Different Spatial Scales of Shape Similarity Representation in Lateral and Ventral LOC

We investigated the relationship between stimulus similarity for a set of parameterized shapes and the spatial scale of neural representation within subregions of the lateral occipital complex (LOC) using a carryover functional magnetic resonance imaging design. In ventral but not lateral LOC, a linear recovery from adaptation proportional to shape dissimilarity was seen. In contrast, a strong correspondence of the distributed neural pattern and stimulus similarity was observed in lateral LOC but not ventral LOC. Further, ventral LOC voxels were found to be broadly tuned and represent all aspects of stimulus similarity, whereas lateral LOC voxels were narrowly tuned and preferentially represented the shape of small features rather than their orientation within the shape. The results, indicating a coarse spatial coding of shape features in lateral LOC and a more focused coding of the entire shape space within ventral LOC, may be related to hierarchical models of object processing.

Keywords: fMRI, neural adaptation, object perception, pattern classification, shape perception

Introduction

A fundamental problem in vision is the need to reduce the myriad dimensions of input provided by the eye into an organized, structured representation (Edelman 1998). Regularities in what we perceive allow the representation of a high-dimensional world by a lower dimensional space; these representations are instantiated in the neural codes maintained within cortical visual areas. Different stages of processing in the visual stream presumably represent objects in different ways. In this paper, we measure the representation and spatial organization of these codes for simple 2-dimensional shapes, and what the relationship is between stimulus similarity and neural representation.

The spatial scale of neural codes for objects has been a subject of debate over the last decade. Spatially small regions of cortex have been found which preferentially respond to particular categories of visual stimuli; for example, faces (Kanwisher et al. 1997; Mccarthy et al. 1997), places (Epstein et al. 1999), general objects (Haxby et al. 2001; Haxby 2006), and body parts (Kanwisher et al. 1997; Mccarthy et al. 1997; Downing et al. 2001). These results imply a fine-scale, within-voxel representation of these stimuli, where small regions of cortex contain populations capable of representing the entire space of images in a category. Such a representation corresponds to Edelman’s (1998) “chorus of prototypes”: sets of neurons that represent specific regions in a shape space and, taken together, represent an entire space.

A counterpoint to this apparent specialization has been the demonstration that information regarding object category is also contained in the distributed pattern of voxel responses across and between these specialized regions (Haxby et al. 2001; O'Toole et al. 2005). These results, in contrast, demonstrate the existence of coarse-scale representations, where neurons in one region of cortex preferentially respond to one region in a representational space, whereas neurons in another correspond to a different region. This type of representation might correspond to the “chorus of fragments” model of Edelman and Intrator (1997), where individual properties of objects are represented by separate neural populations.

Our focus here is upon the representation of variations in stimulus identity within a simplified object category. Within the domain of behavioral studies of object perception, this has been approached by relating the perceptual similarity of stimuli to the properties of their underlying mental spaces (Attnave 1950; Shepard 1964; Garner and Felfoldy 1970; Garner 1974; Sattath and Tversky 1977; Shepard and Arabie 1979). The relative perceptual similarity of a group of stimuli can be mapped to a representation that has metric properties, in that similar stimuli are closer together in an abstract, representational space. Aspects of the underlying representational space, such as its dimensionality and distortions, inform us as to the nature of the representation. In practice, the structure of a parameterized space of shapes can be recovered from human behavioral responses (e.g., reaction times or similarity judgments) to pairs of those shapes, even when the subjects do not see the entire space in its veridical configuration.

This isomorphism between perceptual and behavioral similarity may extend to the neural representation of variations in object appearance as well; that is, the similarity of neural activity patterns evoked by stimuli may map onto the perceptual similarity of the stimuli (Edelman 1998). Op de Beeck et al. (2001) examined this possibility by recording responses of macaque inferotemporal (IT) neurons to 2D shapes. They found that the pattern of neural responses across IT neurons reflected the perceptual similarity of the stimuli and was ordinally faithful to the veridical parametric configuration of the shapes, never seen by the subjects (Op de Beeck et al. 2001).

Does a similar system of neural representation exist within human visual cortex? The human lateral occipital complex (LOC) shows similar functional properties to those previously ascribed to IT structures in the macaque. This region responds more strongly when a viewer is presented with images of parseable objects, as opposed to images that have no 2- or 3-dimensional interpretation, and appears largely indifferent to the method of object perception, for example, objects may be defined by luminance, texture, motion, or stereo difference.
(Grill-Spector et al. 1998). The LOC appears to be composed of 2 distinct, bilateral cortical areas: a more lateral region near the lateral occipital sulcus and a more ventral area near the posterior fusiform gyrus and occipital-temporal sulcus (Malach et al. 1995; Grill-Spector et al. 1999). The ventral LOC is also referred to as the posterior fusiform sulcus (pFS). These lateral and ventral subdivisions of LOC may have different functional properties, as suggested by the greater degree of neural adaptation to object identity that has been observed in the ventral region (Grill-Spector et al. 1999; Kourtzi and Kanwisher 2001). One significant goal in this study was the investigation of the representations hosted by these distinct areas of LOC.

In the current study, we applied the framework of stimulus similarity to investigate the neural representation of shape variation in human subjects and the spatial scale on which that representation occurs. Two recent studies have demonstrated a relationship between the perceptual similarity and the distributed pattern of neural activity in LOC (Op de Beeck et al. 2008; Haushofer et al. 2008), both using synthetic novel shapes to dissociate these patterns from categorical, semantic representations. In the current study, we examine neural representation of perceptual similarity on both a distributed, as well as focal, cortical scale within 2 subregions of LOC. We used a “carryover” functional magnetic resonance imaging (fMRI) design (Aguirre 2007) in which the stimuli to be examined are presented in a counterbalanced, unbroken stream, while the subject performs an orthogonal attention task. A continuous modulation of neural response proportional to stimulus similarity indicates the presence of a within-voxel population code using neural adaptation. Simultaneously, the distributed pattern of neural response evoked by each stimulus across voxels may be measured. This provides a measure of the similarity of neural representations for the stimulus set at different spatial scales, as indexed by neural adaptation and distributed pattern analysis.

