7 Population Coding of Object Contour Shape in V4 and Posterior Inferotemporal Cortex

Anitha Pasupathy and Scott L. Brincat

In primates, visual object perception and recognition is based on information processing within the ventral visual pathway (Ungerleider and Mishkin, 1982; Felleman and Van Essen, 1991). This multistage, hierarchical network runs from area V1 to V2, V4, and finally to the posterior and anterior subregions of inferotemporal (IT) cortex (Felleman and Van Essen, 1991). In each of these processing stages, visual information is encoded in the patterns of activity across populations of neurons. The neural code reflected in these activity patterns is gradually transformed across the ventral pathway into a representation thought to be optimized for robust object recognition, association, and memory. The nature of the neural code in the lowest-level stage, V1, is now relatively clear. Research over the last half-century has identified local orientation and spatial frequency as the basis dimensions of V1 form representation (Hubel and Wiesel, 1959, 1965, 1968; Baizer, Robinson, and Dow, 1977; Burkhalter and Van Essen, 1986; Hubel and Livingstone, 1987). Specifically, each small patch of the retinal image is represented by a local population of V1 neurons tuned to different orientations and narrow spatial frequency (scale) bands. The distribution of activity within these local V1 populations encodes the local orientation and scale of image features, and the full retinal image is represented by a tiling of the visual field with thousands of such local populations. Representation in terms of local orientation and spatial frequency first arises in V1, and these representational bases are fundamental to our understanding of what V1 does.

The bases of representation in subsequent processing stages are less clear, though past results suggest that activity in higher-level neurons reflects a progressively more nonlinear and complex function of the retinal image. V2 neurons, like their V1 inputs, are tuned for local orientation and spatial frequency, but also show some selectivity for line conjunctions (Ito and Komatsu, 2004, Hegde and Van Essen, 2000), orientation combinations (Anzai, Peng, and Van Essen, 2007), and illusory contours (von der Heydt and Peterhans, 1989). V4 neurons show selectivity for radial or concentric gratings and moderately complex shapes (Gallant, Braun, and Van Essen, 1993, Kobatake and Tanaka 1994). And in IT, many neurons exhibit

strong, position-invariant selectivity for complex objects like faces, hands, and bodies (Gross, Rocha-Miranda, and Bender, 1972; Perrett, Rolls, and Caan, 1982; Desimone et al., 1984; Tanaka et al., 1991). But a comprehensive understanding of neural coding in these areas—identification of the representational basis dimensions and the nature of the population code, as in V1—has not yet emerged. Quantitative analysis of higher-level object coding has been frustrated by (1) the high-dimensionality of "shape space," (2) the nonlinearity of response properties beyond V1, and (3) the practical constraints of neurophysiology experiments.

The ideal approach to investigating the basis of shape representation in any neuron would be to study its responses to a large set of complex naturalistic stimuli that uniformly sample the space of all possible object shapes. Naturalistic stimuli approximate the milieu in which visual neurons have evolved and developed, and are therefore likely to probe those shape dimensions most relevant to their complex response selectivities. Unbiased sampling would allow the use of analytical approaches with minimal assumptions—such as spike-triggered covariance—to extract the shape dimensions along which the responses vary maximally. However, since object shape varies along an extremely large number of dimensions, even a coarse sampling of shape space would require thousands of stimuli. Such an approach is impossible due to the time constraints on maintaining neuronal isolation and animal alertness during neurophysiological experiments, although advances in efficient stimulus sampling (Yamane et al., 2008) and long-term chronic recording (Tolias et al., 2007) may eventually make this approach more feasible. Several alternatives have been tried; white noise and natural stimuli define the two extremes. White noise stimuli have optimal sampling properties—they are completely unbiased and uncorrelated in visual image space—but are ineffective at evoking activity in neurons beyond V1. In contrast, arbitrarily chosen natural stimuli can often drive robust responses in higher-level neurons but tend to sample object shape space unevenly and with strong correlations between different stimulus dimensions, making quantitative determination of basis dimensions difficult (Rust and Movshon, 2005).

An intermediate alternative is to explore shape space in a limited, but systematic fashion in order to investigate a specific neural coding hypothesis. With this approach, stimulus sampling is targeted to the subregions of shape space relevant to a hypothesis of interest. Potential conclusions are limited to this subspace, but the advantages of such an approach are that it allows for (1) designing stimuli that are effective at driving neurons in high-level visual areas (unlike random white noise stimuli) and (2) the controlled variation of one dimension at a time, while holding all other variables constant (unlike arbitrary natural stimuli). Further, targeted stimulus design can provide a dense, unbiased sampling of the region of interest making it amenable to quantitative characterizations. We will describe our successful

use of this approach to discover relevant shape dimensions in V4 and posterior IT cortex.

A Hypothesis-Driven Approach to Identifying Basis Dimensions in V4

Our first task was to identify a plausible candidate hypothesis of what basis dimension(s) might be represented in area V4. For this purpose, we considered available evidence from theoretical work, psychological findings, and preliminary experiments from our lab. Several modern shape theories and computational models achieve object recognition by hierarchical feature extraction—objects are first decomposed into simple parts that are pooled at subsequent stages to form progressively larger and more complex parts. This approach has been reasonably successful and is broadly consistent with what is currently known about shape processing in the primate brain. Inspired by physiology, most shape models invoke local edge orientation as the first level feature. Higher-level primitives are based on either (1) object contour features, such as two-dimensional angles and curves and three-dimensional corners, curved surface patches, and indentations (Attneave, 1954; Milner, 1974; Ullman, 1989; Poggio and Edelman, 1990; Dickinson et al., 1992), or (2) volumetric primitives, including simple three-dimensional shapes such as cylinders, cones, and spheres, defined by the orientation of their medial axes and cross-sectional attributes (Biederman, 1987; Pentland, 1989).

Pilot physiological studies from our lab pointed to contour features as a candidate dimension for shape representation in area V4. In these exploratory experiments, we studied the responses of V4 neurons to a large set of simple shapes such as rectangles, squares, diamonds, triangles, crescents, ellipses, and circles, presented at multiple orientations (unpublished data). Most V4 neurons did not show selectivity to a particular shape or its area. Rather, cells typically responded to a variety of shapes that all contained a consistent contour characteristic at a specific position, such as a sharp convex point to the right. In a second study, V4 neurons showed strong systematic tuning for contour features, that is, angles and curves, presented in isolation in the receptive field of the cell (Pasupathy and Connor, 1999). Psychological findings also imply specialized mechanisms for the perception of contour features: angle perception acuity is higher than that predicted by component line orientation acuity (Heeley and Buchanan-Smith, 1996; Chen and Levi, 1996; Regan, Gray, and Hamstra, 1996), the detection thresholds for curvilinear glass patterns is much lower than that for radial glass patterns (Wilson, Wilkinson, and Asaad, 1997; Andrews, Butcher, and Buckley, 1973), and detection of curved targets among straight distractors is faster than that for straight targets among curved distractors (Treisman and Gormican, 1988; Wolfe, Yee, and Friedman-Hill, 1992). Thus,

psychophysical and preliminary physiological studies suggested that intermediate-level primitives are likely to relate to object boundary curvature.

We therefore set out to test whether contour curvature serves as the basis of representation in area V4. If a stimulus dimension is a basis of a given cortical area's representation, then the area's neural population should encode every visual stimulus in terms of that dimension. Second, within every patch of cortex representing a small region of the visual field in the area, neuronal tuning should span the full range of possible values along the dimension. In V1, for example, each visual stimulus is represented in terms of its local orientation and spatial frequency, and the neural population within each topographical patch (hypercolumn) contains cells tuned to all orientations and the full range of perceivable spatial frequencies. To investigate whether contour curvature meets these criteria in area V4, we asked whether (1) single V4 neurons encode complex shapes in terms of their contour curvature; (2) V4 neurons represent all possible curvature values; and, therefore, (3) a population of such V4 neurons provides a complete and accurate representation of all shapes in terms of their boundary conformation.

