Article:
Watson, David Mark, Young, Andy and Andrews, Tim (2016) Spatial properties of objects predict patterns of neural response in the ventral visual pathway. Neuroimage. pp. 173-183. ISSN: 1053-8119

https://doi.org/10.1016/j.neuroimage.2015.11.043

SPATIAL PROPERTIES OF OBJECTS PREDICT PATTERNS OF NEURAL RESPONSE IN THE VENTRAL VISUAL PATHWAY

David M. Watson, Andrew W. Young, Timothy J. Andrews*

Department of Psychology and York Neuroimaging Centre,

University of York, York, YO10 5DD, United Kingdom

*Corresponding author: timothy.andrews@york.ac.uk

Tables: 2

Figures: 8

Supplementary Figures: 3

Abstract: 200

Introduction: 709

Discussion: 1653

Acknowledgements

We would like to thank Danai Beintari for her assistance with data collection and analysis, and Grace Rice for her work developing the stimulus set.

ABSTRACT

Neuroimaging studies have revealed topographically organised patterns of response to different objects in the ventral visual pathway. These patterns are thought to be based on the form of the object. However, it is not clear what dimensions of object form are important. Here, we determined the extent to which spatial properties (energy across the image) could explain patterns of response in these regions. We compared patterns of fMRI response to images from different object categories presented at different retinal sizes. Although distinct neural patterns were evident to different object categories, changing the size (and thus the spatial properties) of the images had a significant effect on these patterns. Next, we used a computational approach to determine whether more fine-grained differences in the spatial properties can explain the patterns of neural response to different objects. We found that the spatial properties of the image were able to predict patterns of neural response, even when categorical factors were removed from the analysis. We also found that the effect of spatial properties on the patterns of response varies across the ventral visual pathway. These results show how spatial properties can be an important organizing principle in the topography of the ventral visual pathway.

INTRODUCTION

Although a large number of studies have investigated the functional properties of high-level regions in the ventral visual pathway, an overarching framework for how information is represented topographically in this region is not fully resolved (Op de Beeck et al. 2008; Grill-Spector and Weiner 2014). Neuroimaging studies using univariate analyses have shown that regions of the ventral visual pathway are selective for different categories of objects (Kanwisher 2010). The location of these regions is broadly consistent across individuals, suggesting that common organizing principles underpin the topography of this region (Malach et al. 1995; Kanwisher et al. 1997; Epstein and Kanwisher 1998; Cohen et al. 2000; Downing et al. 2001). The importance of this topography is evident in multivariate studies that have shown that the pattern of response across the entire ventral stream can distinguish a greater range of object categories than shown in previous univariate studies (Haxby et al. 2001; Spiridon and Kanwisher 2002; Kriegeskorte et al. 2008b). The potential importance of the pattern of response is demonstrated by the fact that the ability to discriminate particular object categories is still evident when the most category-selective regions are removed from the analysis (Haxby et al. 2001).

The distributed nature of the fMRI response to different categories of objects within the ventral visual pathway has been interpreted as showing a topographic map of object forms; a hypothesis known as 'object form topography' (Haxby et al. 2001). However, it is not clear what dimensions are important for this object form topography. A variety of evidence has suggested that patterns of response in the ventral visual pathway are linked to the categorical or semantic information that the images convey (Kriegeskorte et al. 2008b; Naselaris et al. 2009; Connolly et al. 2012). Evidence for other organizing principles can be

found in the large-scale patterns of response to animacy (Chao et al. 1999; Kriegeskorte et al. 2008b; Clarke and Tyler 2014) or to the real-world size of objects (Konkle and Oliva 2012; Konkle and Caramazza 2013). However, it remains unclear how these higher level categorical or semantic properties in the ventral visual pathway might arise from the image-based representations found in early visual regions (Op de Beeck et al. 2008).

In recent studies, we have asked to what extent image-based representations might underpin category-selective patterns of response in the ventral visual pathway (Rice et al. 2014; Watson et al. 2014). We found that similarities in the patterns of fMRI response between different categories of objects could be predicted by corresponding similarities in their low-level image properties. However, more needs to be known about which image properties are important. Previous studies have suggested that ventral temporal cortex may be sensitive to the spatial properties of the visual scene (Levy et al. 2001; Malach et al. 2002; Golomb and Kanwisher 2012; Cichy et al. 2013; Troiani et al. 2014), but it has not been clear whether this reflects a modification of the underlying categorical organization based on the way natural object categories are viewed or whether spatial properties themselves represent a fundamental organizing principle in the ventral visual pathway (Kanwisher 2001).

The aim of this study was to systematically investigate the importance of spatial properties to the pattern of response across the ventral visual pathway. To address this question, we determined the effect of changes in the actual size of the image on the patterns of response to different types of object using Multi-Voxel Pattern Analysis (MVPA). The image size manipulation is known to have a negligible effect on the perception of objects but will clearly have a profound effect on the spatial energy across the image. If the

pattern of response is invariant to changes in image size then we would expect similar patterns of response to the same object category presented at different image sizes. If, on the other hand, patterns of response are sensitive to changes in image size, we would expect significant differences in the pattern of response to the same object presented at different image sizes. Next, we used a computational approach to more directly measure energy across the image. Our aim was determine whether more subtle differences in spatial properties beyond those carried by the large changes conveyed by size could explain patterns of response in the ventral visual pathway.

METHODS

Participants

20 participants (8 males; median age 26; age range 19-36) took part in the experiment. All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

Stimuli

Visual stimuli were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant with all images subtending approximately 10.7° of visual angle. Objects depicted within the images subtended an average width of 6.94° (SD = 2.40°) and height of 7.38° (SD = 2.12°) at the large size, and an average width of 3.50° (SD = 1.20°) and height of 3.72° (SD = 1.06°) at the small size. On average large and small objects occupied 27.23% (SD = 9.92%) and 6.99% (SD = 2.47%) of pixels within the image respectively. Images of bottles, chairs, faces, and houses were presented on a mid-grey background, with 48 images per category. Face images were taken from the Radboud Face Database (Langner et al. 2010). Images from the other categories came from a variety of internet sources. Small images were created by down sampling the original images to 50% of their width and height and then placing them on a mid-grey background of the same dimensions as the original images. Figure 1 shows examples of the images used in the study.

[Figure 1 near here]

Experimental Design

The experiment comprised a single MRI scan run. During the scan, participants viewed images from 8 stimulus conditions comprising 4 categories (bottle, chair, face, house) across 2 sizes (large and small). Images from each condition were presented in a blocked fMRI design with 9 images per block (8 unique and 1 repeated). Each image was presented for 750ms followed by a 250ms mid-grey screen. Each stimulus block was separated by a 9s period in which a fixation screen with a mid-grey background was presented. Each condition was repeated 6 times in a counterbalanced block design, giving a total of 48 blocks. To maintain attention throughout the scan session participants performed a one-back task in which they were required to detect the repeated presentation of one image in each block, responding with a button press.

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, radiofrequency coil tuned to 127.7MHz. Data were collected from 38 contigual axial slices in an interleaved order via a gradient-echo EPI sequence (TR = 3s, TE = 32.5ms, FOV = 288x288mm, matrix size = 128x128, voxel dimensions = 2.25x2.25mm, slice thickness = 3mm with no inter-slice gap, flip angle = 90°, pixel bandwidth = 39.06 kHz, phase-encoding direction = anterior-posterior).

fMRI Analysis

Univariate analyses of the fMRI data were performed with FEAT v5.98 (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9s of data were 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=18s). Spatial smoothing (Gaussian) was applied at 6mm FWHM. Parameter estimates were generated for each condition by regressing the hemodynamic response of each voxel against a box-car regressor convolved with a single-gamma HRF. Next, individual participant data were entered into higher-level group analyses using a mixed-effects design (FLAME, FSL). Functional data were first registered to a high-resolution T1-anatomical image and then onto the standard MNI brain (ICBM152).

A series of anatomical regions of interest (ROIs) from the Harvard-Oxford cortical atlas included in FSL (Smith et al. 2004) were combined to form a mask that corresponded with a recent definition of ventral temporal cortex (Grill-Spector and Weiner 2014). Specifically, these masks were: inferior temporal gyrus (temporo-occipital portion), temporal-occipital fusiform cortex, occipital fusiform gyrus, and lingual gyrus. The overall ventral temporal mask was defined by a concatenation of the individual anatomical masks. The locations of these masks are shown in Figure 2. Functional data were then transformed to the MNI152 standard space and restricted to these anatomical ROIs for all further statistical analyses.

[Figure 2 near here]

Next, we measured patterns of response to the different stimulus conditions. Parameter estimates were generated for each condition in the experiment scans. The reliability of response patterns across participants was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Shinkareva et al. 2008; Poldrack et al. 2009; Rice et al. 2014; Watson et al. 2014) in which parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each

condition in each voxel. This LOPO process was repeated such that every participant was left out of a group analysis once. The LOPO procedure thus yields 2 patterns (one group, one individual) per condition per subject. These data were then submitted to correlation-based pattern analyses (developed from Haxby et al. (2001, 2014)) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by subtracting the mean response per voxel across all experimental conditions (Figure 4a). For each iteration of the LOPO cross-validation the normalized patterns of response to each stimulus condition were correlated between the group and the left-out participant. Importantly, this allows us to not only test the relative similarities within and between the conditions, but also whether these patterns are consistent across individuals. A Fisher's Z-transformation was then applied to the correlations prior to further statistical analyses. To test whether there may also be patterns of response that are idiosyncratic within each participant we also repeated our analyses at an individual level using an odd/even partition of the stimulus blocks. These analyses produced a pattern of results which was highly consistent with those of the LOPO analyses (Supplementary Figure 1). We provided a further test of this hypothesis using intra-class correlations; results were again highly consistent with the LOPO analyses (Supplementary Figure 2).