During fMRI scanning, subjects viewed 16 different shapes defined by radial frequency components (RFCs; a series of sine waves of various frequencies describing perturbations from a circle; Zahn and Roskies 1972; Fig. 1). RFCs were at one time proposed as an organizing principle of shape recognition (Schwartz et al. 1983). Although this idea was later experimentally rejected (Albright and Gross 1990), RFCs are nevertheless a useful method of creating and parameterizing shapes. Because they are simplified objects, these RFC curves provide an evenly parameterized similarity space, without semantically associated categorical boundaries.

These 2-dimensional, closed contours were varied parametrically by modifying the amplitude (amount of perturbation) and phase (positioning of perturbations) of one particular frequency component. Parametric variations of shape related to changes in amplitude and frequency of a low-frequency component have been found to correspond to a 2-dimensional representational space (as determined by a multidimensional scaling [MDS] of similarity ratings), although changes in the phases alone of 2 low-frequency components were found by Cortese and Dyre (1996) to collapse into a single dimension. The previous work of Cortese and Dyre defined the perceptual properties of these stimuli and demonstrated that human observers organize their perception of these stimuli around the 2 axes. Additionally, although these 2 axes were found to be of equal salience, one axis was found to perceptually correspond to essential shape features, whereas the other axis appeared to change the orientation of features within the shape.

**Figure 1.** Stimuli. Sixteen unfilled, closed contours, constructed from RFCs. The amplitude and phase of frequency 6 was varied; the amplitude and phase of frequencies 2 and 4 were held constant at 0.50 radians and 0 degrees. All other frequency components were zero.

Does the similarity of the stimuli correspond to the similarity of the patterns of neural activity that they evoke? Neural adaptation (Grill-Spector and Malach 2001; Henson 2003; Henson and Rugg 2003) measures the habituation a neural population experiences when a stimulus is repeated. We asked in this study if the degree of recovery from neural habituation at different cortical sites was proportional to the transition in similarity between 2 stimuli. Any voxels with this response property would indicate the presence of a neuronal population able to represent the space of shapes, in a manner analogous to that observed in macaque area IT by Op de Beeck et al. (2001). Such a representation would exist at a relatively fine spatial scale, with the population of neurons within a voxel sufficient to represent the shape space. We might expect ventral area LOC to contain this form of representation.

Additionally, recent studies have shown that the pattern of neural response to visual stimuli, distributed across voxels, contains information regarding the category of stimulus (Haxby et al. 2001; O’Toole et al. 2005). In this study, we investigated if the distributed pattern of response can inform us to the identity of stimulus variation within an object category; results to this effect using man-made objects have recently been reported by Eger et al. (2008). As pointed out by Cox and Savoy (2003), the identification of such distributed patterns, which depend on between-voxel differences, indicates a given perceptual feature must be represented at a relatively coarse scale. We further wished to determine whether the relative similarity of these distributed neural patterns in turn reflects the perceptual similarity of the stimuli. Finally, we asked if the different perceptual features of the 2 stimulus axes are reflected in differences in neural coding at either a focal or distributed scale.
Materials and Methods

Subjects and Scanning Parameters

Five right-handed women aged 20-22 participated in the study. All subjects provided informed consent and the study conformed to the guidelines of the University of Pennsylvania Institutional Review Board. Structural and functional data were collected on a 3.0-T Siemens Trio scanner using an 8-channel head coil. High-resolution T1-weighted structural images were collected in 160 axial slices and near isotropic voxels (0.9766 x 0.9766 x 1.0000 mm; time repetition [TR] = 1620 ms, time echo [TE] = 3 ms, time to inversion [TI] = 950 ms). Functional, blood oxygenation level-dependent (BOLD), echoplanar data were acquired in 3 mm isotropic voxels (TR = 3000 ms, TE = 30 ms). BOLD data were acquired in 42 axial slices, in an interleaved fashion with 64 in plane resolution. The functional data were collected in 5 runs of 159 TRs each. The first 6 s of each run consisted of “dummy” gradient and radio frequency pulses to allow for steady-state magnetization during which no stimuli were presented and no fMRI data were collected.

Stimuli and Behavioral Task

Stimuli were 16, simple closed contours (Fig. 1) constructed from RFCs similar to those used by Cortese and Dyre (1996). Specifically, the RFC-amplitude of frequency 6 was 0.25, 0.50, 0.75, or 1.00 radian, and the RFC-phase of frequency 6 was 0, 40, 80, or 120 degrees. The RFC-amplitude and RFC-phase values of frequencies 2 and 4 were held constant at 0.50 radians and 0 degrees. Each contour was drawn on a mean gray background in either red or purple (randomly selected upon each presentation). The presentation of only the shape outline allowed us to distinguish between the similarity of the stimuli in pixelwise or retinotopic measures and the similarity of the contour implied by the outline. The stimuli were back projected onto a screen viewed by the subject through a mirror mounted on the head coil and subtended 5° x 5° of visual angle. Each stimulus was presented for 1400 ms, with a 100 ms ISI consisting of the mean gray background (Fig. 2). The subject was instructed to indicate on each trial, by button press, whether the contour was drawn in red or purple. The task was assigned solely for the purpose of requiring the subject to attend to every stimulus in the experiment and was constructed so as to not involve an explicit judgment of any aspect of the stimuli that was of experimental interest. All subjects performed above 96% accuracy, and the mean accuracy was 98%, indicating that subjects were alert and monitoring the stimuli as they were presented. There was no effect on RT of stimulus or stimulus-previous similarity (P = 0.32).

