across participants or LOPO iterations was assessed using a simple regression analysis. The image properties significantly predicted the neural response in the IP ($\beta = .28$, $p = .001$) and LOPO analyses ($\beta = .57$, $p < .001$). A similar pattern of results was evident when the ROI was restricted to the standard scene-selective regions (combined PPA + RSC + TOS) for both the IP ($r = .78$, $\beta = .32$, $p < .001$) and LOPO analyses ($r = .58$, $\beta = .33$, $p < .001$; Suppl. Figs. 3C–D. Splitting this ROI into its constituent regions produced a similar pattern of results for the IP analyses (PPA: $r = .76$, $\beta = .46$, $p < .001$; RSC: $r = .75$, $\beta = .21$, $p = .022$; TOS: $r = .73$, $\beta = .21$, $p = .024$) and LOPO analyses (PPA: $r = .64$, $\beta = .42$, $p < .001$; RSC: $r = .55$, $\beta = .17$, $p = .065$; TOS: $r = .78$, $\beta = .26$, $p = .004$).

We next repeated our analysis using pixel correlations as a measure of image properties. Pixel correlations did not significantly predict the neural response for the IP analysis ($r = .12$, $\beta = .04$, $p = .653$). However, a significant relationship was found for the LOPO analysis ($r = .55$, $\beta = .34$, $p < .001$). The pixel correlations were also poor predictors of the neural responses in the standard scene-selective regions for the IP analyses (combined PPA + RSC + TOS: $r = .27$, $\beta = .11$, $p = .221$; PPA: $r = .01$, $\beta = .003$, $p = .973$; RSC: $r = .36$, $\beta = .10$, $p = .281$; TOS: $r = .31$, $\beta = .09$, $p = .339$) and LOPO analyses (combined PPA + RSC + TOS: $r = .17$, $\beta = .10$, $p = .298$; PPA: $r = .34$, $\beta = .12$, $p = .120$; RSC: $r = .24$, $\beta = .22$, $p = .017$; TOS: $r = .25$, $\beta = .08$, $p = .361$). Thus, the pixel correlations measure was outperformed by the GIST descriptor.

Experiment 2

In the second experiment, we compared the patterns of responses to different types of natural landscapes: coasts, forests and mountains.

Fig. 6 shows the normalized group responses to coast, forest, and mountain scenes within scene-selective regions. Again, each category of scene had a distinct pattern of response, which was similar in appearance across the two cerebral hemispheres. Similar patterns of response can be found in the individual participants (Suppl. Fig. 6). The reliability of these patterns of response was measured using the LOPO and IP methods. A 3 × 2 repeated measures ANOVA with Scene (coast, forest, mountain) and Comparison (within, between) as the main factors was used to test statistical significance.

First, we performed the pattern analyses for individual participants (IP). The correlation between different scene categories is shown in Fig. 7A. There was a significant main effect of Comparison ($F(1,19) = 33.3$, $p < .001$), revealing significantly higher within-category compared to between-category correlations. However, there was not a significant Scene * Comparison interaction ($F(2,38) = 2.7$, $p = .079$). A kNN classifier obtained mean decoding accuracy across all scene categories of 53.3%, $p = .001$ (chance = 33%). A similar classification was evident when the ROI was restricted to the standard scene-selective regions (combined PPA + RSC + TOS: 55.8%, $p < .001$). Suppl. Fig. 7A shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed accuracies significantly above chance in PPA and TOS, but not RSC (PPA: 56.7%, $p < .001$; RSC: 37.5%, $p = .362$; TOS: 52.5%, $p = .002$).