A representational similarity analysis (RSA; Kriegeskorte et al. 2008a) utilising multiple regression was used to assess the contributions of size and category to the neural response patterns. For each factor (size and form), a binary regressor was generated representing a model correlation matrix in which ones were placed on those elements where the relevant factor was shared and zeroes on all other elements. The regressors therefore represent the extreme cases where the patterns of response are entirely predicted by either size or by

category; these regressors are illustrated in Figure 5a-b. For each model, elements within the matrix were extracted and flattened to a vector. These vectors were then repeated and tiled to match the number of iterations of the LOPO procedure, and entered into the model as regressors. For each iteration of the LOPO procedure elements of the corresponding MVPA correlations matrix were extracted and flattened to a vector. These vectors were then concatenated across LOPO iterations and entered into the model as the outcome variable. This analysis yielded a beta value for each regressor which would be expected to differ significantly from zero if that regressor was able to explain a significant amount of the variance in the MVPA correlations. A t-contrast was used to assess the significance of the differences between the betas. All regressors and outcomes were Z-scored such that all beta values are given in standardised units. For analysis of the ventro-temporal sub-regions, a Bonferroni correction for multiple comparisons was applied across the number of regions.

Image Properties Analysis

Visual properties of the images were assessed using modified versions of the GIST descriptor (Oliva and Torralba 2001). To generate a GIST descriptor, each image was passed through a series of Gabor filters spanning 8 orientations and 4 spatial frequencies, generating 32 filtered images for each input image. Next, each filtered image was divided into an 8×8 grid and pixel intensities averaged within each grid cell. We then constructed two variants of the GIST descriptor that measured the spatial and non-spatial properties of each image separately (Fig. 3). First, we constructed a Spatial GIST descriptor that remained sensitive to the spatial but not the spectral properties of the image. For each grid cell independently we averaged across the filtered images, yielding a single 8×8 grid which was then flattened to a vector of 64 numbers. This vector therefore represents the spatial energy across the image;

higher values indicate a greater amount of spatial energy at a specific spatial location within the image. Second, we constructed a Spectral GIST descriptor that remained sensitive to the spectral properties (spatial frequency, orientation) but not the spatial properties of the image. For each filtered image we further averaged across the grid cells, such that each filtered image was reduced to a single value, and these values then concatenated into a vector of 32 numbers representing the image's spectral or non-spatial properties; higher values indicate a greater presence of a particular or spatial frequency and orientation within the image.

[Figure 3 near here]

In each case, the spatial and spectral GIST were then correlated across within-condition (e.g. small bottle – small bottle) and between-condition (e.g. small bottle – large bottle or small bottle – large chair) comparisons using a leave-one-image-out cross-validation procedure. An average correlation matrix for each GIST descriptor type (spatial, spectral) was then produced by averaging across iterations of the cross-validation.

A RSA utilising multiple regression was used to assess the ability of each GIST descriptor to predict the neural response patterns. The spatial and spectral GIST correlation matrices were entered as regressors, while the fMRI MVPA correlation matrices, concatenated across LOPO iterations, were entered as the outcome variable. A t-contrast was used to assess the significance of the differences between the betas. All regressors and outcomes were Z-scored such that all beta values are given in standardised units. For analysis of the ventro-temporal sub-regions, a Bonferroni correction for multiple comparisons was applied across the number of regions.

Analyses of Posterior-Anterior and Medial-Lateral Axes

Finally, we tested how the patterns of response varied across the ventro-temporal region. First, the main ventro-temporal mask was split along the posterior-anterior axis into 4 slices each 7 voxels thick (Figure 8a). In addition, the main ventro-temporal mask was first split into 8 slices, each 8 voxels thick, spanning the left-to-right extent of the volume, and then combined across hemispheres to yield 4 splits covering the medial-lateral axis (Figure 8b). The pattern analysis, object-size/object-form regression, and spatial-GIST/spectral-GIST regression procedures were then repeated for each slice separately.

RESULTS

Patterns of neural response were measured for 4 different object categories (bottle, chair, face, house) presented at 2 different sizes (large, small). Behavioural performance on the one-back task was near ceiling (mean accuracy = 97.40%, SEM = 0.41%), indicating participants maintained attention throughout the scan. Figure 4a shows the normalised group responses to each condition across the ventro-temporal ROI. Responses above the mean are shown in red, and responses below the mean are shown in blue. Both size and category manipulations can be seen to modulate the patterns of response. For illustrative purposes, we also normalised the response patterns within each size independently (Figure 4b). When this is done, distinct patterns of response are seen across categories, but these patterns appear highly similar across sizes. This suggests that, although ventro-temporal regions may be very sensitive to image size, they are nevertheless capable of responding with a consistent underlying pattern when the responses are normalized for a specific size.

[Figure 4 near here]

To determine the contributions of size and category to the patterns of neural response within the ventro-temporal region we used a RSA analysis (Kriegeskorte et al. 2008a). First, correlation-based MVPA (Haxby et al. 2001) was performed on the normalized responses. The resulting correlation matrix is shown in Figure 5c. Next, model correlation matrices were generated representing the outcome if the patterns of response were entirely predicted by either the size (Figure 5a) or category (Figure 5b). These were then used as regressors in a multiple regression analysis of the fMRI data. Figure 5d shows the resulting coefficient values for each regressor. Both the size ($\beta = .55$, $t(717) = 25.80$, $p < .001$) and category ($\beta = .54$, $t(717) = 25.52$, $p < .001$) regressors were able to significantly

predict the fMRI responses. A t-contrast revealed no significant difference between the coefficients (t = 0.18, p = .856). This shows both that size of the image and the category make significant contributions to the patterns of response. Given that the effect of category was already well-established (Haxby et al., 2001), we consider the effect of size to be the novel finding here.

[Figure 5 near here]

Next, we used a computational approach to more directly explore the relative importance of spatial and non-spatial image properties on the patterns of response in the ventral visual pathway. In previous studies (Rice et al. 2014; Watson et al. 2014) we have shown that image properties measured by the GIST descriptor (Oliva and Torralba 2001) can be used to predict fMRI patterns of response. In this study, we created two versions of the GIST descriptor (see Fig. 3): one that was sensitive to spatial properties of the image (Spatial GIST), and one that was sensitive to non-spatial properties, such as orientation and spatial frequency content (Spectral GIST). These GIST descriptors were used to calculate the spatial and non-spatial profile of images within each condition. We then compared these profiles within and between different stimulus conditions. Fig. 6a, 6b show the average correlation matrices for spatial GIST and spectral GIST across all stimulus conditions. These correlation matrices were then used as regressors in a multiple regression analysis of the fMRI data (see Fig. 5c). Fig. 6c shows that both the spatial GIST (β = .54, t(717) = 16.93, p < .001) and spectral GIST (β = .24, t(717) = 7.66, p < .001) explained a significant amount of variance in the neural responses, with significantly more variance explained by the spatial GIST (t = 5.31, p < .001). We also tested for differences between the hemispheres; results were

consistent both across hemispheres and with the analysis of the bilateral region (Supplementary Figure 3).

[Figure 6 near here]

These results suggest that image properties in addition to categorical information can predict patterns of response in the ventral visual pathway. However, images drawn from the same condition are likely to have similar low-level properties. So, the link between image properties and patterns of neural response is expected under both categorical and image-based accounts. To address this issue, we repeated the analysis only using the off-diagonal elements of the correlation matrices. As before, both the spatial ($\beta = .41$, $t(557) = 10.90$, $p < .001$) and spectral ($\beta = .15$, $t(557) = 3.84$, $p < .001$) GIST significantly predicted the neural responses within the ventral visual region, but with significantly more variance explained by the spatial GIST ($t = 4.79$, $p < .001$).

Although this suggests that the patterns of response are not driven by the difference between within-condition and between-condition comparisons, the off-diagonal elements themselves still contain correlation values representing within--category comparisons (e.g. small bottle vs large bottle), which again might be expected to be higher than between -category values irrespective of the underlying representational structure. To this end, we repeated our analyses excluding not only the on-diagonal elements, but also those off-diagonal elements representing within-category comparisons. Within the ventral stream region, the spatial GIST again significantly predicted responses ($\beta = .52$, $t(477) = 12.95$, $p < .001$). In contrast, the spectral GIST coefficient was significantly negative ($\beta = -.18$, $t(477) = -4.38$, $p < .001$). The spatial GIST coefficient was significantly greater than the spectral GIST

coefficient (t = 11.12, $p < .001$). Thus, the predictive capability of the spatial GIST remains even in the absence of categorical information.

The ROI used for the ventral temporal mask was composed of 4 anatomical regions from the Harvard-Oxford Atlas. To determine if the representation was different across anatomical regions, we repeated our analyses for each anatomical region. MVPA correlation matrices for each anatomical region are shown in Fig. 7a. The size vs category RSA analyses are shown for each region in Fig. 7b. We found a significant effect of size in the lingual (β = .72, t(717) = 31.43, $p < .001$) and occipital-fusiform (β = .49, t(717) = 22.48, $p < .001$) regions, but not in the temporo-occipito-fusiform (β = .05, t(717) = 1.72, $p = .344$) and inferior temporal (β = .05, t(717) = 1.78, $p = .304$) regions. A significant effect of category was found in the lingual (β = .26, t(717) = 11.29, $p < .001$), occipital-fusiform (β = .57, t(717) = 25.98, $p < .001$), temporo-occipito-fusiform (β = .67, t(717) = 24.39, $p < .001$) and inferior temporal (β = .57, t(717) = 18.44, $p < .001$) regions. More variance was explained by size compared to category in the lingual (t = 13.23, $p < .001$) region, but there was no significant difference between regressors in the occipital-fusiform region (t = 2.30, $p = .086$), and more variance was explained by category compared to size in the temporo-occipital-fusiform (t = 14.89, $p < .001$) and inferior temporal (t = 10.95, $p < .001$) regions.