Stimulus Sequence

Each of the 16 different shapes was presented to each subject 85 times in a fully counterbalanced order. The order of stimulus presentation was determined by an n = 17, “type 1 index 1” sequence, a first-order counterbalanced ordering that arranges the stimuli in permuted blocks (Nonyane and Theobald 2008). The full sequence was divided into 5 parts for scanning as described in Aguirre (2007). The labels 1-16 were assigned to the 16 stimuli and the 17th label indexed the presentation of a blank trial (gray screen with fixation cross), which had a duration of 3 s (Appendix A: Aguirre 2007). This sequence provides for first-order counterbalancing of the stimuli, such that every image appeared in the sequence both before and after every other image as well as before and after 3 s of a blank screen. A particular type 1 index 1 sequence was selected that maximized efficiency (Friston et al. 1999) for detection of adaptation effects proportional to stimulus similarity. This sequence was identified by brute force search of several hundred thousand sequences (Appendix A: Aguirre 2007) and can be obtained from our Web site (http://cfn.upenn.edu/aguirre/projects/premade.shtml).

Image Preprocessing

Offline data analysis was performed using VoxBo (http://www.voxbo.org) and SPM2 (http://www.filion.ucl.ac.uk/) software. Data were since interpolated in time to correct for the slice acquisition sequence, motion corrected with a 6-parameter, least squares, rigid body realignment routine using the first functional image as a reference, and normalized in SPM2 to a standard template in Montreal Neurological Institute (MNI) space. Normalization maintained 3 mm isotropic voxels and used fourth degree B-spline interpolation. In the analysis of adaptation effects, the fMRI data were smoothed in space with a 3 x 3 x 3 voxel isotropic Gaussian kernel. In the support vector machine (SVM), distributed analysis, the data were left unsmoothed. For each dataset (the spatially smoothed and unsmoothed), the average power spectrum across voxels and across scans was obtained, and the power spectrum fit with a 1/frequency function (Zarahn et al. 1997). This model of intrinsic noise was used during regression analyses with the modified general linear model (Worsley and Friston 1995) to inform the estimation of intrinsic temporal autocorrelation.

The results of group analyses were presented (using BrainVoyager; http://brainvoyager.com) atop the MNI anatomical image that served as a template for spatial normalization. Regions of interest (ROIs) corresponded to early retinotopic visual areas (V1, V2, V3, hV4) and categorically organized areas (LOC dorsal and ventral, identified by response to object > scrambled object) were defined from data obtained during separate scans using standard methods (Harris and Aguirre 2008; Radoeva et al. 2008). The ROI analyses reported here combined data from the left and right hemispheres with the exception of the ventral LOC ROI for which a difference between the left and right hemisphere responses was found.

Statistical Analysis of Adaptation Effects

In order to analyze the adaptation effects, we created a set of 3 covariates modeling the interstimulus distance along each axis in the

---

**Figure 2.** Stimulus presentation. A sequence of contours and blank trials were shown for 1400 ms each, separated by 100 ms intertrial intervals. The stimulus on each trial was randomized to appear in either a red or purple hue on a gray background. The subject was instructed to indicate on each trial, by button press, the color of the stimulus.
shape space at each point in time (Aguirre 2007). Two covariates modeled the distance along the axes RFC-amplitude and RFC-phase, respectively. We assumed a linear, 4 by 4 spacing of stimuli. In MDS analysis of behavioral data, Cortese and Dyre (1996) found their stimuli to be placed in a reasonable simulacrum of a linear grid; we replicated this study in 11 subjects who performed a similarity rating task for the stimuli. We found that all subjects reliably produced a grid-like arrangement of the stimuli. As the average distance matrix generated by this behavioral data correlates $R = 0.93$ with a simple linear grid, we felt justified in using the simplified model; this also permitted decomposition of the adaptation response into the RFC-amplitude and RFC-phase components. Additional covariates, not of interest in this study, modeled the main effect of stimulus presentation as compared with the blank trials, the effect of exact repetition of stimulus identity, and the effect of a stimulus following a blank trial (for details, see Aguirre 2007). Nuisance covariates, corresponding to the effects of global signal, motion, and the orthogonal attention task, were included in both this analysis and the analysis of distributed effects. In an additional analysis, the 2 covariates corresponding to RFC-amplitude and RFC-phase were replaced with a set of 6 covariates that modeled the 6 possible sizes of city-block transitions in the shape space.

Using these covariates, we examined the linear relationship between shape similarity and neural adaptation within the functionally defined subregions of LOC. Exploratory group results were also obtained for a whole-brain, random-effects analysis and thresholded at a map-wise significance of $\alpha = 0.05$ as determined by a permutation test (Nichols and Holmes 2002) ($t > 3.5$ with a cluster $> 50$ voxels). For the analysis within ROIs, the voxels in each ROI with the largest main effect (i.e., the contrast of all stimuli vs. blank) were selected and averaged. This was done to maintain parity across the ROIs and with the analysis of distributed patterns.

**Statistical Analysis of Distributed Effects**

In a separate analysis, the distributed pattern of neural activity associated with each stimulus was obtained. Each stimulus appeared 85 times during the experiment. These 85 presentations were randomly assigned to 1 of 5 groups, with the constraint that an equal number of presentations were included from each scan (to avoid scan effects being a learnable factor in the subsequent analysis). A set of covariates modeled the identity of the stimulus being viewed for each group. These 80 covariates (10 stimuli \( \times \) 5 groups) modeled each stimulus presentation as a neural impulse, convolved with a standard hemodynamic response function (Aguirre et al. 1998).

An SVM was then used to classify the average brain activation map associated with each stimulus. For each voxel, we obtained the 5 average responses to a given shape over the groupings of its presentations. For all possible pairings of the stimuli, the linear SVM classifier (Joachims 1999) was trained and tested in a leave-one-out manner using 4 of the 5 groups for each stimulus (Our code for queuing SVM analyses and organizing the results is available at http://cfn.upenn.edu/aguirre/wiki/support_vector_machines). For each voxel, the classifier exposes a value related to the amount of variation of that voxel between conditions; for example, how useful that voxel was at discriminating between conditions, across all the pairwise comparisons. This is called the $w$-map; the map of these across the brain is called the "w-map." The $w$-map for each subject was transformed across voxels. A group $w$-map was created by smoothing (with a 3-voxel Full Width Half Maximum kernel) the $w$-map from each subject and then averaging across subjects.