To determine whether the pattern of response was consistent across participants, we repeated the analysis using the LOPO method (Fig. 7B). There was a significant main effect of Comparison ($F(1,19) = 114.4$, $p < .001$) and a significant Scene * Comparison interaction ($F(2,38) = 18.18$, $p < .001$). This interaction was due to larger within- versus between-category comparisons for the coast and mountain conditions compared to the forest condition (coast: $p < .001$, forest: $p = .009$, mountain: $p < .001$). A kNN classifier obtained mean decoding accuracy across all scene categories of 67.5%, $p < .001$ (chance = 33%). A similar classification was evident when the ROI was restricted to the standard scene-selective regions (combined PPA + RSC + TOS:
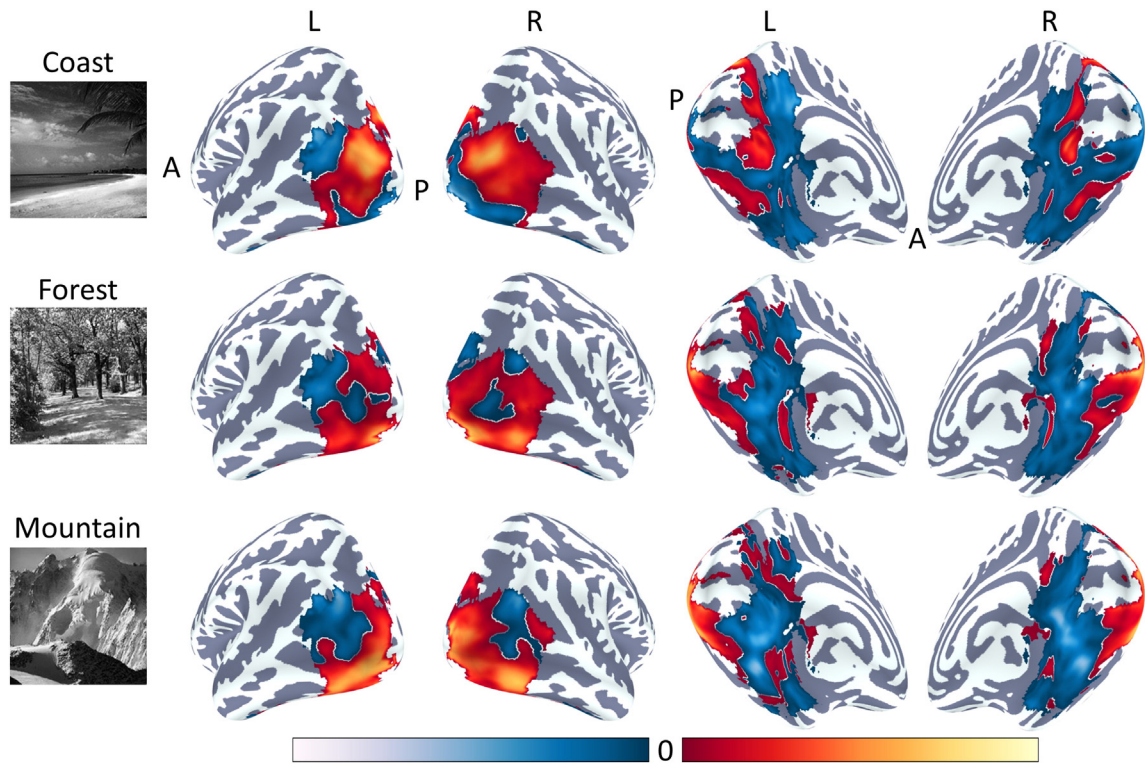
Fig. 6. Experiment 2: Group patterns of response to coast, forest, and mountain conditions on lateral (leftmost panels) and ventro-medial surfaces (rightmost panels). Patterns are restricted to regions defined by the response of mixed scenes > scrambled scenes. Red and blue colors indicate normalized values above and below the mean respectively.

50.8%, $p < .001$). Suppl. Fig. 7B shows the corresponding correlations matrix for this region. Splitting this ROI into its constituent regions revealed a similar pattern of results (PPA: 49.2%, $p = .002$; RSC: 50.0%, $p = .002$; TOS: 49.2%, $p = .002$).