Next, we repeated the spatial GIST/spectral GIST analysis for each region (Fig. 7c). We found a significant effect of spatial GIST in the lingual (β = .61, t(717) = 17.38, $p < .001$), occipital-fusiform (β = .45, t(717) = 13.84, $p < .001$), temporo-occipito-fusiform (β = .20, t(717) = 5.66, $p < .001$) and inferior temporal (β = .15, t(717) = 4.12, $p < .001$) regions. We found a significant effect of spectral GIST in the occipital-fusiform (β = .31, t(717) = 9.55, $p < .001$), temporo-occipito-fusiform (β = .44, t(717) = 12.26, $p < .001$) and inferior temporal (β = .4.23, t(717) = 11.30, $p < .001$) regions, but not the lingual (β = -.04, t(717) = -1.21, $p = .900$)

region. More variance was explained by the spatial compared to the spectral GIST in the lingual region (t = 10.65, $p < .001$) region, whilst more variance was explained by the spectral compared to the spatial GIST in the inferior temporal (t = 4.11, $p < .001$) and temporo-occipito-fusiform regions (t = 3.79, $p = .001$). There was no significant difference between regressors in the occipital-fusiform region (t = 2.46, $p = .057$).

[Figure 7 near here]

Next, we tested how the patterns of response might change along the posterior-anterior axis of the ventro-temporal region. The main ventro-temporal mask was split into 4 slices spanning the posterior (slice 1) to anterior (slice 4) extent of the mask. The MVPA and RSA analyses were then repeated for each slice independently. The results of these analyses are illustrated in Figure 8a. We first contrasted the category and size models using multiple regression analyses. The size model significantly predicted responses in the most posterior regions, however the amount of variance explained declined towards more anterior regions and failed to reach significance in the most anterior slice (slice1: $\beta = .50$, $t(717) = 20.36$, $p < .001$; slice2: $\beta = .33$, $t(717) = 11.78$, $p < .001$; slice3: $\beta = .09$, $t(717) = 3.23$, $p = .004$; slice4: $\beta = .05$, $t(717) = 1.86$, $p = .190$). The category model significantly predicted neural responses throughout the ventro-temporal region (slice1: $\beta = .49$, $t(717) = 19.85$, $p < .001$; slice2: $\beta = .54$, $t(717) = 19.43$, $p < .001$; slice3: $\beta = .64$, $t(717) = 22.68$, $p < .001$; slice4: $\beta = .65$, $t(717) = 23.19$, $p < .001$), and indeed a slight increase in the amount of variance explained was observed moving anteriorly. This suggests that neural representations of visual objects show an increasing degree of size invariance along the posterior to anterior regions of the ventral visual pathway. We then repeated this analysis with the spatial and spectral GISTs. Both spatial and spectral GISTs explained a significant amount of variance

throughout the ventro-temporal region. The spatial GIST explained progressively less variance moving anteriorly (slice1: β = .46, t(717) = 13.43, $p < .001$; slice2: β = .40, t(717) = 11.41, $p < .001$; slice3: β = .21, t(717) = 5.95, $p < .001$; slice4: β = .23, t(717) = 6.52, $p < .001$). Conversely, the spectral GIST explained progressively more variance moving anteriorly (slice1: β = .24, t(717) = 7.08, $p < .001$; slice2: β = .28, t(717) = 8.05, $p < .001$; slice3: β = .43, t(717) = 11.93, $p < .001$; slice4: β = .43, t(717) = 12.21, $p < .001$). Finally, removing the within-category comparisons from the analysis did not affect the ability of the spatial GIST to predict patterns of response, but did affect the ability of the spectral GIST to predict patterns of neural response. Results of these analyses are shown in Table 1.

[Figure 8 & Table 1 near here]

We next tested how the patterns of response change along the medial-lateral axis of the ventro-temporal region. The results of these analyses are illustrated in Figure 8b. The size model significantly predicted the most variance in the most medial region, but gradually declined towards more lateral regions (slice1: β = .64, t(717) = 26.13, $p < .001$; slice2: β = .46, t(717) = 20.98, $p < .001$; slice3: β = .17, t(717) = 6.13, $p = .004$; slice4: β = .07, t(717) = 2.03, $p = .129$). In contrast, the category model significantly predicted neural responses throughout the ventro-temporal region (slice1: β = .31, t(717) = 12.43, $p < .001$; slice2: β = .61, t(717) = 28.12, $p < .001$; slice3: β = .64, t(717) = 23.27, $p < .001$; slice4: β = .49, t(717) = 14.98, $p < .001$). We next contrasted the spatial and spectral GISTs along the medial to lateral axis of the ventral stream. The spatial GIST explained progressively less variance moving from medial to lateral (slice1: β = .61, t(717) = 17.35, $p < .001$; slice2: β = .43, t(717) = 13.12, $p < .001$; slice3: β = .27, t(717) = 7.35, $p < .001$; slice4: β = .13, t(717) = 3.26, $p =$

.005). Conversely, the spectral GIST explained progressively more variance moving from medial to lateral (slice1: $\beta < .01$, $t(717) = 0.03$, $p = 1$; slice2: $\beta = .34$, $t(717) = 10.51$, $p < .001$; slice3: $\beta = .38$, $t(717) = 10.56$, $p < .001$; slice4: $\beta = .38$, $t(717) = 9.83$, $p < .001$). This suggests neural representations of visual objects show an increasing degree of size and spatial invariance moving from medial to lateral through the ventral temporal region. Removing the on-diagonal elements from the analysis had a greater effect on the spectral GIST compared to the spatial GIST (Table 2).

[Table 2 near here]

DISCUSSION

The aim of this study was to investigate how spatial properties of images influence patterns of neural response in the ventral visual pathway. We found that patterns of neural response to images of different object categories are influenced by the actual size of the images. We then used a computational approach to more directly test whether more subtle variance in the spatial properties of images could explain the different patterns of response to objects in the ventral temporal cortex. Our results suggest that the spatial properties of images are an important organizing principle in the topography of the ventral visual pathway.

The distributed nature of the fMRI response to different categories of objects within the ventral visual pathway has been interpreted as showing a topographic map of object form (Haxby et al. 2001) that is analogous with the continuous, topographic maps found in early stages of visual processing (Hubel and Wiesel 1968; Bonhoeffer and Grinvald 1991; Wandell et al. 2007). However, it is not clear what dimensions might be important for the hypothesised object form topography of higher visual areas. By varying the size and category from which images were presented, it was possible for us to compare the relative contribution of these factors to the patterns of response across the ventral visual pathway. Our results show, consistent with previous studies (Haxby et al. 2001; Hanson et al. 2004), that there are indeed distinct patterns of response to different object categories. However, we also show that these patterns of response across the ventral stream are significantly affected by the retinal size of the objects. That is, the pattern of response to one object category is significantly reduced when the images have a different size.

The marked effect of size on the patterns of response shown in the current study suggests that spatial properties are an important organizing dimension in the topography of some regions of the ventral visual pathway. To more directly explore the importance of spatial properties as an important organizing principle in the topography of the ventral visual pathway, we used a computational approach in which we directly measured spatial and non-spatial properties of the image (see Fig. 3). We produced modified versions of the GIST descriptor (Oliva and Torralba 2001): one version of the GIST descriptor was sensitive to spatial properties of the stimulus (spatial GIST), while the other version was sensitive to non-spatial properties, such as orientation and spatial frequency content (spectral GIST). The use of the spatial GIST allowed us to take into account more subtle differences in spatial properties beyond those carried by the large changes conveyed by size. For example, there are significant differences in the spatial properties of objects from different categories even when they are scaled to a similar size (Fig. 1). We found that both the spatial GIST and the spectral GIST were able to predict neural responses throughout the ventral region. However, we found that the spatial GIST explained more variance.

Our results are consistent with previous studies in which we showed that patterns of response in high-level visual areas may be better explained by the response to image properties that are characteristic of different object categories (Rice et al. 2014; Watson et al. 2014). Nevertheless, it was not evident from these data which image dimensions were important. The importance of spatial properties is consistent with previous studies demonstrating the importance of spatial visual properties in ventral visual regions, such as biases for eccentricity (Levy et al. 2001) and retinotopic location (Golomb and Kanwisher 2012; Cichy et al. 2013). However, it has not been clear until now whether spatial

properties represent a fundamental organizing principle in the organization of the ventral visual stream or whether they reflect a modification of the underlying category-selective representation due to the statistics of natural viewing (Kanwisher 2001).

In principle, the link between image properties and the patterns of neural response we have shown in the ventral visual pathway need not be entirely counter to categorical representations reported in previous studies (Kriegeskorte et al. 2008b; Naselaris et al. 2009). This is because object categories typically contain objects that are visually similar. So, the effect of category that was evident throughout the ventral stream could reflect sensitivity to visual properties besides object size, that were not directly manipulated here, but nevertheless differ reliably between object categories (Rice et al. 2014). For example, image properties were still able to predict the patterns of neural response when all the within-category correlations were removed from the analysis. If the organization of the ventral temporal cortex were solely dependent on categorical principles, then the relationship between neural responses and the spatial properties of the images should not extend to between-category correlations. Interestingly, this effect was greater for the spatial than the spectral properties. Therefore, the apparently high-level selectivity for object categories in ventral visual cortex could, at least in part, arise from biases for the lower-level, visual components of the image (Op de Beeck et al. 2008).