To test the hypothesis of curvature coding in V4, we designed a set of complex shape stimuli based on systematic sampling and combination of contour curvatures. We started with a set of five different contour segments: three convex (sharp, medium, and broad; see enlarged stimulus 2 in figure 7.1) and two concave (shallow and medium). We then created closed shapes based on all geometrically feasible combinations of four to eight of these contour segments separated by 90°, 135°, or 180°. Each of these shapes was presented at eight global orientations, separated by 45° intervals. The resulting stimulus set (figure 7.1) consisted of 366 complex shapes that vary systematically in boundary contour shape. This set represents a dense, unbiased sampling of the subregion of shape space defined by solid silhouette shapes that radiate out from the center of the object. Of course, other stimulus attributes, such as 3D-depth, texture, shading, color, or more complex contour topologies (e.g., holes or folds) are not sampled in this set—these attributes are not directly relevant to the hypotheses being tested here, and we leave it to future work to explore how they might also modulate V4 responses.

Shape Representation in Terms of Contour Curvature in Area V4

We studied 109 V4 neurons using this set of parameterized shape stimuli. Figure 7.1 shows the responses of a typical V4 cell. The background grayscale of each icon reflects the neuron's average response to the overlaid stimulus object. This neuron responded strongly to a wide variety of shapes containing a sharp convexity in the lower left corner of the object (angular position 225°) with an adjoining concavity

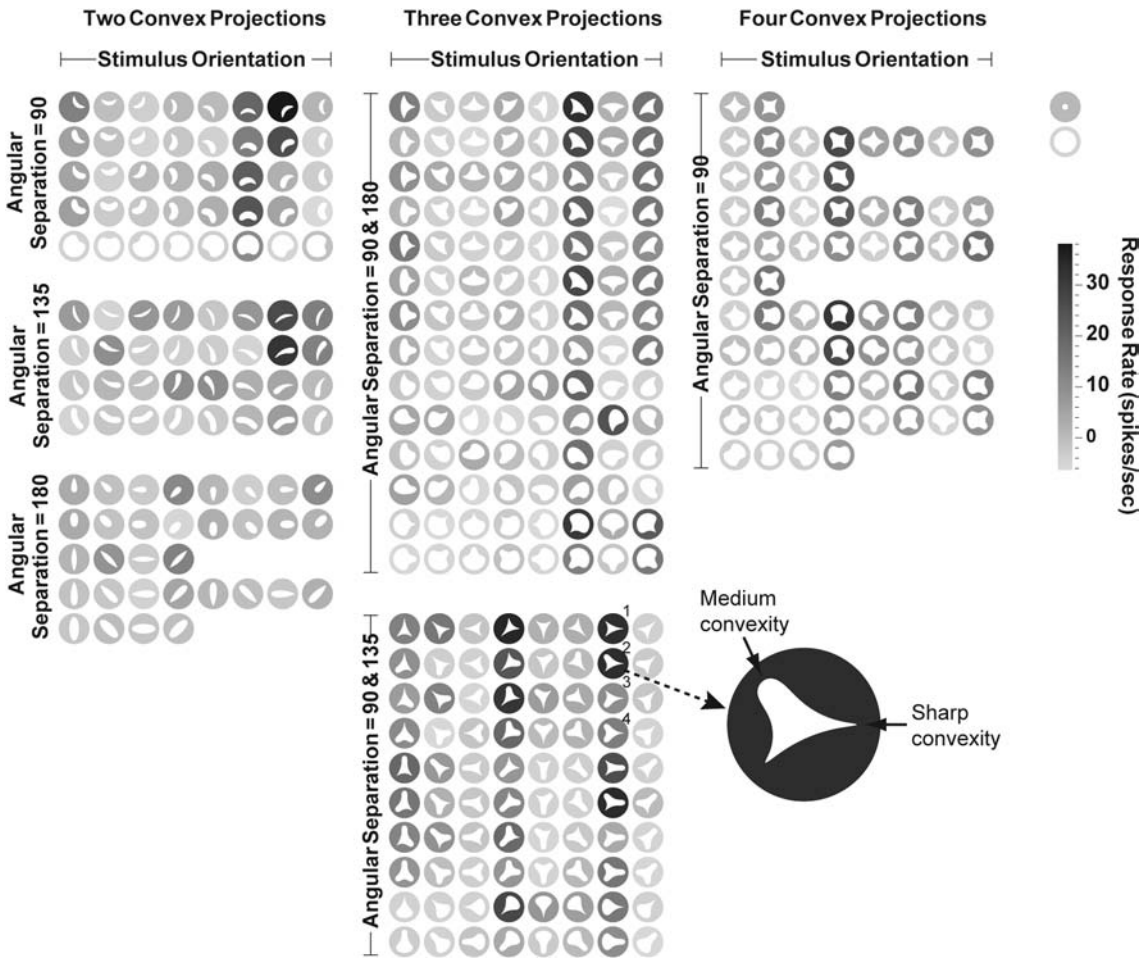

Figure 7.1
Example V4 neuron whose responses are dictated by curvature of the bounding contour at the lower-left corner and bottom of the object. The background grayscale of each icon shows the average response of the cell to the overlaid stimulus object. Each stimulus (shown in white) was presented in the optimal color for the cell, in the center of the receptive field (represented by the surrounding circle which was not part of the visual display). This neuron responds preferentially to shapes with a sharp convexity at the lower-left corner adjoined by a concavity at the bottom of the object. Numbers (1–4) pertain to references in text. Modified from Pasupathy and Connor (2001).

at the bottom (270°) of the object. Stimuli with a medium convexity at 225° elicited a weaker response (e.g., compare responses *1* vs. *3* or *2* vs. *4* in figure 7.1, whose shapes differ only in the 225° position). Stimuli with a broad convexity or a concavity at this position failed to drive the cell. Contour curvature values at other positions along the boundary (from 0° to 180°, right, top and left) varied widely across stimuli evoking strong responses, and they did not modulate the neuron's activity. These results suggest that V4 neurons are tuned for contour curvature, and that curvature values at specific positions along the boundary strongly dictate neural responses to complex shapes.

To rigorously evaluate how object contour characteristics influence responses of V4 neurons, we sought to derive a quantitative relationship between the two. To do this, we needed to describe our stimuli in a parametric "shape space," and we chose curvature and angular position as the defining dimensions. The stimuli in our experiments can be uniquely represented in terms of a continuous curvature × angular position function. (This is not the case for all two-dimensional closed contours. For a general unconstrained two-dimensional closed contour, unique representation would require additional dimensions such as radial position relative to the center of the stimulus and local tangential orientation. Here, since several of these dimensions co-vary, representation in terms of just curvature and angular position is unique). Because of how the stimuli were originally constructed (a smooth curve drawn through 4–8 constant curvature segments), we could simplify the representation of each shape down to 4–8 discrete curvature × angular position values along the contour. This discretized representation was less cumbersome and captured all the information in the continuous curvature function. Therefore, each shape in our set could be uniquely represented using four to eight points in the curvature × angular position space. For example, stimulus 2 (figure 7.1) is represented by six points: one for each of the sharp convex points at 0° and 225°, one for the medium convexity at 135°, and one for each of the intervening concavities. We modeled the neural response as a 2D Gaussian function (product of two 1D Gaussians with no correlation terms) of contour curvature and angular position. A neuron's response to a stimulus was modeled as the maximum of the responses predicted by its component contour segments (cf. Riesenhuber and Poggio, 1999). By this scheme, if a cell were strongly driven by a particular boundary element, the tuning function would predict high responses to all shapes containing that element, independent of other stimulus characteristics. Parameters (Gaussian amplitude, and peak position and standard deviation for each of the two dimensions) were estimated by minimizing the sum of squared errors between predicted and observed responses.