A further analysis examined the similarity of distributed patterns of neural activity associated with the 16 different stimuli within different ROIs. Within each of the studied ROIs, we selected the 50 most discriminatory voxels (i.e., those with the highest $w$-value as identified by SVM). The average vector of stimulus beta values across these voxels was then obtained and subtracted from each voxel vector. A neural similarity matrix was constructed by calculating the Pearson correlation between the vector of beta values across voxels for each stimulus and every other stimulus. As the matrix is symmetric about the diagonal, only the lower triangle was retained, and the diagonal elements were excluded (as these have an obligatory value of unity and are thus uninformative). We then asked how well the neural similarity matrix in each region was correlated with the stimulus similarity matrix as defined by Cortese and Dyre (1996) and with its decomposed elements (RFC-amplitude and RFC-phase).

**Measurement of Voxel Tuning**

For each voxel in ventral and lateral LOC for each subject, we identified the average amplitude of BOLD fMRI response to each of the 16 stimuli, expressed as a 4 by 4 matrix (the response profile). The number of voxels with maximal responses to particular stimuli defined a histogram of peak stimulus responses for each ROI. To determine mean voxel tuning, all the response profiles in a given ROI for a subject were averaged together after aligning each 4 by 4 matrix within a 7 by 7 matrix, such that the center cell held the maximum value. The response to each of the 15 nonpeak shapes was scaled as a proportion of the range prior to averaging. The region tuning functions were then averaged across subjects. The center value was omitted from plots as it had an obligatory value of unity. Finally, we examined the degree to which neural adaptation varied as a function of the tuning of a voxel. The stimulus that elicited the maximum response within the response profile of each voxel was identified. Separate covariates modeled the adaptation associated with transitions between stimuli that were adjacent in the stimulus space but either included (i.e., were proximate to) or excluded (i.e., were distant from) the stimulus identified as eliciting the largest average response. The average difference in the degree of adaptation elicited by proximate and distant stimulus transitions was obtained for a given region and subject and the average across subjects then obtained.

**Results**

**Continuous Neural Adaptation in Ventral LOC Is Proportional to Shape Similarity**

We first measured the strength of the relationship between within-voxel neural adaptation and stimulus similarity. Consistent with previous work and our predictions, the largest and most significant effect was found within ventral area LOC on the right ($t_4 = 10.5$, $P = 0.0005$), whereas the lateral component of area LOC showed essentially no consistent adaptation effect proportional to perceptual similarity ($t_4 = 1.0$, $P = 0.4$). The difference between these subregions of area LOC was significant ($t_4 = 4.9$, $P = 0.008$; Fig. 3A). These effects were present in all 5 subjects and individually significant in 4 of the 5 (linear effect in ventral LOC for each subject: $P = 0.05$, 0.01, 0.03, 0.09, and 0.04), as well as at the population level. The adaptation effect for exact repetition was not significant in either ROI. This result is consistent with recent findings that infrequent, exact repetition is associated with an attenuated neural adaptation response (Summerfield et al. 2008).

An exploratory whole-brain group analysis of these data (Fig. 3D) supported the result found within the LOC ROIs. Most prominent was a significant proportional adaptation effect seen in the right ventral pFS corresponding to ventral area LOC. No adaptation effects were seen within the area of the lateral LOC.

To confirm that the modulatory effect of stimulus context upon neural response was linearly related to the change in stimulus similarity, we obtained the average BOLD response to each stimulus as a function of the size of transition in the shape space from the prior stimulus. The steady increase in neural response seen in the right, ventral LOC (Fig. 3C) is well fit by a linear function. In contrast, the lateral LOC did not evidence a systematic recovery from adaptation over the range of shape changes.

An alternative explanation for the proportional recovery from adaptation in ventral LOC is that the extreme stimuli (those from the corners of the stimulus space) may evoke...
a larger neural response generally (e.g., Kayaert et al. 2005). As the larger distance stimulus transitions tend to include these extreme stimuli to a greater extent, perhaps the apparent recovery from adaptation is actually a larger response to these extreme stimuli independent of an adaptation effect. To evaluate this possibility, we again measured the degree of BOLD response to each stimulus as a function of the size of transition from the previous stimulus. Additionally, however, transitions that included one of the extreme, corner stimuli were modeled separately from those that did not include these stimuli. We confirmed that a linear recovery from adaptation was seen for transitions that either excluded (t = 3.3, P = 0.03) or included (t = 3.74, P = 0.02) the stimuli from the extremes of the stimulus space and that the degree of recovery did not differ between these 2 sets (t = 0.98, P = 0.4). Therefore, the proportional recovery from adaptation seen in ventral LOC indicates the presence of a population code for stimulus shape and cannot be attributed to a generally greater neural response to extreme stimuli.

Distributed Pattern Responses Distinguish between Shapes

We next used an SVM classifier to analyze our data at a coarse spatial level. We first analyzed the data to identify the across-voxel pattern of activity evoked by each stimulus for each subject. As the order of stimulus presentation was counterbalanced, this measure of the average response across trials is independent of first-order context and thus not influenced by short-term adaptation effects. The distributed activation pattern for each stimulus was then used in leave-one-out training of an SVM classifier to determine classification accuracy and the location of voxels that contributed to successful classification. The ability to classify the stimuli based upon the voxel-wise activity pattern constitutes evidence of a coarse scale of representation of shape, where different stimuli evoke patterns of activation that differ in their amplitude between voxels.