The effect of size on patterns of response in the ventral stream showed a progressive decline from posterior to anterior regions of the ventral stream. Indeed, there was no significant effect of object size in the most anterior regions of interest, suggesting a greater degree of size invariance in more anterior regions of the ventral stream. These findings are consistent with other univariate fMRI analyses showing a degree of size invariance in

regions of the ventral stream. For example, a number of studies have shown fMR-adaptation to objects even when the images vary in size (Grill-Spector et al. 1999; Andrews and Ewbank 2004; Ewbank et al. 2005). The findings are also in line with non-human primate studies, which show an increase in the receptive field size of neurons from posterior to anterior regions of temporal lobe (Rust and DiCarlo 2010; Kay et al., 2015). This increase in receptive field size could explain the decreased effect of size or increased size invariance that we observe in more anterior regions.

To determine whether the influence of image properties varied across the ventral stream, we repeated our analysis along different axes of our mask. Both the spatial GIST and spectral GIST predicted patterns of response in all anterior-posterior subdivisions of the ventral stream mask. However, there was a slight decrease in the amount of variance explained by the spatial GIST and a slight increase in the amount of variance explained by the spectral GIST along the posterior to anterior axis. Gradual changes in the specificity along the posterior-anterior axis of the ventral visual pathway have been argued to reflect a transition from image-based representations in more posterior regions to more categorical and semantic representations in more anterior regions (Rust and DiCarlo 2010; Clarke and Tyler 2014; Shimotake et al. 2014). Our results suggest that these changes in selectivity across may better be interpreted as showing a transition in the way that image properties are represented.

In an additional set of analyses we split the ventral stream mask along a medial-to-lateral axis. We found the predictive capability of the spatial GIST decreased towards more lateral regions while the opposite was true of the spectral GIST. Previous studies have indicated a medial-to-lateral organisation of the ventral stream for stimulus properties such

as animacy (Chao et al. 1999; Konkle and Caramazza 2013), real-world size (Konkle and Oliva 2012; Konkle and Caramazza 2013), and visual eccentricity (Malach et al. 2002). The current findings are therefore consistent with a division of labour between medial and lateral sites, although further investigation will be required to determine if a direct link can be made between the image properties tested here and the stimulus properties identified in previous research.

Our findings do not imply that the representation of image properties in high-level visual areas is similar to the way information is represented in lower-level regions. For example, it seems likely that high-level areas will increasingly selectively represent those low-level image properties that are more commonly found in natural images (Kayaert et al. 2003; Op de Beeck et al. 2008). Moreover, the low-level image description used in our studies may not account for all of the variance in the magnitude and patterns of response. A number of fMRI studies have shown a link between patterns of fMRI response in object-selective regions and both the physical and perceived shape of objects (Haushofer et al. 2008; Op de Beeck et al. 2008; Drucker and Aguirre 2009). So, it seems that models of image properties that incorporate these mid-level properties of objects could provide a more complete account of cortical organization in the ventral visual pathway.

Another important feature of our results is they show that the patterns of fMRI response generalize across participants. Our analysis compared the pattern of response in one participant with the pattern from a group analysis in which that participant was left out (Shinkareva et al. 2008; Poldrack et al. 2009). This approach showed that the topographic patterns of response to different object categories were consistent across individuals and thus reflect the operation of consistent, large-scale topographical organizing principles.

Although this does not preclude the possibility that further information may be represented in a more idiosyncratic manner in individual brains, we none the less found similar patterns of results when we repeated our analysis at the individual level (see Supplementary Figure 1, Figures 5-7).

In conclusion, we have shown that the retinal size of objects has a significant effect on patterns of fMRI response in the ventral visual pathway. We have also shown that variance in the spatial properties of images can explain patterns of response and that the influence of spatial properties was still evident when only between-category comparisons were analysed. Finally, we found that the importance of spatial properties on the topographic patterns of response varied systematically across the ventral visual pathway. Together, these results suggest that spatial properties are an important organizing principle in the topography of the ventral visual pathway.

REFERENCES

Andrews TJ, Ewbank MP. 2004. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. Neuroimage. 23:905–913.

Bonhoeffer T, Grinvald A. 1991. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. Nature. 354:429–431.

Chao LL, Haxby J V, Martin A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat Neurosci. 2:913–919.

Cichy RM, Sterzer P, Heinzle J, Elliott LT, Ramirez F, Haynes J-D. 2013. Probing principles of large-scale object representation: Category preference and location encoding. Hum Brain Mapp. 34:1636–1651.

Clarke A, Tyler LK. 2014. Object-Specific Semantic Coding in Human Perirhinal Cortex. J Neurosci. 34:4766–4775.

Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff M, Michel F. 2000. The visual word form area. Brain. 123:291–307.

Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu Y-C, Abdi H, Haxby J V. 2012. The representation of biological classes in the human brain. J Neurosci. 32:2608–2618.

Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A Cortical Area Selective for Visual Processing of the Human Body. Science. 293:2470–2473.

Drucker DM, Aguirre GK. 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb Cortex. 19:2269–2280.

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

Ewbank MP, Schluppeck D, Andrews TJ. 2005. fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. Neuroimage. 28:268–279.

Golomb JD, Kanwisher N. 2012. Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. Cereb Cortex. 22:2794–2810.

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.

Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci. 15:536–548.

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby J V, Pollmann S. 2009. PyMVPA: a

Python Toolbox for Multivariate Pattern Analysis of fMRI Data. Neuroinformatics. 7:37–53.

Hanson SJ, Matsuka T, Haxby J V. 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? Neuroimage. 23:156–166.

Haushofer J, Livingstone MS, Kanwisher N. 2008. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biol. 6:1459–1467.

Haxby J V, Connolly AC, Guntupalli JS. 2014. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. Annu Rev Neurosci. 37:435–456.

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.

Hubel DH, Wiesel TN. 1968. Receptive Fields and Functional Architecture of Monkey Striate Cortex. J Physiol. 195:215–243.

Kanwisher N. 2001. Faces and places: of central (and peripheral) interest. Nat Neurosci. 4:455–456.

Kanwisher N. 2010. Functional specificity in the human brain: A window into the functional architecture of the mind. Proc Natl Acad Sci. 107:11163–11170.

Kanwisher N, Woods RP, Iacoboni M, Mazziotta JC. 1997. A Locus in Human Extrastriate Cortex for Visual Shape Analysis. J Cogn Neurosci. 9:133–142.

Kay KN, Weiner KS, Grill-Spector K. 2015. Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex. Curr Biol. 25:595–600.

Kayaert G, Biederman I, Vogels R. 2003. Shape tuning in macaque inferior temporal cortex. J Neurosci. 23:3016–3027.

Konkle T, Caramazza A. 2013. Tripartite Organization of the Ventral Stream by Animacy and Object Size. J Neurosci. 33:10235–10242.

Konkle T, Oliva A. 2012. A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. Neuron. 74:1114–1124.

Kriegeskorte N, Mur M, Bandettini PA. 2008a. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci. 2:1–28.

Kriegeskorte N, Mur M, Ruff D, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini P. 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron. 60:1126–1141.

Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. 2010. Presentation and validation of the Radboud Faces Database. Cogn Emot. 24:1377–1388.

Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center – periphery organization of human object areas. Nat Neurosci. 4:533–539.

Malach R, Levy I, Hasson U. 2002. The topography of high-order human object areas. Trends Cogn Sci. 6:176–184.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci. 92:8135–8139.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian Reconstruction of Natural Images from Human Brain Activity. Neuron. 63:902–915.

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.

Op de Beeck HP, Haushofer J, Kanwisher NG. 2008. Interpreting fMRI data: maps, modules and dimensions. Nat Rev Neurosci. 9:123–135.

Poldrack RA, Halchenko YO, Hanson SJ. 2009. Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol Sci. 20:1364–1372.

Rice GE, Watson DM, Hartley T, Andrews TJ. 2014. Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. J Neurosci. 34:8837–8844.

Rust NC, DiCarlo JJ. 2010. Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT. J Neurosci. 30:12978–12995.

Shimotake A, Matsumoto R, Ueno T, Kunieda T, Saito S, Hoffman P, Kikuchi T, Fukuyama H, Miyamoto S, Takahashi R, Ikeda A, Lambon Ralph M a. 2014. Direct Exploration of the Role of the Ventral Anterior Temporal Lobe in Semantic Memory: Cortical Stimulation and Local Field Potential Evidence From Subdural Grid Electrodes. Cereb Cortex. 1–16.

Shinkareva S V, Mason RA, Malave VL, Wang W, Mitchell TM, Just MA. 2008. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS One. 3:e1394.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 23:S208–S219.

Spiridon M, Kanwisher N. 2002. How Distributed Is Visual Category Information in Human Occipito-Temporal Cortex? An fMRI Study. Neuron. 35:1157–1165.

Troiani V, Stigliani A, Smith M, Epstein R. 2014. Multiple object properties drive scene-selective regions. Cereb Cortex. 24:883–897.

Wandell BA, Dumoulin SO, Brewer AA. 2007. Visual field maps in human cortex. Neuron. 56:366–383.

Watson DM, Hartley T, Andrews TJ. 2014. Patterns of response to visual scenes are linked to the low-level properties of the image. Neuroimage. 99:402–410.