Figure 7.2 shows the best-fitting curvature × position tuning surface for the figure 7.1 example neuron. The horizontal axes represent angular position (0° to 360°) and

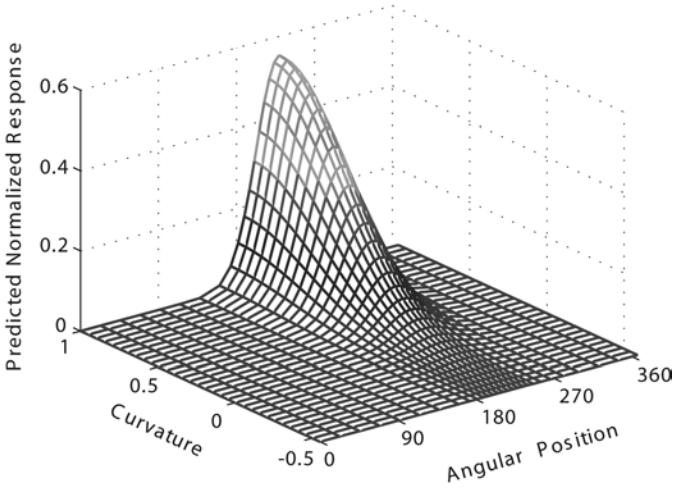

Figure 7.2
2D Gaussian tuning function that best describes responses in figure 7.1. Horizontal axes represent angular position and contour curvature. Vertical axis represents normalized response predicted by the tuning function. This tuning function (mean: contour curvature = 1.0; angular position = 229.6°) predicts strongest responses to sharp convexity at the lower left corner of the object consistent with the response pattern in figure 7.1. Modified from Pasupathy and Connor (2001).

contour curvature (ranging from –0.31 to 1.0; negative, zero, and positive values correspond to concave, flat, and convex curvatures, respectively). The Z axis and surface shading represent the predicted normalized response. The peak of the Gaussian surface (contour curvature = 1.0 and angular position = 229.6°), predicts the strongest responses to stimuli with a sharp convex projection near the lower-left corner (~225°) of the object, consistent with the observed responses for this cell. Tuning along the angular position dimension was narrow (Gaussian standard deviation of 27°) indicating that only a small range of positions of the sharp convexity evoked strong responses. The model accurately predicted responses close to zero for stimuli with broad convexities and concavities at the lower left. Overall, there was an excellent correlation (r = 0.7) between the observed and predicted responses across the full stimulus set.

A large fraction of V4 neurons that we studied showed similar systematic shape responses that could be well described by a 2D Gaussian tuning function. Figure 7.3A shows a distribution of the goodness of fit values (correlation between observed and predicted responses) for the best fitting 2D Gaussians for the entire population. The predictions of the 2D model are significantly improved if the curvatures of two adjoining segments are also included in a 4D Gaussian model of curvature × angular position (Pasupathy and Connor, 2001). These results strongly suggest that many

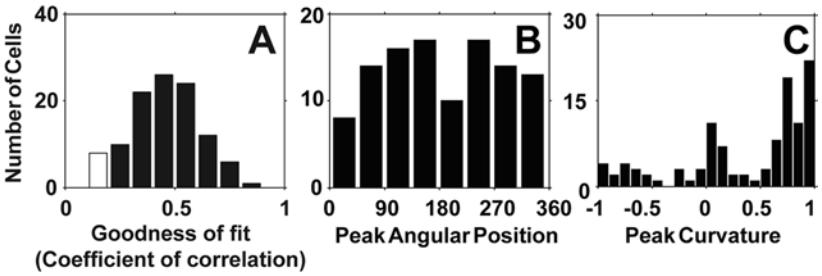

Figure 7.3
V4 population results: goodness of fit and tuning parameters. (A) Goodness of fit for the 2D Gaussian model. Distribution of coefficient of correlation values between observed and predicted responses for the 109 V4 neurons. (B, C) Distributions of the mean parameters for the angular position (B) and curvature (C) dimensions of the 2D Gaussian model.
Modified from Pasupathy and Connor (2001).

V4 neurons encode complex shapes in terms of the contour curvature at specific locations along the bounding contour of the object. Extensive analyses also revealed that, for a large majority of V4 neurons, responses could not be explained in terms of stimulus edge orientation, contrast polarity, or the orientation of the axis of elongation and aspect ratio of the stimulus (Pasupathy and Connor, 2001).

Another important requirement for a basis dimension is that neural tuning peaks span the entire range of possible values of the dimension, at every location in the visual field. The distributions of tuning peaks across our studied population of V4 neurons suggest that this criterion may also be met for the curvature and angular position dimensions (figure 7.3B,C; tuning peaks collapsed across receptive field locations). The tiling of tuning peaks along the angular position dimension (figure 7.3B) is not significantly different from uniform. While there is a strong bias toward representation of sharp convex curvatures—perhaps reflecting the enhanced perceptual salience of convexities relative to convexities and straight edges (Kanizsa and Gerbino, 1976; Subirana-Vilanova and Richards, 1996)—all tested curvature values are represented. However, since most of the recorded neurons had receptive fields in the 3°–6° range, we cannot assess whether the observations made here hold at other visual field locations. Second, since our stimuli effectively sampled only five points along the curvature axis (three convex and two concave), further experiments with denser sampling along the curvature axis are required to draw strong conclusions about tiling densities. Nevertheless, given the observed distribution of tuning peak locations and breadth of tuning (Pasupathy and Connor, 2001), it seems likely that V4 neurons span the curvature and angular position dimensions.

Ensemble Representation of 2D Contours in Area V4

We have demonstrated that single V4 neurons encode complex shapes in a piecewise fashion—some neurons encode information about a sharp convexity off to the left, others about a concavity at the bottom, etc. An obvious next step is to ask how such piecewise information is distributed across the V4 population: do V4 neurons, as a population, provide a complete and accurate representation of the entire shape? If so, this would be another strong indicator that piecewise contour curvature and angular position serve as a basis dimensions in V4.

We assessed the completeness and accuracy of shape information across the V4 population by deriving a population code for each stimulus object from the responses of all studied single neurons and their tuning functions, using a linear basis function decoding method. This method estimates a population code for a given stimulus as a weighted sum of basis functions, where the bases are the set of single-neuron tuning functions (in this case, 2D Gaussian functions on the curvature × angular position domain, as in figure 7.2), and the weights are each corresponding neuron's response to the stimulus in question. The method assumes that the neurons that make up the population are independent processes, but the tuning functions can be of any arbitrary shape. The well-known population vector decoding method (Georgopoulos, Schwartz, and Kettner, 1986) is a special case of basis function decoding (Zhang et al., 1998).

Mathematically, if $f_i(\theta,c)$ represents the ith neuron's tuning function in the angular position (θ) × curvature (c) space and r_{ij} represents its response to a specific object j, then the population reconstruction of that object, $p_j(\theta,c)$, is given by:

$$p_j(\theta,c) = \sum_i r_{ij} f_i(\theta,c) \tag{7.1}$$

p_j, is a probabilistic surface in the 2D angular position × curvature space and its local maxima represent the locations of the most probable features of the given object in this space. Thus, p_j, represents the population code for a given object and its peaks provide a decoded estimate of the object's shape.

Figure 7.4 (plate 4)shows the resulting population code for an example stimulus object (figure 7.4A) as a pseudocolor surface (figure 7.4B)—colors closer to red reflect stronger evidence for the presence of shape features with the corresponding curvature/angular position combinations within the object. The population surface contains peaks (red) corresponding to the major boundary features of this object: the sharp convexity at 90°, medium convexities at 0° and 180°, broad convexity at 270°, and concavities at 45° and 135°.