Within each subject, the SVM analysis was successful at distinguishing between pairs of the 16 stimuli using the whole-brain response maps. For each subject, a decision matrix is generated containing the 120 values (hyperplane separation distances) associated with each pairwise decision, where this value is larger when the classifier is more certain of its ability to separate the pair of neural patterns. On average, in a given subject, the SVM classifier was able to distinguish between the neural patterns elicited by a given pair of stimuli with 71% accuracy (standard deviation ±11%). When the decision matrices are summed (i.e., the classifier is permitted to use the hyperplane distances from all subjects in its decisions), the pairwise accuracy rises to 89%.

To identify which cortical areas most contributed to classification accuracy, we obtained the average (across pairwise comparisons) of the discriminatory contribution (t value) for each voxel for each subject. These t-maps were normalized to z values across voxels, spatially smoothed, and

---

**Figure 3.** Focal pattern similarity. (A) The average across subjects of the amplitude of linear recovery from adaptation associated with stimulus transitions in ventral and lateral LOC. (B) Random-effects analysis of recovery from adaptation, linearly related to perceptual shape similarity. Warm colors indicate significant recovery from adaptation associated with stimulus change. (C) In a 4 by 4 configuration of stimuli, 6 (city block) step sizes are possible. In ventral but not lateral LOC, increasing step sizes are associated with increasing BOLD response (zero on the y-axis corresponds to the average response across all stimulus transitions). Greater error in higher step sizes is due to fewer possible exemplars of those changes. Error bars reflect variability across subjects. The R value of the best linear fit is given for each region.
then averaged across subjects. The resulting map (Fig. 4B) indicates the cortical location of voxels that tended, across subjects and across stimulus pairs, to be most informative regarding stimulus identity in the SVM analysis.

Bilateral discriminant patches were found in the lateral component of LOC, although not in the ventral areas. Given their identification in this analysis, these lateral regions arguably contain a coarse neural code in which individual voxels have different levels of activation associated with the different stimuli, thus allowing the classifier to associate a specific stimulus with a specific neural activation pattern. Thus, in these areas, the success of the SVM classifier reveals the spatial nature of the neural population: a coarse, across-voxel pattern carries information about shape.

The Similarity of Distributed Pattern Responses Reflects the Similarity of Shapes

The accuracy of the SVM analysis and the identified patch within lateral LOC indicates that the distributed voxel pattern of activity in that area carries information about shape. However, the pattern difference between shapes need not reflect the similarity of the stimuli or indeed have any particular structure. The SVM requires only that patterns be different in order to distinguish them—no assumptions about similarity structure are made or used. We wished to test the further hypothesis that the similarity of the distributed neural pattern evoked by any pair of stimuli would reflect their similarity.

Behavioral testing (Cortese and Dyre 1996) can be used to define the perceptual similarity matrix for a set of stimuli. The left panel of Figure 4C shows the similarity matrix for our stimuli (Fig. 1). Each cell of this $16 \times 16$ matrix expresses (by color scale) the similarity of a pair of stimuli. The visible structure to the matrix follows directly from the arrangement of the stimuli within the perceptual space (Fig. 1). We created a “distributed neural similarity matrix” for lateral and ventral LOC for each subject. The neural similarity matrix is constructed by assigning each cell of the matrix to the Pearson correlation of the across-voxel pattern of response (for the 50 most discriminant voxels within a region) associated with a particular stimulus pair. The entire matrix captures the pairwise similarity structure of the entire set of stimuli, as

![Figure 4. Distributed pattern similarity. (A)](image)

(A) The average across subjects of the correlation of the distributed neural similarity matrix with the stimulus similarity matrix is shown. The distributed pattern within lateral LOC was significantly more correlated than in ventral LOC. (B) Average $w$-map indicating the amount of variation (informativeness) in each voxel’s activity due to change in stimulus. The map from each subject was z-transformed across voxels, spatially smoothed, and then the average across subjects obtained. (C) Shown are stimulus and neural similarity matrices. The color of each cell reflects how similar a given pair of stimuli are as measured by behavioral responses (left) or distributed neural codes (right). The diagonal of the matrix is undefined for identical pairs of stimuli. Shown below the 2 neural similarity matrices is the correlation of the entire neural matrix with the stimulus similarity matrix. Ventral LOC carries little information about similarity in the distributed pattern and is therefore weakly correlated with the perceptual matrix. By contrast, in lateral LOC, the similarity of the distributed patterns associated with each stimulus pair is strongly and linearly related to their perceptual similarity, yielding strong correlation with the perceptual matrix. The diagonal symmetry of the matrices is enforced by the nature of their construction.
instantiated in the neural responses they evoke. The correspondence between the measured neural similarity matrix for each region and the stimulus similarity matrix for the stimuli was then obtained.

Within lateral LOC, the strongly discriminant responses seen in the SVM analysis were found to also reflect stimulus similarity consistently across subjects (t_R = 10.0, P = 0.001). In contrast, the distributed pattern of response in ventral LOC had a weaker correlation with the perceptual similarity of the stimuli (t_R = 1.2, P = 0.3) (Fig. 4A). The difference between these subregions of area LOC was significant (t_R = 11.4, P = 0.0003).

This regional difference is also visible in the average neural similarity matrix obtained across subjects from within the ventral and lateral LOC. Figure 4C demonstrates that the average neural similarity matrix from the lateral LOC has definite structure and a strong correlation (R = 0.65) with the stimulus similarity matrix. Notably, there are aspects of the structure of the neural similarity matrix that do not seem reflected in the stimulus matrix; the source of this difference is explored in the next section. The average neural similarity matrix for ventral LOC had a weaker correspondence to stimulus similarity (R = 0.19). Using all voxels in each of these ROIs, rather than only the 50 most discriminant, produced similar results (lateral LOC R = 0.64; ventral LOC R = 0.21).