To address the spatial scale of the patterns we repeated the LOPO and IP analyses with no spatial smoothing. Consistent with a coarser scale representation, we found a similar pattern of results (Suppl. Fig. 8). To determine the extent to which our findings generalize to
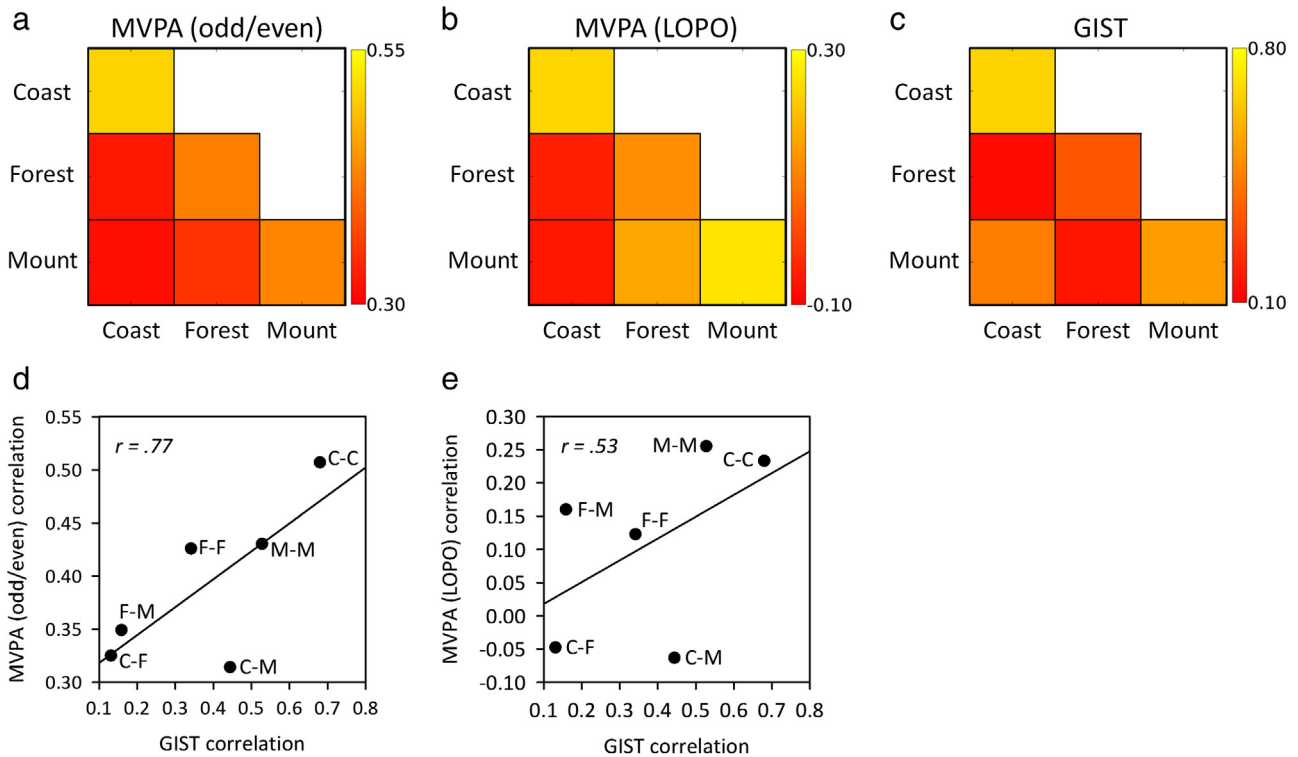

Fig. 7. Experiment 2: Relationship between fMRI response and low-level image properties. Within- and between-category correlations for coast, forest, and mountain conditions as determined by the individual participant (a) and LOPO (b) MVPA analyses, and by the GIST image descriptor (c). Scatter-plots (d and e) showing strong positive correlations of the correlation matrices in (a) and (b) with (c) respectively.

regions outside the ROI, the LOPO and IP analyses were repeated using a whole-brain searchlight paradigm. Significant spheres fell within the scene-selective ROI, particularly along the lateral regions, that included the TOS, and along medial regions that included the PPA and RSC. Suppl. Fig. 9 shows the resulting searchlight group-average statistical maps.

Next, we used the GIST description to measure the statistics of each image used in the fMRI experiment. Fig. 7C shows the within- and between-category correlations in image properties for different categories of visual scenes. We found higher within-category than between-category correlations (coast: $p < .001$, forest: $p < .001$, mountain: $p < .001$). To determine whether there was a relationship between image properties of the stimuli and patterns of brain activity, the GIST correlations for each combination of scene were then correlated with the corresponding neural correlations for both the IP and LOPO analyses. Figs. 7D and E show the relationship between the similarity in image properties and the similarity in the pattern of response across different scenes for the IP and LOPO analyses. Positive correlations were evident for both the IP ($r = .77$) and LOPO analyses ($r = .53$). The significance of this relationship across participants/LOPO iterations was assessed using a simple regression analysis. The image properties significantly predicted the neural response in the IP ($\beta = .27$, $p = .003$) and LOPO analyses ($\beta = .36$, $p < .001$). A similar pattern of results was evident when the ROI was restricted to the standard scene-selective regions (combined PPA + RSC + TOS) for both the IP ($r = .85$, $\beta = .27$, $p = .003$) and LOPO analyses ($r = .84$, $\beta = .37$, $p < .001$); Suppl. Figs. 7C and D. When the scene-selective ROI was split into its constituent regions, for the IP analysis the relationship between image properties and fMRI response was significant for the PPA and TOS, but not for the RSC (PPA: $r = .70$, $\beta = .26$, $p = .004$; RSC: $r = .49$, $\beta = .08$, $p = .394$; TOS: $r = .91$, $\beta = .32$, $p < .001$). The LOPO analysis showed a significant relationship for the TOS and RSC, but not in the PPA (PPA: $r = .32$, $\beta = .13$, $p = .172$; RSC: $r = .56$, $\beta = .23$, $p = .012$; TOS: $r = .90$, $\beta = .24$, $p = .008$).