We obtained similar results for all shapes in our stimulus set—the V4 population code had peaks associated with all of their major boundary features (Pasupathy and

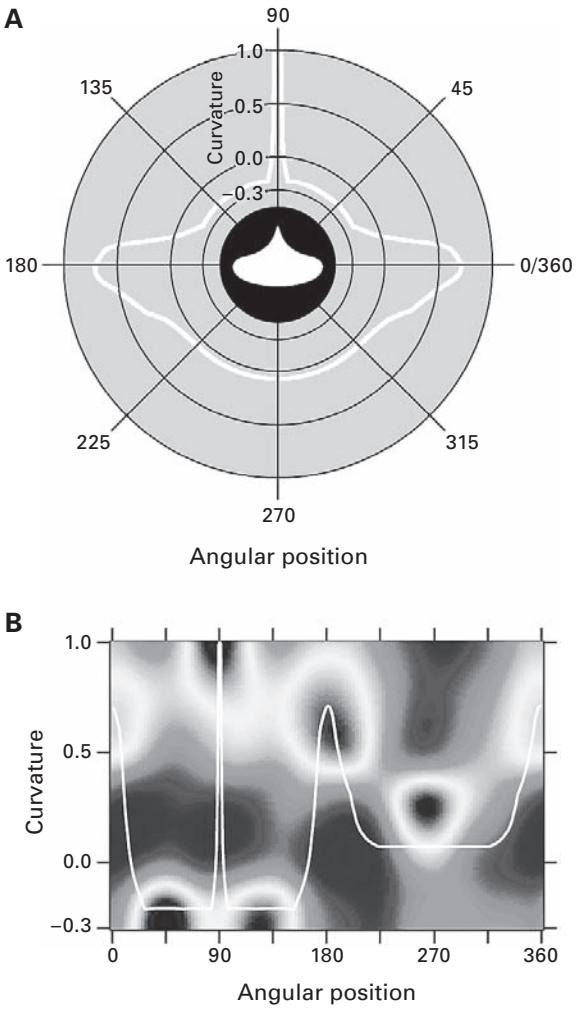

Figure 7.4 (plate 4)
Reconstruction of V4 population code for a given stimulus object. (A) The stimulus object is shown in the center. The surrounding white line plots boundary curvature (radial dimension) as a function of angular position (angular dimension). (B) Estimated population code across the curvature × angular position domain (colored surface); true curvature function superimposed (white line) for comparison. X axis represents angular position; Y axis represents curvature. Color scale runs from 0.0 (blue) to 1.0 (red). Peaks in the population code (red) correspond quite well with the peaks and troughs in the curvature function. Reproduced from Pasupathy and Connor, 2002.

Connor, 2002). This was confirmed by quantitative comparisons showing that the peaks in the population code matched up remarkably well to the extrema in the corresponding true curvature × position function (compare to the white curve in figure 7.4B). Thus the V4 population code provides a complete and accurate representation of 2D object shapes in terms of contour curvature and angular position.

Shape Coding in Posterior Inferotemporal Cortex

The majority of area V4 outputs project to the next stage in the ventral pathway, posterior inferotemporal cortex (PIT), which in turn projects to the highest level of form processing, anterior IT. Though PIT therefore sits at a key point in the transformation from intermediate-level to high-level shape processing, it has received scant experimental attention (Kobatake and Tanaka,1994). We therefore adapted our experimental paradigm to examine the shape selectivity of PIT neurons (Brincat and Connor, 2004). Neuronal activity was measured in response to a large set of parametrically defined shapes that expanded on the stimulus set used in our V4 experiments. Though this set sampled exact contour-curvature magnitudes more sparsely than the previous one, it featured a much larger array of more complex contour configurations (see figure 7.5A,B for a subset of the full stimulus set), reflecting our a priori expectation that PIT cells would exhibit selectivity for more complex object structure.

The example neuron in figure 7.5A shows exactly this expected pattern—it responds robustly to shapes containing a combination of concave contour segments pointing to the lower left and lower right (figure 7.5A, bottom). Objects containing only one of these contour segments, however, elicited virtually no response (figure 7.5A, top). This neuron is therefore highly selective for a specific configuration of contour segments, which together define a curved "v"-like shape. Many other PIT neurons, such as the one in figure 7.5B, were not such an obvious match with our expectations. This cell also responded optimally to shapes with a similar combination of concave contour segments to the previous example. However, unlike the first neuron, shapes containing only one component of the optimal combination also evoked moderately strong responses. Thus, while the first example neuron approximates a "logical AND" computation, responding only when an object contains a specific *combination* of contour segments, the second cell exhibits a much less specific, more distributed, response profile around the preferred contour segment combination.

We found that a single class of tuning function models could parsimoniously explain both types of observed PIT neural response patterns. The models consisted of one or more excitatory or inhibitory Gaussian tuning peaks on a 4D contour shape (curvature and orientation) × object-relative position (x, y relative to the

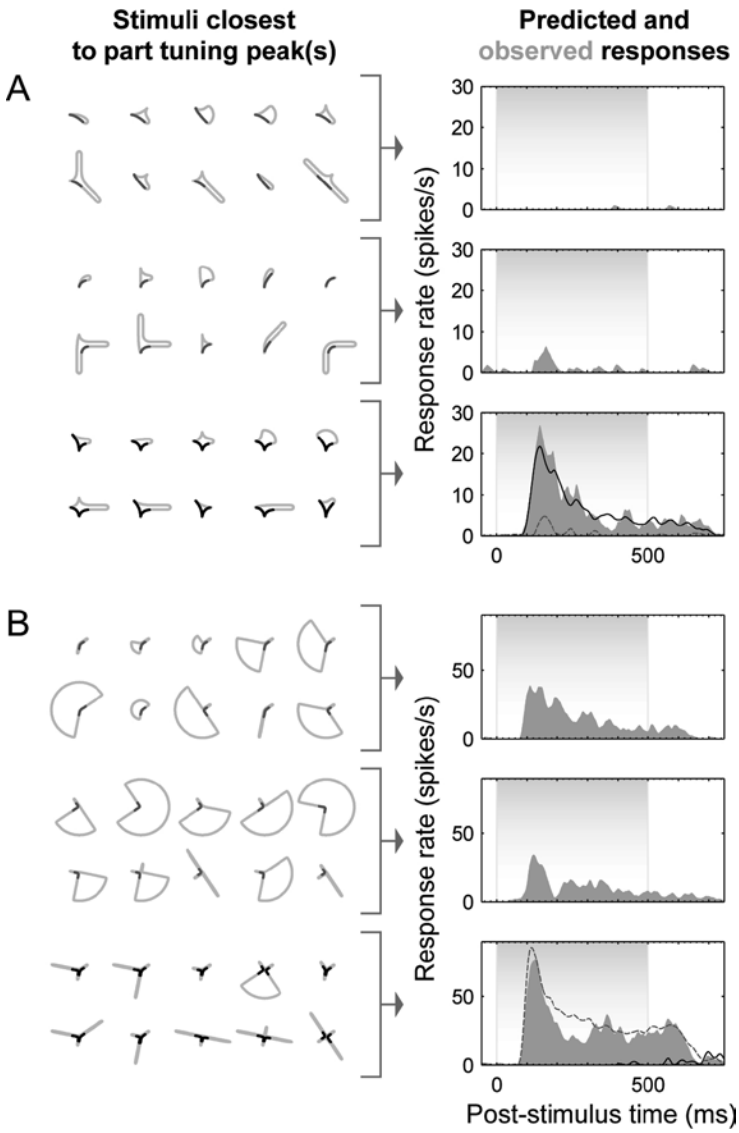

Figure 7.5
Example neurons illustrating PIT linear and nonlinear response patterns. (A) Neuron with linear tuning for two types of contour parts. Left: Subset of shape stimuli used to study this cell that contain contour fragments (dark gray) most closely matching its shape-space tuning peaks. Right: Averages across example stimuli of observed neural responses (gray histogram), and responses predicted by linear (gray dotted line) and nonlinear (black solid line) components of fitted models. Light gray shading denotes stimulus presentation period. (B) Neuron with nonlinear tuning for a specific multipart configuration (same conventions as in A). Modified from Brincat and Connor (2006).