Could the distributed pattern in lateral LOC be explained entirely by simple retinotopic organization? Because we used unfilled shape outlines, the pixel-wise similarity between shape pairs was quite low, and the correlation between the pixel-wise similarity of the shapes and the distributed neural pattern in lateral LOC was 0.18, far lower than the 0.65 correlation seen with the perceptual similarity of the shapes (Fig. 4C). Another possibility is that the shape outlines are perceptually "filled in," as has been observed to occur in neuronal responses (Lamme et al. 1999). The correlation between the pixel-wise similarity of the contours filled and distance in the shape space is fairly high (R = 0.85), making this explanation plausible.

Other aspects of the results, however, render a solely retinotopic account incomplete. First, there is a weak correspondence between distributed neural and perceptual similarity in adjacent ventral cortical areas with much stronger retinotopic organization (V2/V3v R = 0.23; hV4 R = 0.24). Second, as will be discussed in the next section, differences in the strength of representation of the 2 stimulus axes (RFC-amplitude and RFC-phase) cannot be explained on a retinotopic basis.

The RFC-Amplitude and RFC-Phase Axes Are Differentially Represented at Coarse and Fine Neural Scales

Although the distributed neural similarity matrix measured from lateral LOC was strongly correlated with the stimulus similarity matrix, there appeared to be aspects of the structure of the neural response not evident in the stimulus matrix (Fig. 4C). We considered the possibility that this difference is explained by differences in the neural representation of 2 axes that define the stimulus space. Cortese and Dyre (1996) showed that vectors in their behavioral MDS fitted for RFC-amplitude and RFC-phase were approximately orthogonal and equal. This indicates that the 2 perceptual axes define the perceived shape space and that each axis is equally perceptually salient. It is not the case, however, that the axes are perceived equivalently. Cortese and Dyre found that parame

Neural Adaptation within Lateral LOC Is Modulated by Narrow Tuning for Shape

Based upon the differential sensitivity to shape identity for the adaptation and distributed pattern methods, we argue that although both the lateral and ventral components of area LOC contain neural population codes for shape, the spatial scale of these representations differ. Specifically, the absence of a distributed pattern effect within ventral LOC is evidence for a homogeneous representation of the shape space, such that the average response of any one voxel does not differentiate between the shapes, whereas the presence of a distributed code and the absence of an adaptation effect in lateral LOC suggests that there is a heterogenous distribution of shape representation, such that any one voxel tends to respond only to a limited area of the shape space.
An objection to this account is that adaptation and distributed pattern methods might be expected to discriminate between the stimuli regardless of the underlying spatial distribution of the neural populations. For example, whereas the neurons responsive to particular line orientations are distributed within V1 on a scale smaller than individual voxels, multi-voxel pattern methods are nonetheless capable of recovering stimulus information from fMRI data (Kamitani and Tong 2005), presumably because voxels retain some tuning to particular line orientations. Further, one might expect that even in the presence of a coarse code for shape identity, the small adaptation effects obtained in different voxels would, on average, produce a recovery from adaptation for a region that reflects stimulus similarity. Finally, the results so far cannot distinguish between the possibility that lateral LOC has clusters of neurons tuned to a particular location in the shape space and the possibility that clusters of neurons are tuned to particular parts of the shapes.

To address these issues, we examined the tuning of individual voxels to the shape space and the effect that tuning had upon neural adaptation. First, our proposal that ventral and lateral LOC represent the RFC-shapes on different spatial scales would predict that voxels drawn from lateral LOC would have narrow tuning for particular shapes within the shape space, whereas voxels from ventral LOC would be broadly tuned. If ventral LOC voxels are so broadly tuned to shape identity that the amplitude of the BOLD response is roughly the same across shapes, then the relatively weak distributed representation within this region will be explained.

For each voxel for each subject, we obtained the response profile that measures the amplitude of the BOLD fMRI response to each stimulus (independent of adaptation effects). An example response profile for one voxel is shown in Figure 6A. For each voxel, one shape will have evoked the maximum observed response. Figure 6B presents the histogram of peak responses for ventral and lateral LOC across subjects. Stimuli from the extremes of the stimulus space tended to have a greater representation in maximal voxel responses (effect of stimulus upon proportion of voxels: \( F_{15,159} = 6.08, P < 0.0001 \)). However, the distribution of maximally responsive voxels was not different between ventral and lateral LOC (stimulus by region effect: \( F_{15,159} = 1.07, P = 0.39 \)).

The aspect of voxel response most relevant to a distributed pattern analysis, however, is the tuning of the response of the voxel across stimuli. We obtained the average tuning of voxels in each ROI across subjects (Fig. 6C). These plots show the decline in BOLD response for an average voxel for the presentation of shapes that are progressively more distant from the shape that evokes the greatest response for the voxel. Within ventral LOC, no meaningful tuning for the shape space...
can be identified: The amplitude of the response is no different for different shapes. This indicates that ventral LOC voxels are broadly tuned for shape identity. In contrast, lateral LOC voxels show relatively narrow tuning: there is a progressive decline in the response of a voxel for shapes more distant from the shape for which the voxel is best tuned (which was frequently a stimulus from the edges of the stimulus space). Moreover, lateral LOC voxels appear more narrowly tuned for the RFC-amplitude, as compared with the RFC-phase dimension of the shape space, consistent with our previous observation that the RFC-amplitude dimension is more strongly represented in the distributed pattern of response within this region.