		Slice 1 (posterior)	Slice 2	Slice 3	Slice 4 (anterior)
Off-diagonal (DoF = 557)	Spectral GIST (β)	.20***	.17***	.26***	.30***
	Spatial GIST (β)	.34***	.29***	.04 (ns)	.11*
	Contrast (t)	2.41 (ns)	2.07 (ns)	3.65**	3.17**
Between-category (DoF = 477)	Spectral GIST (β)	-.02 (ns)	-.27***	-.29***	-.26***
	Spatial GIST (β)	.41***	.42***	.18***	.26***
	Contrast (t)	6.46***	10.47***	6.82***	7.41***

Table 1. Multiple regression analysis of ventro-temporal region along anterior-posterior axis, restricted to off-diagonal or between-category elements of correlation matrices. The Off-diagonal analysis removes all within-condition comparisons (e.g. small bottle – small bottle or large bottle – large bottle). The Between-category analysis additionally removes all within-category comparisons (e.g. small bottle – large bottle or small chair – large chair). Values of regression coefficients and t-contrasts between regressors and corresponding significance levels are shown (*** $p < .001$, ** $p < .01$, * $p < .05$).

		Slice 1 (medial)	Slice 2	Slice 3	Slice 4 (lateral)
Off-diagonal (DoF = 557)	Spectral GIST (β)	-.04 (ns)	.27***	.25***	.35***
	Spatial GIST (β)	.53***	.32***	.08 (ns)	< .01 (ns)
	Contrast (t)	10.69***	0.96 (ns)	2.82*	6.03***
Between-category (DoF = 477)	Spectral GIST (β)	-.06 (ns)	-.06 (ns)	-.26***	.15**
	Spatial GIST (β)	.57***	.43***	.20***	.03 (ns)
	Contrast (t)	10.28***	7.54***	6.65***	1.68 (ns)

Table 2. Multiple regression analysis of ventro-temporal region along medial-lateral axis, restricted to off-diagonal or between-category elements of correlation matrices. Values of regression coefficients and t-contrasts between regressors and corresponding significance levels are shown (*** $p < .001$, ** $p < .01$, * $p < .05$).

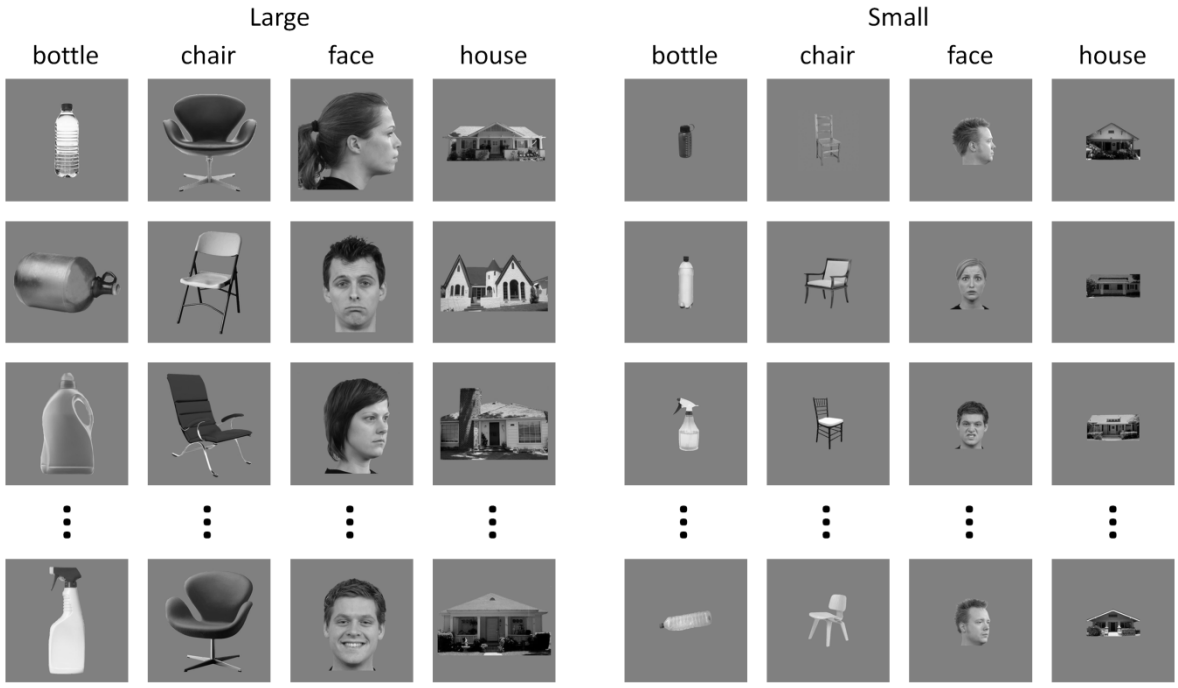


Figure 1. Examples of images from each condition in the experiment.

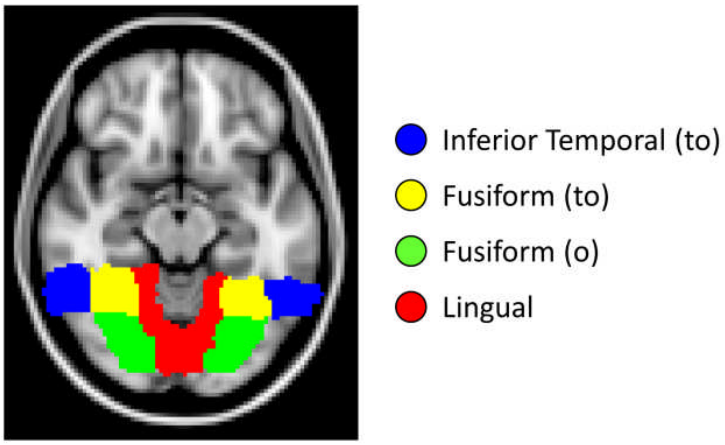


Figure 2. Harvard-Oxford anatomical masks used for pattern analyses: Inferior Temporal Gyrus (temporo-occipital portion), Temporo-Occipital-Fusiform Cortex, Occipital-Fusiform Gyrus, Lingual Gyrus. The ventro-temporal region was defined by the concatenation of these masks.

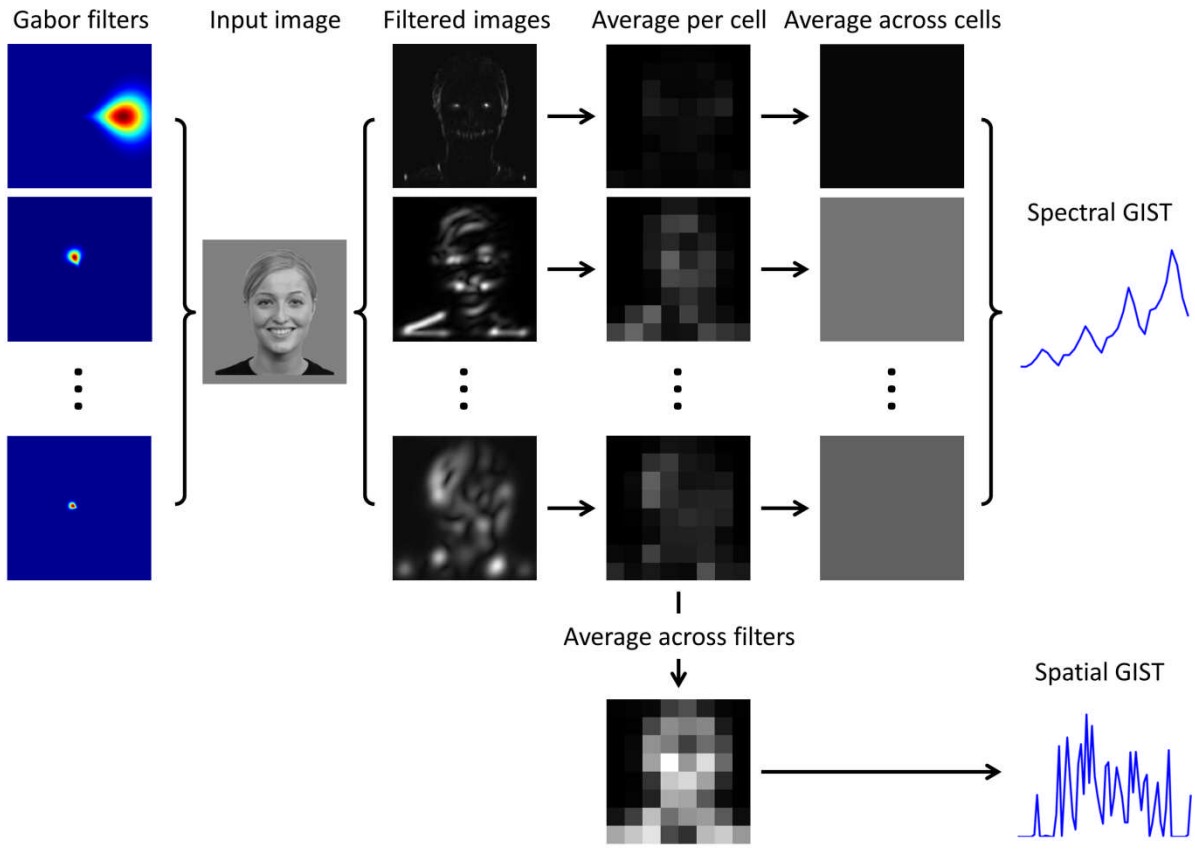


Figure 3. Schematic illustration of the calculation of the modified GIST descriptors. A series of Gabor filters (8 orientations × 4 scales) were applied to the image, yielding 32 filtered images. Each of these was then windowed by an 8×8 grid and the pixel intensities within each grid cell averaged. The Spatial GIST descriptor was constructed by averaging across filtered images whilst retaining the 8×8 grid, and then reshaping the resulting grid to yield a vector of 64 values. The Spectral GIST descriptor was constructed by further averaging across windows for each filtered image separately, and then concatenating these to yield a vector of 32 values.