object center) domain. (Because of the added complexity of this stimulus set, two object-relative position dimensions, as well a local contour orientation dimension, were necessary to fully describe object shape). Each distinct peak in this space can be thought of as a subunit of a complex, high-dimensional PIT receptive field, and may correspond to distinct V4 inputs. The full model (equation (7.2)) contained terms that combined the responses (R_s) of multiple subunits either through simple linear summation (first term) or by a multiplicative interaction that captured non-linear, logical-AND-like selectivity for specific contour combinations (second term).

$$\text{Response}_{\text{overall}} = \left[\sum_{s}^{\text{all subunits}} (w_s R_s) + w_{NL} \prod_{s}^{\text{excitatory subunits}} (R_s) + b_0 \right]^{+} \tag{7.2}$$

Least-squares model fits to the responses of each PIT neuron across ~1,000 stimuli yielded optimal values for the baseline response (b_0), for the mean (center) and standard deviation (tuning width) of Gaussian subunits, and for weights on the linear (w_s's) and nonlinear (w_{NL}) terms, whose relative magnitude reflects a cell's degree of contour-combination nonlinearity. The optimal number of subunits was determined through a model selection procedure that required the n^{th} model subunit to explain a criterion additional fraction of response variance over the optimal model with $n - 1$ subunits (this procedure was considerably more conservative than other common model selection criteria, such as the partial-F test, AIC, or BIC).

As in V4, most of the 109 PIT neurons we studied displayed graded, systematic tuning for contour shape and object-relative position, which was well characterized by our tuning function models (average cross-validated correlation between predicted and observed responses of 0.7). Across the sampled population, PIT tuning functions collectively covered essentially all contour curvatures, orientations, and object-relative positions (figure 7.6). These results suggest that PIT neurons, similar to V4, encode the shape of complex objects in terms of basis dimensions related to local contour shape (curvature and orientation) and object-relative position.

However, PIT appeared to be selective for more complex object structure on average than V4—the majority of cells were tuned for ~2–4 distinct contour segments along object boundaries, and many were also significantly inhibited by specific contour segments. With respect to their degree of contour-combination nonlinearity, our sample of PIT neurons split into two fairly distinct subpopulations. One group of neurons, like the one in figure 7.5B, showed almost purely linear summation across contour segments ($w_{NL} \approx 0$ in equation 7.2). These cells act as complex, multimodal linear filters on the contour shape × object-relative position domain. As such, they are sensitive to specific patterns of contour segments within objects, but not exclusively so—moderate responses could correspond to either an off-optimal pattern of contours or to only a single near-optimal contour segment. Another

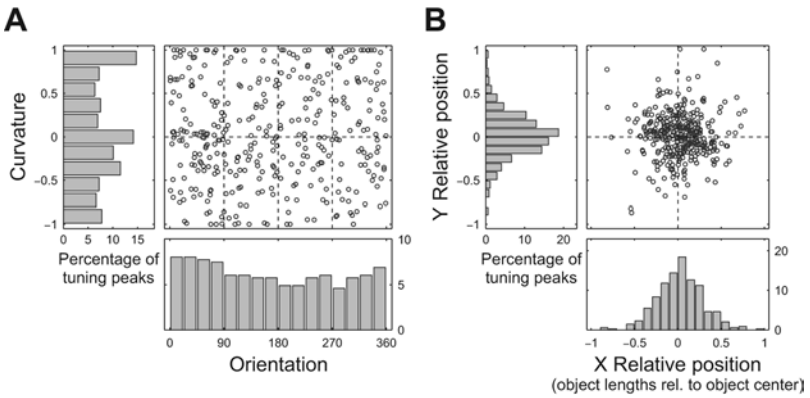

Figure 7.6
Distributions of tuning peaks indicate that all contour curvature, orientation, and object-relative position values are represented in PIT. (A) Distribution of all PIT tuning peaks (all subunits of all neuronal tuning functions) on the contour shape dimensions (curvature × orientation), and marginal histograms. (B) Distribution of all PIT tuning peaks on the object-relative position dimensions (X × Y). Peak positions are normalized to the length of the longest object in the stimulus set used to study each neuron, and are plotted in units of object lengths relative to the object center. A symmetric object would therefore maximally extend from –0.5 to +0.5 on these axes. Modified from Brincat (2005).

group, exemplified by figure 7.5A, showed varying degrees of nonlinear AND-like selectivity for specific contour combinations ($w_{NL} > w_s$ in equation 7.2). Unlike the linear subpopulation, these neurons respond almost exclusively to specific configurations of contour segments, and can be thought of as either low-order nonlinear filters in shape × relative position space, or equivalently, as unimodal filters in a high-dimensional "contour-configuration space." This type of nonlinear summation is essential for generating complex shape selectivity in a hierarchical system, since a series of linear transforms can be reduced to a single linear transform. As elaborated later herein, linear cells may simply act as an intermediate computational stage in the biological instantiation of this nonlinear transform. In summary, individual PIT neurons represent object shape in terms of configurations of specific contour segments at multiple positions along the object boundary, either via nonlinear AND-like combination or via linear summation across contour segments.

Population Coding of 2D Contours in PIT

Though we studied neurons sequentially across recording sessions, as in V4, we can infer the basic structure of the PIT population code for contour shape based on single-neuron tuning properties. Our results show that PIT cells represent contour-segment configurations within whole objects, and that their tuning functions

cover essentially all contour curvatures, orientations, and object-relative positions. Whole-object shape must therefore be represented by an ensemble of neurons coding for overlapping combinations of contour segments. Each contour-segment-combination within an object would be represented in a graded fashion by a subpopulation of neurons with nearby configural tuning peaks. These neurons effectively comprise a local Gaussian-like bump of activation in a high-dimensional contour-configuration space. The peak location of this activation bump codes for the combined shape and relative positions of the actual stimulating configuration, while its width reflects the tuning widths of the contributing single units. An entire object will be represented by several such local population activity bumps, each coding for a distinct, though overlapping, combination of contour segments within the object. Such a population code would have several computational advantages—it could flexibly encode a large number of objects with a limited vocabulary of features (cf. Marr and Nishihara, 1978; Biederman, 1987; Edelman and Intrator, 2003), but since the features are fairly complex, objects could be represented with a relatively sparse code that is efficient for recognition and memory storage (Rolls and Treves, 1990).