Figure 6. Voxel tuning and proximate versus distant adaptation. (A) The average BOLD fMRI response in a single voxel (shown inset) to each of the 16 stimuli. Data points are indicated by black dots and connected by a red surface. The shape that elicited the largest response is indicated in yellow. (B) The proportion of voxels within each ROI that responded maximally to each of the 16 different shapes. (C) The average response of voxels to stimuli that differ from the stimulus that elicited the largest response, expressed as a proportion of the range of response. Black data points and the red surface are the average across subjects, and the transparent blue surfaces are ± standard error of the mean across subjects. The x-axis and y-axis are expressed in units of steps along the stimulus axes away from the maximum stimulus. The central value is omitted and had an obligatory value of unity. (D) For the response profile shown in panel (A), the stimulus indicated in yellow elicited the maximum response on average across all presentations. We might consider that the neural adaptation associated with small stimulus transitions that involve this stimulus (red) will differ from that measured for stimulus transitions that do not include this stimulus (black). The top figure illustrates these different transitions in the stimulus space, whereas the bottom figure shows the transitions in the context of a similarity matrix. (E) The neural adaptation associated with small stimulus transitions was measured in the lateral and ventral LOC. Shown is the across subject average difference in adaptation between transitions that were proximate to, or distant from (red and black in panel D, respectively), the center of tuning for each voxel.
The differential tuning of voxels within lateral and ventral LOC to shape is a sufficient explanation for the differences in distributed pattern representation in the 2 regions. The narrow tuning observed in lateral LOC may also explain the absence of a linear adaptation response in this region to transitions in shape space. If a given voxel is narrowly tuned to a particular region of the shape space, then it may only show recovery from adaptation for stimulus transitions within its tuned area. The example voxel shown previously (Fig. 6A) is tuned to maximally respond to a particular stimulus, indicated in yellow. Some transitions between stimuli will occur within the center of the tuning for this voxel (Fig. 6B, indicated in red), whereas some stimulus transitions of equal magnitude will occur distant from the tuned shape (indicated in black). For each voxel for each ROI, we measured the degree of neural adaptation associated with stimulus transitions of equal perceptual sizes that were located either proximate to or distant from the tuned center of the response profile for a voxel (Fig. 6E).

Within ventral LOC, the tuning of a voxel had little effect upon the degree of adaptation seen for stimulus transitions. This is in keeping with the broad tuning of the response of each voxel and is again consistent with the notion that voxels within ventral LOC contain populations of neurons capable of representing the entire shape space. In contrast, voxels in lateral LOC showed greater adaptation to small stimulus transitions when those transitions occurred proximate to the center of the tuning of voxel. The absence of a linear recovery from adaptation for the entire lateral LOC region can therefore be understood as a consequence of the restricted tuning of the neurons contained in any one voxel. Each lateral LOC voxel is only able to represent by adaptation small stimulus transitions that occur within its tuned area. A reexamination of the recovery from adaptation related to the magnitude of stimulus change for lateral LOC (Fig. 3C, right panel) is informative in this regard. A roughly linear increase in recovery from adaptation is seen for small stimulus transitions (step size 4 or less) but not for larger stimulus transitions.

Discussion
It has been suggested (Edelman 1998) that the neural representation of objects may be best characterized as representation of similarity, encoded by a chorus of prototypes of neural responses tuned to different regions of a shape space. In a 1998 study that anticipated the subsequent application of multivoxel analysis methods in fMRI by several years, Edelman et al. (1998) demonstrated that the pattern of voxel responses could be used to reconstruct the similarity representation of stimuli within and across object categories.

In our current study, we return to these ideas with the aim of examining the correspondence of perceptual and neural similarity in the representation of a set of simple shapes. This relationship has been explored at the across-voxel scale in 2 recent studies (Op de Beeck et al. 2008; Haushofer 2008). By using a continuous carryover design, our study was capable of examining neural similarity both on a coarse, across-voxel scale by distributed pattern analysis, as well as on a fine, within-voxel scale using continuous neural adaptation. We can thus compare the information provided at distributed and focal levels.

We found that shape similarity is represented on both scales, although the cortical sites that carry information at these 2 scales differ. Our focus here was upon the components of the LOC, a visual area with specific responses to formed visual objects. Within the ventral portion of LOC, we identified neural adaptation proportional to shape similarity, suggesting the existence of a population code for shape distributed on a fine, within-voxel spatial scale. This result parallels that of Op de Beeck et al. (2001) in their measurement of neural response similarity in macaque IT cortex to 2-dimensional shape variation. The weaker distributed (across-voxel) similarity pattern for shape within ventral LOC indicates that each voxel responds roughly equally to each shape, suggesting that the within-voxel code in this region generally spans the entire shape space. It is of course possible that scanning with higher spatial resolution or the use of other shape variations, perhaps with greater differences along the axes, would reveal heterogeneity in the focal population code across voxels and thus improve the performance of the distributed pattern measure. Indeed, this kind of heterogeneity has been demonstrated within ventral temporal cortex for objects of different categories, for which the pattern of across-voxel responses is discriminative even outside of categorically defined cortical areas (Haxby et al. 2001). A study that compared the recoverability of distributed pattern information for large and small ranges of shape variation would be one way to unify these findings.

In contrast, the lateral component of area LOC demonstrated a coarse (across-voxel) pattern of response that reflected shape similarity, the result of relatively narrow tuning of individual voxels to the shape space. This parallels the finding of Op de Beeck et al. (2008) of high correlation between similarity ratings and distributed neural patterns in area LO as compared with area PF. Several possible forms of coarse coding within lateral LOC could account for this result. Each voxel may contain neurons tuned to a particular subregion of the shape space, resulting in voxels that demonstrate a "receptive field" for a region of the shape space coded by firing rate. A related coding scheme would have patches of lateral LOC cortex tuned to particular shape features, with some voxels (e.g.) preferring tight curves and others concave line segments. It is also possible that simple retinotopic organization within lateral LOC is the basis of the distributed similarity found here, either alone or in combination with sensitivity to shape features. We regard it as unlikely, however, that retinotopic organization can alone explain this result for the reasons discussed earlier. Unlike ventral LOC, the lateral portion of LOC did not show adaptation responses that were linearly related to shape similarity. We found that the narrow tuning of lateral LOC voxels could explain this finding, indicating that each particular voxel has a population of neurons that are tuned to one specific region of the shape space. Consequently, most of the transitions between stimuli would not induce neural adaptation within the voxel as they would be transitions between stimuli not within the voxel's receptive field.