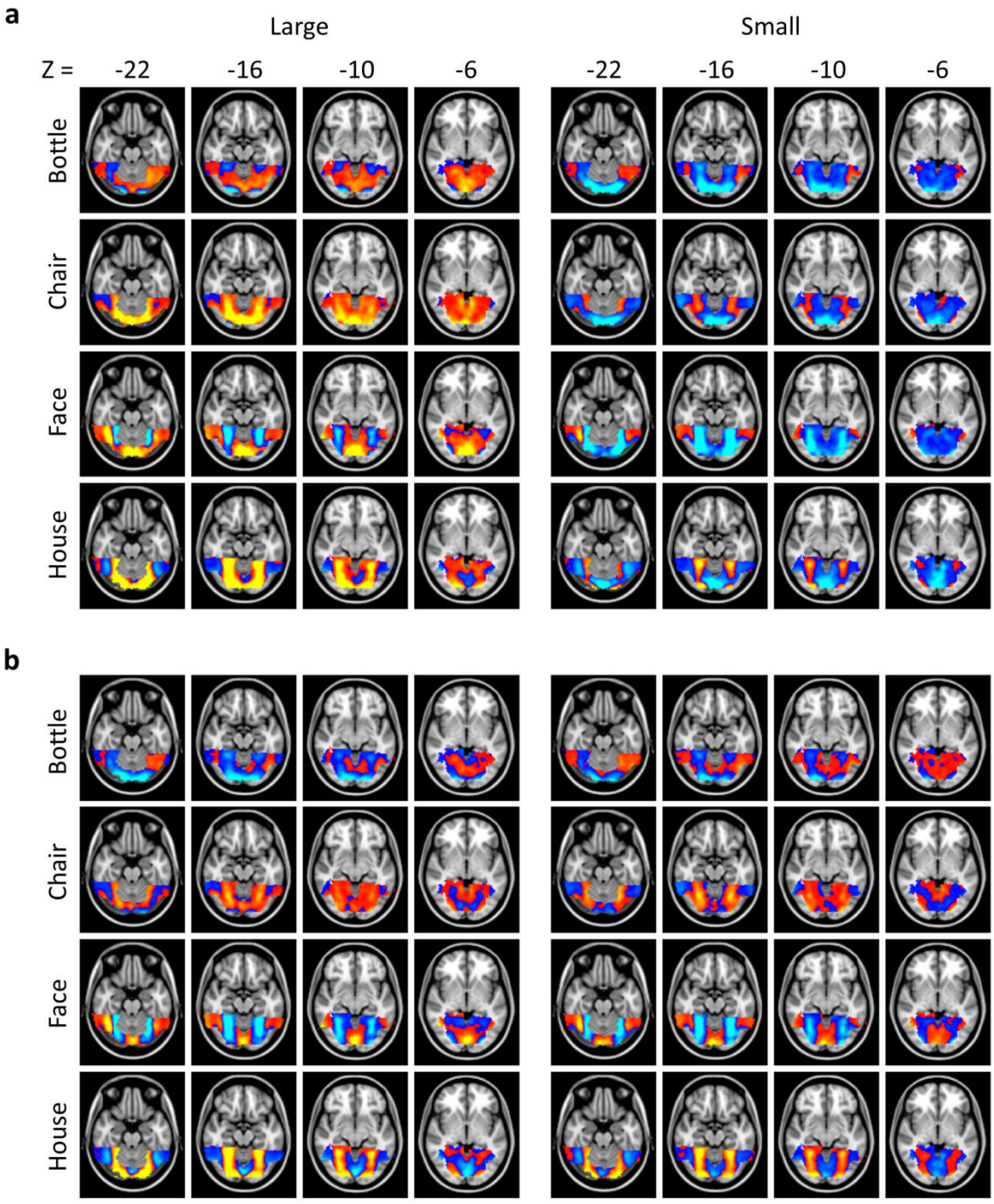


Figure 4. Group-level patterns of response to each condition, restricted to the ventral temporal ROI. Red and blue colours indicate normalized values above and below the voxel-wise mean respectively. (a) Patterns normalised across all conditions. (b) Patterns normalised within large and small conditions independently.

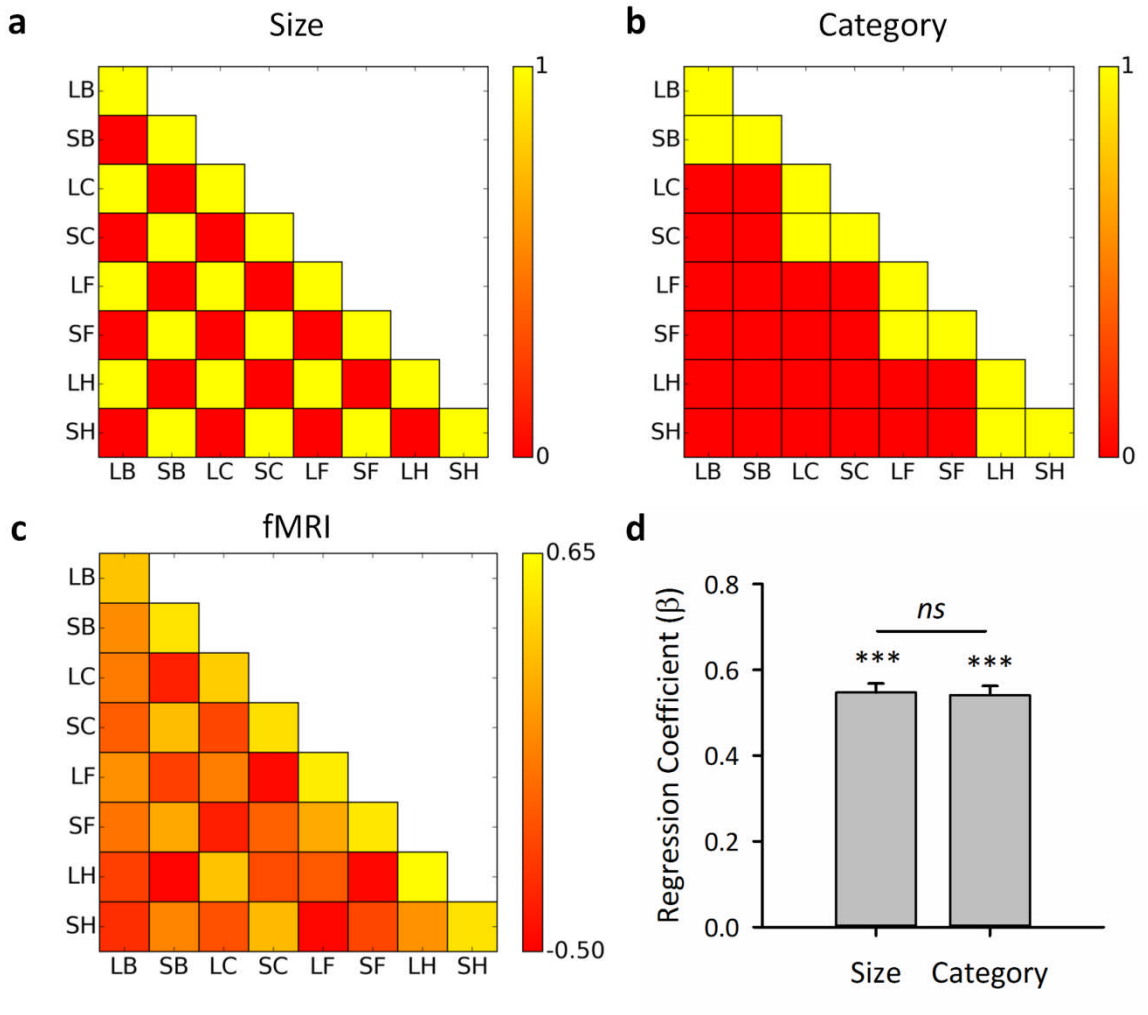


Figure 5. Ventro-temporal region: Representational similarity analysis on the effect of size and category on the patterns of response in ventral temporal cortex. Acronyms: S = small, L = large, B = bottle, C = chair, F = face, H = house. Binary model correlation matrices were defined representing the extreme cases where patterns of response are entirely predicted by either the size (a) or category (b) manipulations. These were entered as regressors into a multiple regression analysis, whilst the fMRI MVPA correlation matrix (c) was entered as the outcome variable. The resulting standardised regression coefficients are shown in (d). Error bars represent ±1 SEM.