If this characterization of the PIT population code is correct, we should be able to accurately decode the shape of an object from the PIT population responses evoked by it, as we did in V4. This task is considerably more difficult here, however—the stimulus shapes and neural tuning functions are more complex, and the stimuli used to study different neurons were not identical (the global shapes were the same, but the sampling of exact convex and concave curvature magnitudes was optimized for each neuron). Nevertheless, we attempted to decode population responses with a basis function method similar to that used in our V4 work. To reconstruct the population response to a given object, we performed a weighted summation of the tuning functions of all recorded PIT neurons, with the weight on each neuron's tuning function ($f_i(\theta,c,x,y)$, in equation 7.3) given by its response (r_{ij}) to the object. Because of the strongly combinatorial nature of most PIT cells' responses, each neuron's full, multipeaked tuning function was used for this decoding analysis, with the nonlinear terms split evenly among the n constituent tuning peaks to simplify the computations (equation 7.3, top, second term), and the resulting full function normalized to a unit integral to make it a proper probability distribution (equation 7.3, bottom).

$$f_i(\theta,c,x,y) = \sum_{s}^{n} w_s f_{is}(\theta,c,x,y) + \sum_{s}^{n} (w_{NL}/n) f_{is}(\theta,c,x,y)$$
$$p_j(\theta,c,x,y) = \sum_{i}^{\text{all cells}} \left[r_{ij} f_i(\theta,c,x,y) / \int |f_i(\theta,c,x,y)| \right] \quad (7.3)$$

The result of this decoding analysis was a complex, multipeaked function $p_j(\theta,c,x,y)$ on the 4D contour shape (orientation θ, curvature c) × object-relative position (x,y) domain that represents the PIT population code for a given object j (figure 7.7), in terms of the probability of object features at locations within this space. A peak-finding algorithm was used to extract local maxima in this function (figure 7.7, black squares), representing the decoded estimates of the most probable shapes and positions of this object's constituent contour segments (figure 7.7, white circles). As was the case for many stimuli, these estimates were reasonably close to the actual values. Although the spread of the local peaks in these functions suggests that part shape and (especially) position are coarsely coded in PIT neural responses, they demonstrate that 2D object structure can be accurately recovered from the PIT population code.

Transformation between V4 and IT Representations

Our work has contributed to at least a first-pass understanding of how single neurons and neural populations encode object shape within two intermediate-level ventral pathway areas. A critical next question is how coding transformations across successive processing stages are actually generated by networks of neurons. Theoretical work suggests two broad classes of potential mechanisms. One possibility is that more complex selectivity is generated at each hierarchical level by selective feedforward convergence of simpler inputs from the previous stage. This scheme was originally proposed to explain the generation of V1 orientation selectivity from thalamic afferents (Hubel and Wiesel, 1962), and was later suggested as a general mechanism for all ventral pathway shape transformations (Riesenhuber and Poggio, 1999). Alternatively, feedforward processing may provide only a coarse, weak representation that is amplified and refined by recurrent processing within each ventral pathway area (Douglas et al., 1995; Salinas and Abbott, 1996; Chance, Nelson, and Abbott, 1999). Parsimony and the speed of object category perception (Thorpe, Fize, and Marlot, 1996) have tended to favor feedforward models, and few results have been produced that they have not been able to accommodate (Sugase et al., 1999).

To distinguish between these alternatives, we examined the time course of shape coding in PIT neurons. We found that, in the PIT population, the transformation from linear contour-segment summation to an explicit nonlinear code for contour configurations develops gradually over a duration of ~60 ms following stimulus onset. This is partially due to a ~20 ms lag in response onset between neurons exhibiting predominately linear and nonlinear responses, which might be accommodated in an elaborated feedforward model. However, there are also many individual PIT neurons that initially show linear summation but gradually evolve

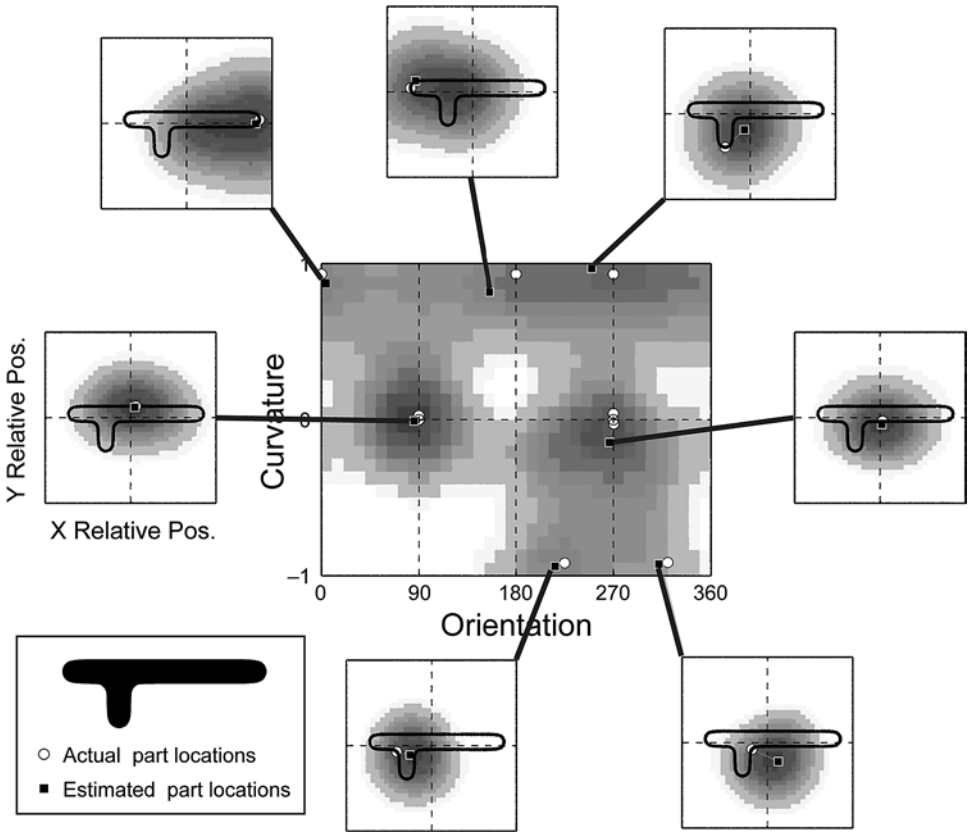

Figure 7.7
Reconstruction of PIT population code for a given stimulus shape. Each pseudocolor plot is 2-dimensional projection of a 4-dimensional (orientation × curvature × x,y position relative to object center) function representing the activation of the PIT neural population in response to the shape depicted at the lower left. Center: orientation × curvature domain activation averaged across all object-relative positions. Periphery: object-relative position domain activations taken at locations of local peaks in reconstruction function. White circles indicate actual locations of stimulus contour fragments within 4D space (curvature values are the median values used across the population), and black squares indicate population reconstruction estimates of these points based on the locations of local peaks within the 4D reconstruction function. Note that these reconstruction peak points closely match the actual shape space locations of the stimulating object's parts.

nonlinear selectivity over tens of milliseconds (e.g., responding at short latency to either of a pair of optimal contour segments, but with a sustained response only to shapes containing their combination, as in figure 7.8). It is not obvious how purely feedforward models could account for these results, though such complex dynamics arise naturally in recurrent networks (Pugh et al., 2000).