We next repeated our analysis using pixel correlations as a measure of image properties. The pixel correlations significantly predicted the neural response for the IP ($r = .70$, $\beta = .24$, $p = .008$) but not the LOPO analyses ($r = .23$, $\beta = .16$, $p = .084$). When the ROI was restricted to the standard scene-selective regions, a more variable pattern of results was observed: IP analyses (combined PPA + RSC + TOS: $r = .72$, $\beta = .22$, $p = .014$; PPA: $r = .59$, $\beta = .22$, $p = .014$; RSC: $r = .43$, $\beta = .07$, $p = .450$; TOS: $r = .76$, $\beta = .27$, $p = .003$) and LOPO analyses (combined PPA + RSC + TOS: $r = .69$, $\beta = .31$, $p = .001$; PPA: $r = .09$, $\beta = .04$, $p = .672$; RSC: $r = .38$, $\beta = .16$, $p = .088$; TOS: $r = .83$, $\beta = .22$, $p = .014$). Although pixel correlations accounted for significant variance in the similarity of neural responses in some of the ROIs and analyses, performance was typically inferior to that of the GIST descriptor.

Discussion

The aim of this study was to understand the principles that underlie the organization of scene-selective regions of the human brain. We found that the patterns of response to images from the same scene category were more similar than the patterns of response from different categories of scene. However, there were differences in the magnitude of both the within- and between-category correlations. Next, we investigated the extent to which this variation in the categorical pattern of response to different scenes could be explained by systematic differences in image properties. We found a strong, linear relationship between the pattern of neural response in scene-selective regions and the image statistics of the scenes.

Our results show that the within-category correlations in fMRI response to scenes were higher than the between-category correlations. These results are consistent with previous neuroimaging studies that have used pattern classification techniques to show distinct patterns of response to different categories of scene (Walther et al., 2009,

2011). However, our results also show that there was marked variation in the capacity of MVPA to distinguish different categories of real-world scenes. In Experiment 1, although we found distinct patterns of neural response to different categories of scenes, the pattern of response to natural landscapes was more distinct than to cityscapes or indoor scenes. In Experiment 2, we asked whether the patterns of response in scene-selective regions could discriminate between more subtle differences in scene type using different types of natural landscapes (coasts, forests, mountains). The results again showed that, although within-category responses were higher than between-category responses, but that there were also differences in the patterns of response to different types of natural scenes. For example, coastal scenes could be accurately distinguished from other scene categories on the basis of the pattern of brain activity they evoked, but the pattern of response to forests was often confused with the responses to the other scenes.

The variability in the ability of the pattern of response to discriminate different scenes suggests that factors other than category membership may contribute to the organization of scene-selective regions. Other studies have found that classification of fMRI responses is impaired when poor exemplars of a scene are used (Torralbo et al., 2013). This suggests that the image properties may also be important. This conclusion is supported by other MVPA studies that have shown that variation in the pattern of response in scene-selective regions is not reflected by categorical differences in scenes, but rather by the spatial layout of the scene (Kravitz et al., 2011; Park et al., 2011). However, these studies do not provide a statistical account of how the spatial layout of the scene is linked to the patterns of response.

To directly address this issue, we determined the low-level properties of the images used in our experiment using the GIST descriptor (Oliva and Torralba, 2001). This determines the orientation energy at different spatial frequencies and spatial positions in the image and generates a list of values for each image that could be used to determine the similarity of images within and across different categories of scenes. The results showed that the properties of individual images of a scene were more similar to the average of images from the same category than they were to average image of different categories. However, like the neural patterns of response, there were also differences in the consistency or homogeneity of the image properties within different categories of scenes.

The main finding from this study was that the similarity of patterns of response to different categories of scenes showed a strong positive correlation with the similarity of their low-level image statistics. This relationship between the neural response and image properties was found in both experiments with two different methods of pattern analysis (IP, LOPO). The correlation is based not only on the variation within each category of scene, but also reflects systematic variation in the between-category confusions. Our findings contrast with those of Walther et al. (2009) who found no significant correlation between neural responses and image similarity. However, their analysis involved a different measure of image similarity based on correlating pixel values across images. We likewise found, consistent with these previous results, that pixel correlations did not reliably predict the similarity of neural responses. The difference in results may reflect the fact that the GIST descriptor used in our main analysis more accurately reflects statistics encoded by the human visual system and was expressly devised to capture the critical spatial variables used to distinguish scene categories (Oliva and Torralba, 2001).