The presence of a coarse spatial code for shape within lateral LOC and a finescale code within ventral LOC suggests a processing hierarchy in which cortical patches within lateral LOC tend to represent features, whereas population codes in ventral LOC represent entire integrated shapes. This coarse lateral LOC representation would correspond to the chorus of fragments model (Edelman and Intrator 2000) in which representations of features are combined with a representation of fragment orientation or retinotopy to capture shape information. In support of this model, we found that the
distributed code for shape identity within lateral LOC reflected changes along the RFC-amplitude axis, which perceptually corresponded to changes in features, but not changes along the RFC-phase dimension, which corresponded to the orientation of features. Fragment or feature orientation may be implicitly represented in the low-resolution retinotopy that is present in lateral LOC (Dumoulin and Wandell 2008), in other visual areas, or in an aspect of lateral LOC response that is below the resolving power of our method. In any case, the form of the RFC-phase representation does not reflect shape similarity on a coarse scale.

Within ventral LOC, shape changes along both the RFC-amplitude and RFC-phase dimensions were strongly represented in the within-voxel population code, as indexed by continuous neural adaptation. The representation of shape in this ventral region is thus arguably more integrated, no longer representing individual features but instead the entire shape. This is not to say that a distributed pattern could not, in principle, be recovered from ventral LOC, given, for example, better resolving power, smaller voxels, or less spatial smoothing by the hemodynamic response function. Indeed, a modest correlation ($R = 0.21$) was found for the distributed pattern in ventral LOC in our study. However, our results demonstrate at least a difference between the scales of a neural pattern in the 2 ROIs.

Our results differ from prior studies of the neural correlates of shape adaptation (e.g., Kourtzi and Kanwisher 2001) in the use of a continuous adaptation design. Instead of paired presentations of stimuli that were either identical or different, we presented a continuous stream of stimuli. Although a linear recovery from adaptation proportional to stimulus dissimilarity was observed in ventral LOC, there was not significant adaptation for perfect stimulus repetition. This is not an anticipated feature of continuous carryover designs (Aguirre 2007). Behavioral studies have long demonstrated that different mental operations accompany the detection of same and different stimulus pairings (Sternberg 1998). Moreover, perfect stimulus repetition in this design is an infrequent and salient event, which has been observed to modulate neural adaptation (Summerfield et al. 2008). Although our inferences here do not depend upon identical stimulus repetition effects, we can imagine modifications of our design that would allow this component of the response to be more interpretable. For example, the addition of a stimulus variation that does not alter the similarity structure under study but does disrupt the salience of perfect repetition (e.g., rotation or misalignment of sequential stimuli) may be beneficial in this regard.

Our results may have relevance to the intriguing finding of paired, categorical regions of extrastriate visual cortex. Other categorically responsive visual areas, such as those responsive to faces, places, and body parts, appear to have both a ventral and a more lateral or dorsally located component (Schwarzlose et al. 2008). In the case of faces, a subregion of the LOC—the occipital face area (OFA)—has been identified in addition to the more ventral fusiform face area (FFA). It has been proposed previously that the OFA and the FFA differentially represent facial features and holistic gestalt, respectively (Haxby et al. 2000). Recently, we have shown that facial features are represented in both the OFA and the FFA, although only the FFA shows an adaptation response sensitive to holistic representation of familiar faces (Harris and Aguirre 2008). It remains to be seen if representation of features and wholes is a common property across these paired categorical visual areas.

A correspondence between shape similarity and focal and distributed neural similarity was also observed in cortical areas other than LOC. There was a significant linear adaptation effect in the vicinity of the transverse occipital sulcus bilaterally, corresponding in location to visual area V3a. This region also showed a distributed pattern for RFC-amplitude that weakly reflected shape similarity. Previous studies (Larsson and Heeger 2006) have observed object-preferential responses in area V3a, although responses to visual motion also drive this region (Wandell et al. 2005).

A recent study of distributed pattern similarity in ventral and lateral LOC (Haushofer et al. 2008) offers an interesting contrast to our results. Using a set of 4 shapes, the authors found that whereas the distributed response in lateral LOC reflected the physical similarity of their stimuli, the distributed response in ventral LOC (labeled pFS in their study) reflected a perceptual similarity. Interestingly, the ventral LOC correspondence was found to be quite sensitive to the idiosyncratic similarity judgments of different subjects. It is possible that the weak correspondence between stimulus similarity and distributed neural response in ventral LOC in our study is the result of between-subject differences in the perceptual similarity of the stimuli that were not modeled. Another possibility is that, as Haushofer et al. used only 4 stimuli and a Pearson correlation to judge correlation, a distinctive perceptual judgment and neural response to a single outlier stimulus was sufficient to produce the modest positive correlations in ventral LOC that they observed (as is shown for perceptual similarity in Figure 1B of their paper).

With 16 stimuli and thus 120 unique stimulus pairings in our study, a consistent distributed coding across the entire shape space would be needed to observe a correspondence between neural and perceptual similarity.

In summary, our results demonstrate generally that the similarity of patterns of neural response within higher order visual cortical areas can reflect stimulus similarity. By examining both within-voxel neural adaptation and across-voxel distributed patterns, we were able to identify substantial differences in the spatial scale and form of these representations across cortical areas. These differences, in turn, may be related to a hierarchy of the visual processing of shape that moves from a spatially coarse representation of features to an integrated representation of shape.

Funding
National Institutes of Health (K08 MH72926); Burroughs Wellcome Fund.

Notes
We thank Alison Harris for assistance in editing and Wesley Kerr for implementation of the SVM analysis. Sarah Drucker provided many insightful comments. Conflict of Interest: None declared.

Address correspondence to Geoffrey K. Aguirre, Neurology Department, University of Pennsylvania, 3400 Spruce Street, 3 West Gates, Philadelphia, PA 19104, USA. Email: aguirreg@mail.med.upenn.edu.

References

35:1480-1494.