To illustrate how such a process might work, we built a simple recurrent network model of the V4 to PIT transformation (figure 7.9A) with an architecture based on previous models of cortical processing (Salinas and Abbott, 1996; Chance, Nelson, and Abbott, 1999). Each model PIT unit received both feedforward inputs (modeling V4 afferents), which conferred additive selectivity for two Gaussian ranges of contour segment shapes, and recurrent inputs from other PIT units. The recurrent connection weights were structured such that units with similar contour-combination tuning were mutually facilitative, while those with dissimilar tuning were mutually suppressive. Recurrent processing therefore effectively made comparisons across populations of active neurons to infer the presence of multicontour-segment configurations. The relative weights of feedforward and recurrent inputs were varied continuously across units independently of their shape-tuning properties (figure

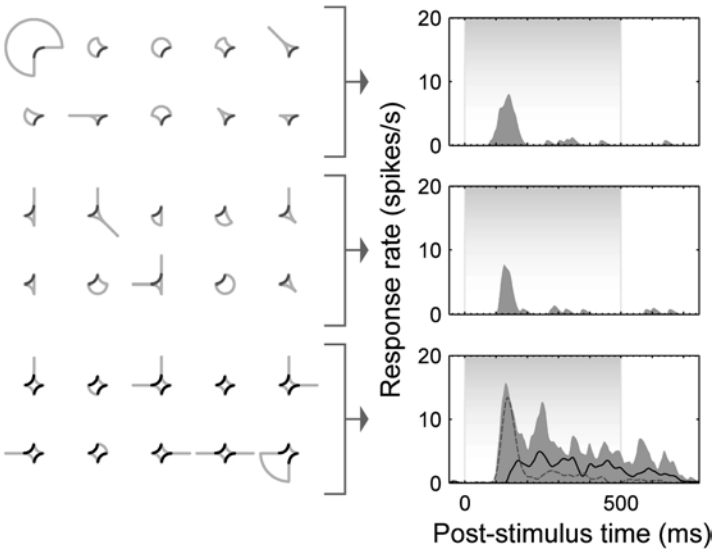

Figure 7.8
Example neuron illustrating PIT linear to nonlinear transitioning response pattern. Same format as figure 7.1. At short latencies (< 180 ms), this neuron's response to multipart stimuli (bottom; gray histogram: observed response, gray dashed line: linear component) can be predicted from the sum of its responses to each optimal part in isolation (top). At later time points, the cell responds only to the full optimal part configuration. Modified from Brincat and Connor (2006).

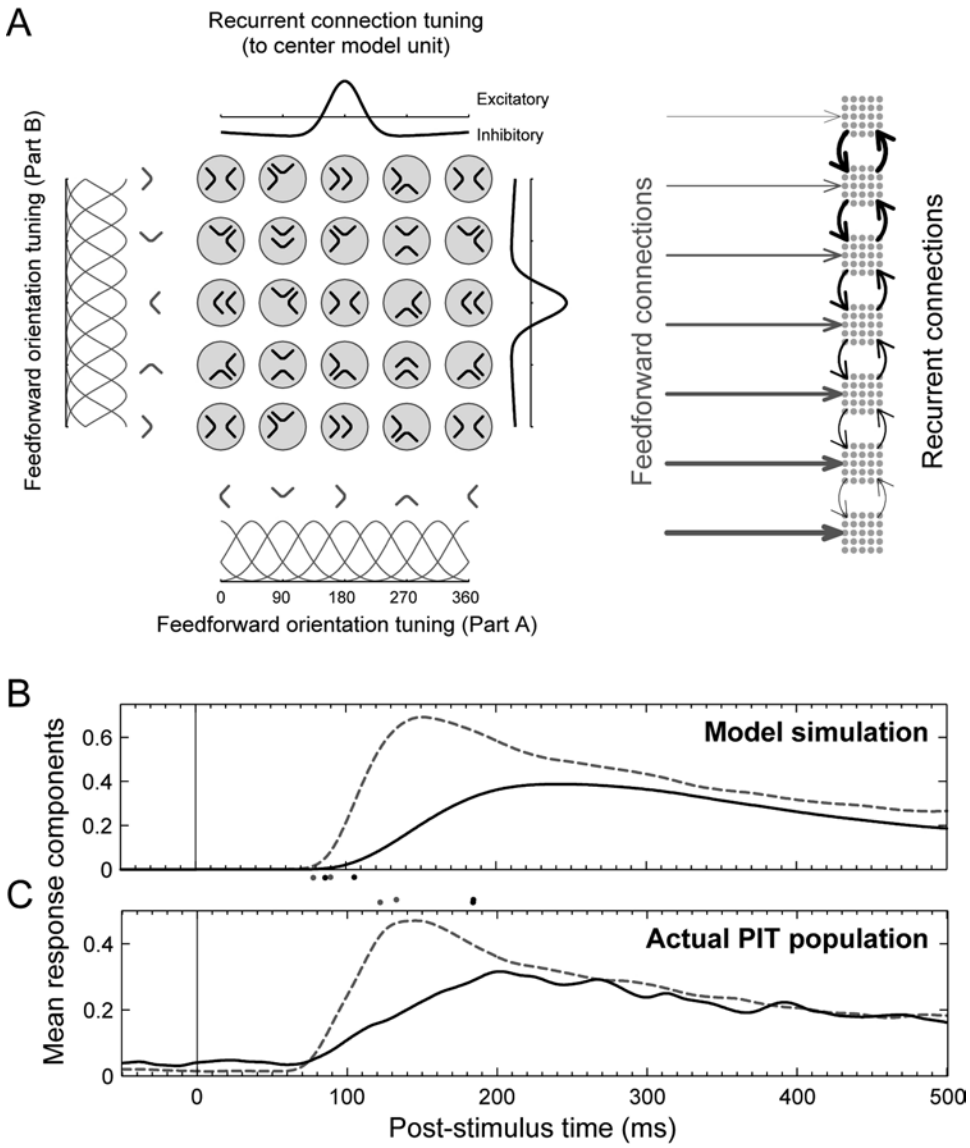

Figure 7.9
A recurrent network model can explain observed PIT response dynamics. (A) Illustration of network architecture. Left: Each model unit receives Gaussian-tuned feedforward inputs (light gray curves at left and bottom) for two ranges of object parts (in this model, concave contour fragments differing in orientation). The population of model units spans all combinations of tuning for two parts (black illustrations). The black curves (top and right) represent strength of recurrent inputs to a unit at the center of the model tuning grid. The difference-of-Gaussians or "Mexican-hat" profile of recurrent connection weights produces mutual excitation between active units with similar tuning and mutual suppression between units with dissimilar tuning. Right: The relative strengths (arrow thickness) of feedforward (gray) and recurrent (black) connections vary continuously across the population of model units. (B, C) Average temporal profiles of linear (gray dotted line) and nonlinear (black solid line) response components across the populations of model-simulated (B) and actual (C) PIT units. The recurrent network model accurately captures the delayed, gradual onset of the nonlinear signal observed in the PIT population. Modified from Brincat and Connor (2006).

7.9A, right). Model unit responses were probed by activating feedforward inputs representing one or both of their optimal segments, with the temporal profile of linear responses measured in the actual PIT population. Unsurprisingly, units with predominately feedforward inputs retained this rapid-onset linear representation in their responses. However, units with strong recurrent inputs developed a delayed nonlinear contour-configuration selectivity (similar to the actual neuron in figure 7.5A), and units with balanced feedforward and recurrent inputs exhibited a clear linear to nonlinear temporal shift (similar to the figure 7.8 cell). The time course of nonlinear selectivity averaged across the entire set of model units (figure 7.9B) closely matched the gradual temporal profile observed in the actual PIT population (figure 7.9C).

These simulations demonstrate that our data are at least consistent with a simple recurrent network mechanism for generating complex shape selectivity, though our results alone cannot be interpreted as conclusive evidence in favor of this particular model or against other recurrent or feedback architectures. More fundamentally, our data indicate that the transformation to an explicit PIT population code for complex contour configurations does involve a gradual, dynamic network process. This poses a serious challenge to purely feedforward models of ventral pathway transformations, but it is consistent with the increased processing time required for configural perception (Wolfe and Bennett, 1997; Arguin and Saumier, 2000) and with the gradual transformations to more complex representations observed in other brain areas (Pack, Berezovskii, and Born, 2001; Smith, Majaj, and Movshon, 2005).

Conclusion

Our results in V4 and PIT provide the first quantitative evidence for contour curvature as a basis dimension in the intermediate processing stages of the ventral shape–processing pathway. Single neurons encode complex shapes in terms of the curvature at a specific boundary location in V4, and in terms of configurations of contour curvatures at multiple locations in PIT. In both areas, neural tuning spans the entire range of possible contour curvature values and object boundary locations, making possible a complete and accurate population representation of a variety of complex shapes. This was confirmed by decoding a reasonably accurate estimate of object shape from the pattern of neural population responses in each area. Finally, we have shown evidence that the transformation between the V4 and PIT population codes is a gradual, dynamic process that may involve recurrent processing, in addition to classical feedforward circuitry.