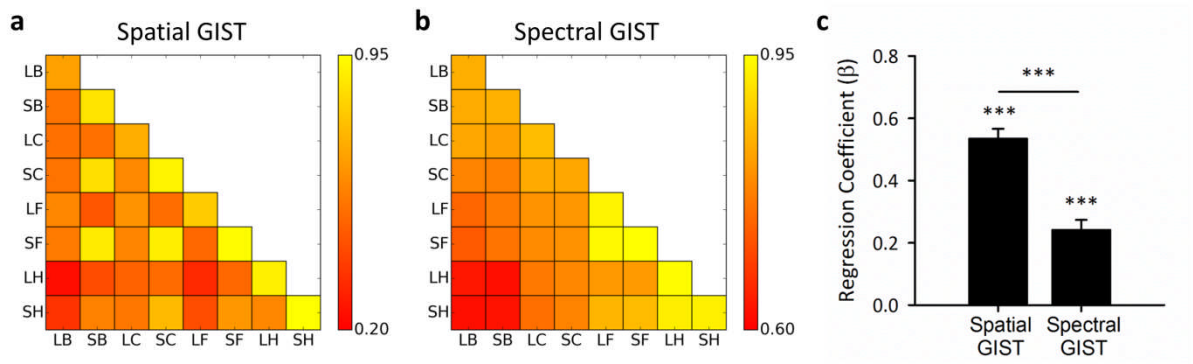


Figure 6. Ventro-temporal region: Representational similarity analysis between fMRI response and image properties defined by the Spatial GIST and Spectral GIST (see Fig. 3). Within- and between-condition correlations for each condition were determined for Spatial GIST (a) and Spectral GIST (b) descriptors. These were then entered as regressors into a multiple regression analysis, whilst the fMRI MVPA correlation matrix (Figure 5c) was entered as the outcome variable. The resulting standardised regression coefficients are shown in (c). Error bars represent ±1 SEM.

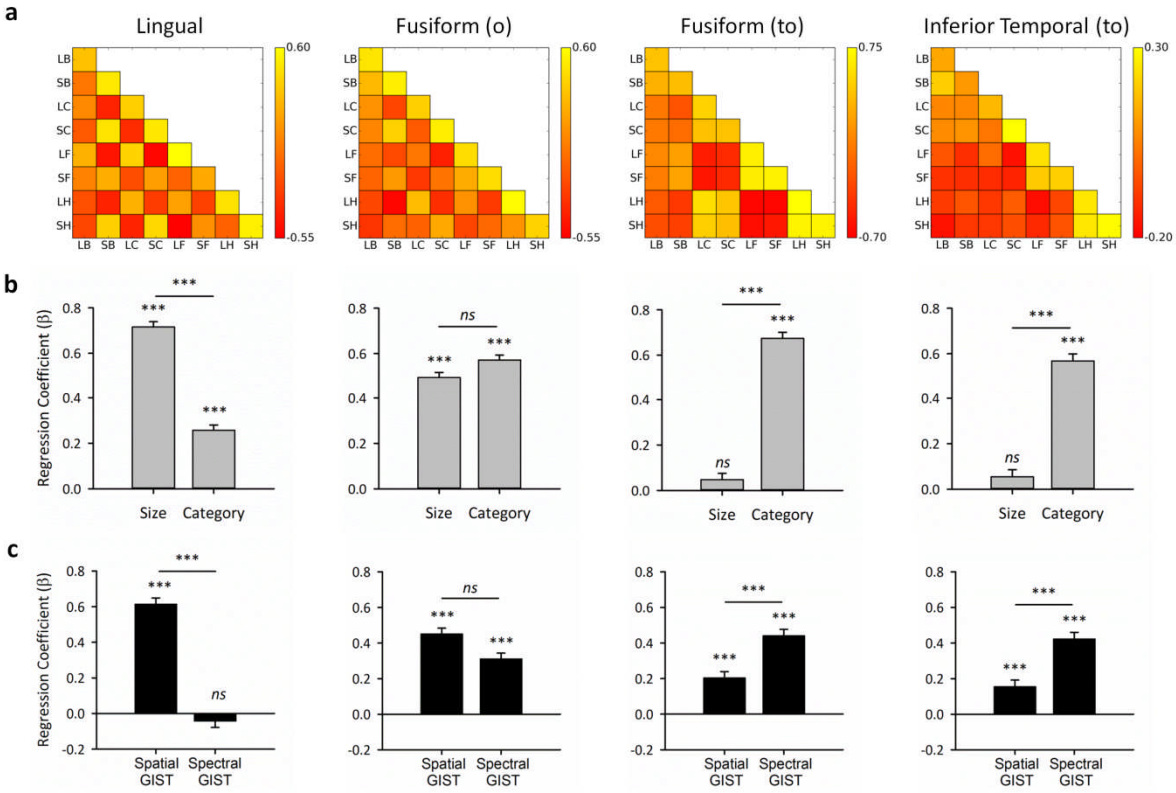


Figure 7. Representational similarity analyses for anatomical subdivisions of the ventro-temporal mask. (a) fMRI MVPA correlation matrices for each anatomical region. Standardised coefficient values of multiple regression analyses contrasting size and category models (b), and Spatial GIST and Spectral GIST descriptors (c). Error bars represent ±1 SEM.

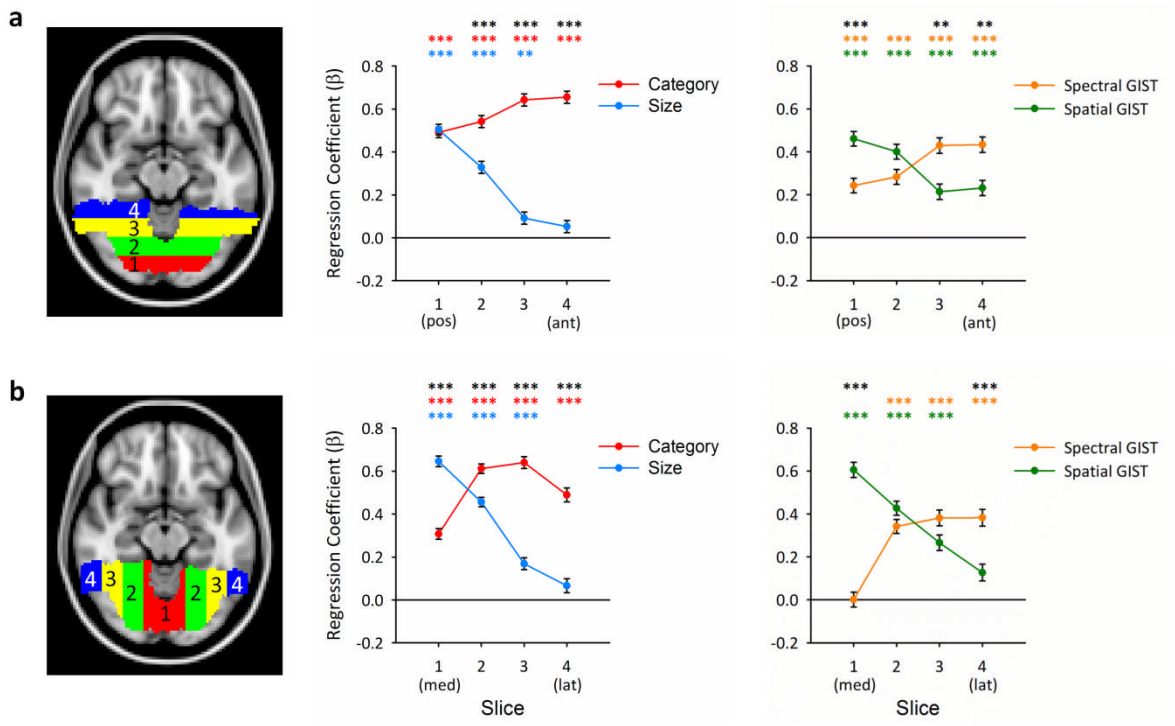
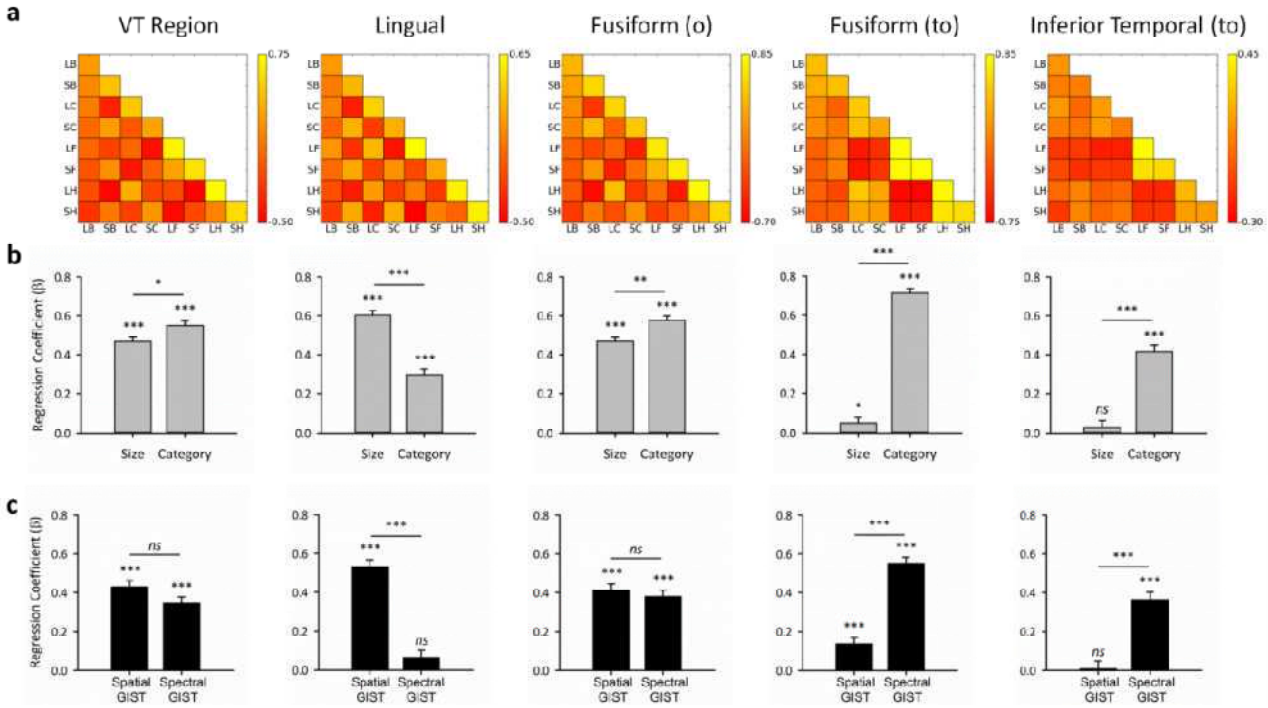
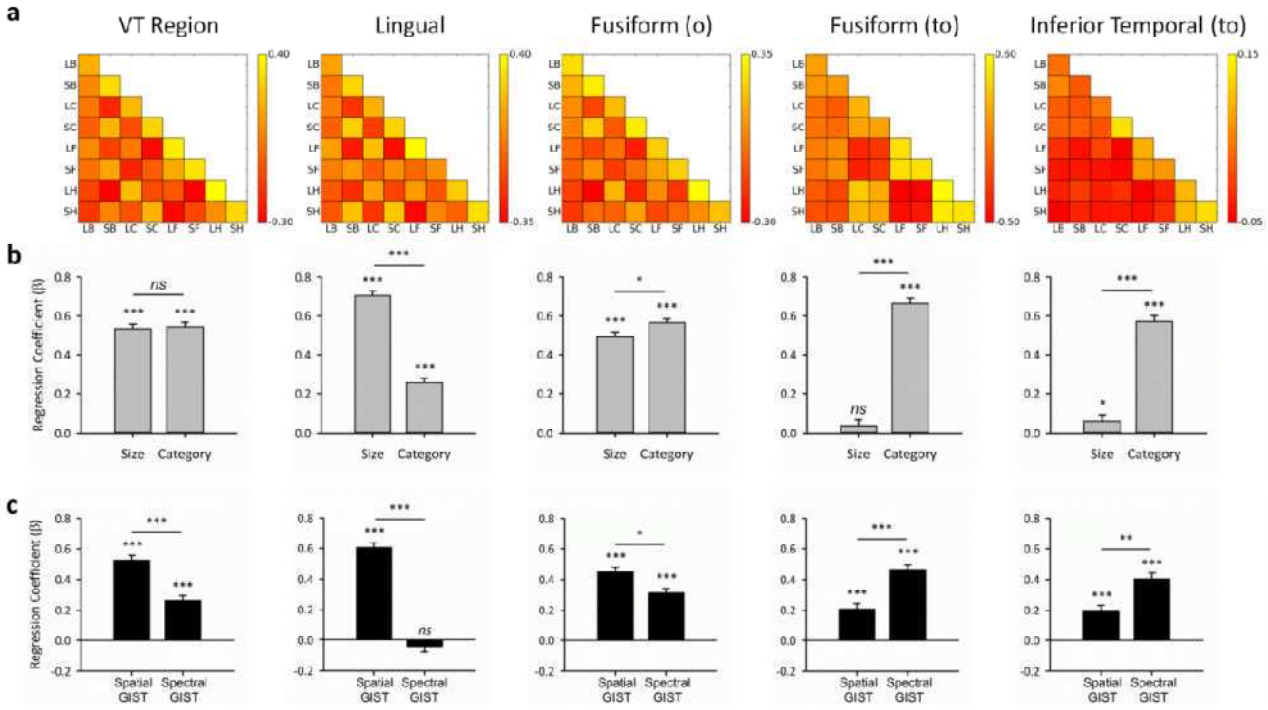