Whether we consider the ventral stream as a whole or whether we restrict our analysis to the standard scene-selective regions, the current findings suggest that the pattern of response to different categories of scenes is linked to the low-level properties of the image. This conclusion is consistent with other work showing that low-level image biases may be encoded in scene-selective regions. For example, spatial frequency (Rajimehr et al., 2011) and orientation (Nasr and Tootell, 2012) biases, along with visual field representations (Arcaro et al., 2009) have been reported in these regions.

Our results show that the neural patterns were not specific to individual participants; rather they reflect a more consistent functional organization. Using a modified cross-validation analysis (Haxby et al., 2001) we compared the pattern of response in one participant with the pattern from a group analysis in which that participant was left out. This leave-one-participant-out (LOPO) approach indicates that patterns of response to different visual scene categories are consistent across individuals (see also Haxby et al., 2011; Poldrack et al., 2009; Shinkareva et al., 2008). We found that the LOPO method often outperformed equivalent individual participant (IP) analyses. These observations are significant in that they suggest that our findings reflect the operation of consistent, large-scale organizing principles, rather than an arbitrarily distributed representation in each individual.

In conclusion, our results showed that the pattern of response in scene-selective regions of the brain can be used to discriminate different categories of scene. However, there was systematic variation in the within- and between-category similarity of neural responses across different scenes. We found that low-level image properties could explain these variations in response to visual scenes in scene-selective regions of the human brain.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.05.045.

Acknowledgments

References

Aguirre, G.K., D'Esposito, M., 1997. Environmental knowledge is subserved by separable dorsal/ventral neural areas. J. Neurosci. 17, 2512–2518.

Aguirre, G.K., D'Esposito, M., 1999. Topographical disorientation: a synthesis and taxonomy. Brain 122, 1613–1628.

Arcaro, M.J., McMains, S.A., Singer, B.D., Kastner, S., 2009. Retinotopic organization of human ventral visual cortex. J. Neurosci. 29, 10638–10652.

Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33, 1331–1336.

Epstein, R.A., 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn. Sci. 12, 388–396.

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392, 598–601.

Greene, M.R., Oliva, A., 2009. The briefest of glances: the time course of natural scene understanding. Psychol. Sci. 20, 464–472.

Grill-Spector, K., 2003. The neural basis of object perception. Curr. Opin. Neurobiol. 13, 159–166.

Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7, 37–53.

Haxby, J.V., Gobbini, M., Furey, M., Ishai, A., Schouten, J., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72, 404–416.

Joubert, O.R., Rousselet, G.A., Fize, D., Fabre-Thorpe, M., 2007. Processing scene context: fast categorization and object interference. Vis. Res. 47, 3286–3297.

Kravitz, D.J., Peng, C.S., Baker, C.I., 2011. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. J. Neurosci. 31, 7322–7333.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. 103, 3863–3868.

Maguire, E., 2001. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scand. J. Psychol. 42, 225–238.

Mendez, M.F., Cherrier, M.M., 2003. Agnosia for scenes in topographagnosia. Neuropsychologia 41, 1387–1395.

Nasr, S., Tootell, R.B.H., 2012. A cardinal orientation bias in scene-selective visual cortex. J. Neurosci. 32, 14921–14926.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. 42, 145–175.

Park, S., Brady, T.F., Greene, M.R., Oliva, A., 2011. Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. J. Neurosci. 31, 1333–1340.

Poldrack, R.A., Halchenko, Y.O., Hanson, S.J., 2009. Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol. Sci. 20, 1364–1372.

Potter, M.C., 1975. Meaning in visual search. Science 187, 965–966.

Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., Tootell, R.B.H., 2011. The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol. 9, e1000608.

Shinkareva, S.V., Mason, R.A., Malave, V.L., Wang, W., Mitchell, T.M., Just, M.A., 2008. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS One 3, e1394.

Torralba, A., Oliva, A., 2003. Statistics of natural image categories. Netw. Comput. Neural Syst. 14, 391–412.

Torralbo, A., Walther, D.B., Chai, B., Caddigan, E., Fei-Fei, L., Beck, D.M., 2013. Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. PLoS ONE 8, e58594.

Walther, D.B., Caddigan, E., Fei-Fei, L., Beck, D.M., 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. J. Neurosci. 29, 10573–10581.

Walther, D.B., Chai, B., Caddigan, E., Beck, D.M., Fei-Fei, L., 2011. Simple line drawings suffice for functional MRI decoding of natural scene categories. Proc. Natl. Acad. Sci. 108, 9661–9666.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F., Tanaka, J.W., 2010. Controlling low-level image properties: the SHINE toolbox. Behav. Res. Methods 42, 671–684.