These findings provide a preliminary understanding of object shape coding within intermediate ventral pathway stages, and will serve as a starting point for future

investigations into the representation of more complex shape properties, multiple overlapping objects, and, eventually, entire complex natural scenes. Ultimately, a central goal of visual neuroscience is to understand the neural codes for object shape—and the transformations between them—through the entire object vision system. We expect that the methods and concepts we have described in this chapter will also serve as a basis for extending our current understanding to the transformation between low-level (V1/V2) edge orientation signals and intermediate-level curvature selectivity, and to the transformations that take place in the highest processing stages in anterior IT.

Acknowledgments

Ideas and experiments presented in this chapter were developed and performed at the Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD in collaboration with Charles E. Connor.

References

Andrews DP, Butcher AK, Buckley BR. 1973. Acuities for spatial arrangement in line figures: human and ideal observers compared. *Vision Res* 13: 599–620.

Anzai A, Peng X, Van Essen DC. 2007. Neurons in monkey visual area V2 encode combinations of orientations. *Nat Neurosci* 10: 1313–1321.

Arguin M, Saumier D. 2000. Conjunction and linear non-separability effects in visual shape encoding. *Vision Res* 40: 3099–3115.

Attneave F. 1954. Some informational aspects of visual perception. *Psychol Rev* 61: 183–193.

Baizer JS, Robinson DL, Dow BM. 1977. Visual responses of area 18 neurons in awake, behaving monkey. *J Neurophysiol* 40: 1024–1037.

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94: 115–147.

Brincat SL. 2005. Neural coding of object contour shape in primate posterior inferotemporal cortex. Baltimore, MD: PhD Dissertation, Johns Hopkins University.

Brincat SL, Connor CE. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7: 880–886.

Brincat SL, Connor CE. 2006. Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49: 17–24.

Burkhalter A, Van Essen DC. 1986. Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J Neurosci* 6: 2327–2351.

Chance FS, Nelson SB, Abbott LF. 1999. Complex cells as cortically amplified simple cells. *Nat Neurosci* 2: 277–282.

Chen S, Levi DM. 1996. Angle judgement: is the whole the sum of its parts? *Vision Res* 36: 1721–1735.

Desimone R, Albright TD, Gross CG, Bruce C. 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062.

Dickinson SJ, Pentland AP, Rosenfield A. 1992. From volumes to views: An approach to 3-D object recognition *CVGIP:Image Understanding* 55: 130–154.

Douglas RJ, Koch C, Mahowald M, Martin KA, Suarez HH. 1995. Recurrent excitation in neocortical circuits. *Science* 269: 981–985.

Edelman S, Intrator N. 2003. Towards structural systematicity in distributed, statically bound visual representations. *Cogn Sci* 27: 73–109.

Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47.

Gallant JL, Braun J, Van Essen DC. 1993. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259: 100–103.

Georgopoulos AP, Schwartz AB, Kettner RE. 1986. Neuronal population coding of movement direction. *Science* 233: 1416–1419.

Gross CG, Rocha-Miranda CE, Bender DB. 1972. Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol* 35: 96–111.

Hegde J, Van Essen DC. 2000. Selectivity for complex shapes in primate visual area V2. *J Neurosci* 20: RC61.

Heeley DW, Buchanan-Smith HM. 1996. Mechanisms specialized for the perception of image geometry. *Vision Res* 36: 3607–3627.

Hubel DH, Livingstone MS. 1987. Segregation of form, color, and stereopsis in primate area 18. *J Neurosci* 7: 3378–3415.

Hubel DH, Wiesel TN. 1959. Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148: 574–591.

Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160: 106–154.

Hubel DH, Wiesel TN. 1968. Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195: 215–243.

Ito M, Komatsu H. 2004. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J Neurosci* 24: 3313–3324.

Kanizsa G, Gerbino W. (1976) Convexity and symmetry in figure-ground organization. In: *Art and Artefacts*, edited by M. Henle. New York: Springer, p. 25–32.

Kobatake E, Tanaka K. 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867.

Marr D, Nishihara HK. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond B Biol Sci* 200: 269–294.

Milner PM. 1974. A model for visual shape recognition. *Psychol Rev* 81: 521–535.

Pack CC, Berezovskii VK, Born RT. 2001. Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature* 414: 905–908.

Pasupathy A, Connor CE. 1999. Responses to contour features in macaque area V4. *J Neurophysiol* 82: 2490–2502.

Pasupathy A, Connor CE. 2001. Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86: 2505–2519.

Pasupathy A, Connor CE. 2002. Population coding of shape in area V4. *Nat Neurosci* 5: 1332–1338.

Pentland A. 1989. Shape information from shading: a theory about human perception. *Spat Vis* 4: 165–182.

Perrett DI, Rolls ET, Caan W. 1982. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47: 329–342.

Poggio T, Edelman S. 1990. A network that learns to recognize three-dimensional objects. *Nature* 343: 263–266.

Pugh MC, Ringach DL, Shapley R, Shelley MJ. 2000. Computational modeling of orientation tuning dynamics in monkey primary visual cortex. *J Comput Neurosci* 8: 143–159.

Regan D, Gray R, Hamstra SJ. 1996. Evidence for a neural mechanism that encodes angles. *Vision Res* 36: 323–330.

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019–1025.

Rolls ET, Treves A. 1990. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1: 407–421.

Rust NC, Movshon JA. 2005. In praise of artifice. *Nat Neurosci* 8: 1647–1650.

Salinas E, Abbott LF. 1996. A model of multiplicative neural responses in parietal cortex. *Proc Natl Acad Sci USA* 93: 11956–11961.

Smith MA, Majaj NJ, Movshon JA. 2005. Dynamics of motion signaling by neurons in macaque area MT. *Nat Neurosci* 8: 220–228.

Subirana-Vilanova JB, Richards W. 1996. Attentional frames, frame curves and figural boundaries: the inside/outside dilemma. *Vision Res* 36: 1493–1501.

Sugase Y, Yamane S, Ueno S, Kawano K. 1999. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400: 869–873.

Tanaka K, Saito H, Fukada Y, Moriya M. 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66: 170–189.

Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature* 381: 520–522.

Tolias AS, Ecker AS, Siapas AG, Hoenselaar A, Keliris GA, Logothetis NK. 2007. Recording chronically from the same neurons in awake, behaving primates. *J Neurophysiol* 98: 3780–3790.

Treisman A, Gormican S. 1988. Feature analysis in early vision: evidence from search asymmetries. *Psychol Rev* 95: 15–48.

Ullman S. 1989. Aligning pictorial descriptions: an approach to object recognition. *Cognition* 32: 193–254.

Ungerleider L, Mishkin M. 1982. Two cortical visual systems. In *Analysis of visual behavior*, ed. DJ Ingle, MA Goodale, RJW Mansfield, pp. 549–586. Cambridge, MA: MIT Press.

von der Heydt R, Peterhans E. 1989. Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J Neurosci* 9: 1731–1748.

Wilson HR, Wilkinson F, Asaad W. 1997. Concentric orientation summation in human form vision. *Vision Res* 37: 2325–2330.

Wolfe JM, Bennett SC. 1997. Preattentive object files: shapeless bundles of basic features. *Vision Res* 37: 25–43.

Wolfe JM, Yee A, Friedman-Hill SR. 1992. Curvature is a basic feature for visual search tasks. *Perception* 21: 465–480.

Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. 2008. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11: 1352–1360.

Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ. 1998. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J Neurophysiol* 79: 1017–1044.

Q