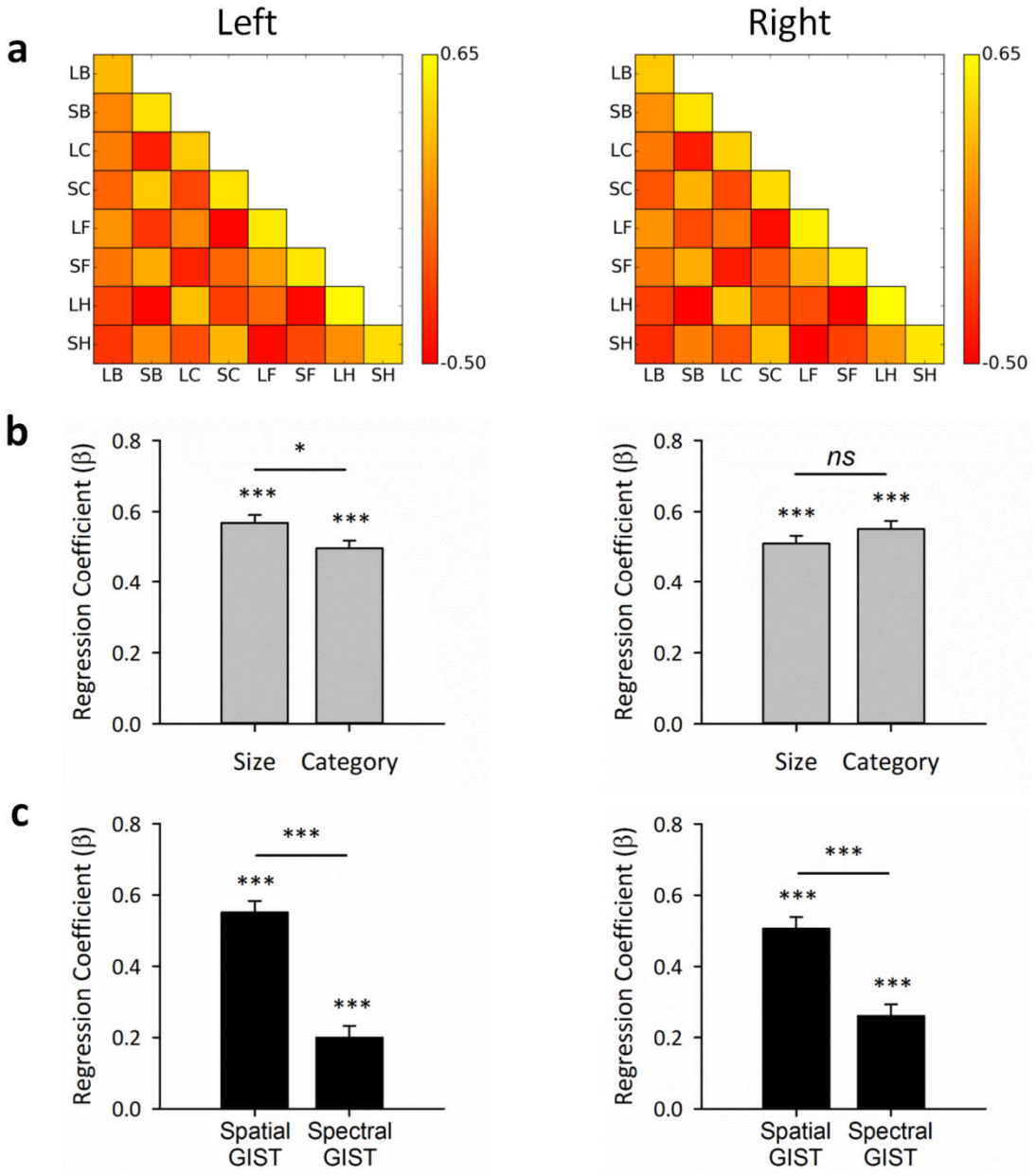
Figure 8. Analyses of the ventral temporal mask along posterior-anterior (a) and medial-lateral (b) axes. The ROI was split into 4 slices running along each axis, and pattern analyses run for each split separately. Mask divisions are shown in the leftmost column. Also shown are standardised coefficient values of multiple regression analyses contrasting category (red) and size (blue) models (middle column), and Spatial GIST (green) and Spectral GIST (orange) descriptors (right column). Coloured asterisks indicate the significance of the regressor coefficients, while black asterisks indicate the significance of the t-contrast between regressor coefficients at each slice. Error bars represent ±1 SEM.



Supplementary Figure 1. Results of within-participant analyses for ventro-temporal region and corresponding anatomical sub-regions. Data were split into odd and even blocks of the stimulus presentation and the neural response patterns correlated across the splits within each participant separately. (a) Correlation matrices averaged across participants. Standardised coefficient values of multiple regression analyses contrasting size and category models (b), and Spatial GIST and Spectral GIST descriptors (c). Error bars represent ±1 SEM.



Supplementary Figure 2. Intra-class correlations. A given participant was designated as a left-out participant, and the remaining participants designated as left-in participants. Response patterns of the left-out participant were correlated against response patterns of each of the 19 left-in participants in turn, yielding a total of 19 correlation matrices which were then averaged together into a single correlations matrix. This process was then repeated for using each participant as the left-out participant in turn. Thus response patterns are compared between individuals and the resulting correlation matrices are averaged across left-in participants, whereas the LOPO procedure averages the neural response patterns across left-in participants and then yields a single correlations matrix per iteration. (a) Correlation matrices averaged across iterations of the intra-class cross-validation procedure. Standardised coefficient values of multiple regression analyses contrasting size and category models (b), and Spatial GIST and Spectral GIST descriptors (c). Error bars represent ±1 SEM.



Supplementary Figure 3. Analysis of the ventro-temporal ROI split by hemisphere. (a) Correlation matrices averaged across participants. Standardised coefficient values of multiple regression analyses contrasting size and category models (b), and Spatial GIST and Spectral GIST descriptors (c). Error bars represent ±1 SEM.