Dynamic Shape Synthesis in Posterior Inferotemporal Cortex

Report

Scott L. Brincat[1,2] and Charles E. Connor[1,*]
[1] Zanvyl Krieger Mind/Brain Institute
and Department of Neuroscience
Johns Hopkins University
Baltimore, Maryland 21218

Summary

How does the brain synthesize low-level neural signals for simple shape parts into coherent representations of complete objects? Here, we present evidence for a dynamic process of object part integration in macaque posterior inferotemporal cortex (IT). Immediately after stimulus onset, neural responses carried information about individual object parts (simple contour fragments) only. Subsequently, information about specific multipart configurations emerged, building gradually over the course of ~60 ms, producing a sparser and more explicit representation of object shape. We show that this gradual transformation can be explained by a recurrent network process that effectively compares parts signals across neurons to generate inferences about multipart shape configurations.

Introduction

In primate visual cortex, object shape information is processed in the ventral pathway, a hierarchical network of areas in the occipital and temporal lobes (Ungerleider and Mishkin, 1982). At earlier processing stages in the ventral pathway (V1, V2, and V4), objects are represented in terms of their constituent parts. Individual neurons signal simple properties of local edge fragments (e.g., position, orientation, curvature) (Gallant et al., 1993; Hubel and Wiesel, 1962; Ito and Komatsu, 2004; Pasupathy and Connor, 1999), and objects are encoded by highly distributed patterns of neural activity (Pasupathy and Connor, 2002). In contrast, at the final processing stages in anterior IT, individual neurons show nonlinear selectivity for global object shape, and objects are encoded by a sparser pattern of active neurons (Baker et al., 2002; Freedman et al., 2003; Fujita et al., 1992; Gross et al., 1972; Logothetis et al., 1995; Tsunoda et al., 2001; Young and Yamane, 1992). This kind of sparse global shape coding increases metabolic efficiency, enhances memory storage potential, and simplifies decoding for purposes of object recognition (Barlow, 2001; Fiete et al., 2004; Rolls and Treves, 1990). However, the neural transformation from distributed parts coding in V1, V2, and V4 to sparse global shape coding in anterior IT is not yet understood at a mechanistic level.

Theoretical models suggest two basic alternative mechanisms for synthesizing parts information into global shape representations. One possibility is selective feedforward convergence, as in Hubel and Wiesel's model of V1 orientation tuning based on patterned summation of thalamic inputs (Hubel and Wiesel, 1962). Feedforward models of object recognition depend critically on some kind of nonlinearity in the summation mechanism at each hierarchical processing stage (Riesenhuber and Poggio, 1999). Purely feedforward summation of parts information would facilitate rapid object recognition (Thorpe et al., 1996). The alternative is that recurrent processing within each hierarchical stage may iteratively refine feedforward stimulus signals that are initially weak or ambiguous (Chance et al., 1999; Douglas et al., 1995; Pugh et al., 2000; Salinas and Abbott, 1996). Although recurrent neural networks operate more slowly than feedforward networks, they provide a biologically plausible mechanism for generating the kind of nonlinear selectivity for element combinations (Salinas and Abbott, 1996) that global shape selectivity requires.

Here, we distinguish between these alternatives by analyzing shape information dynamics in posterior IT (TEO and posterior TE), which mediates between distributed, parts-level coding in lower-level ventral pathway areas and sparse, global coding in anterior IT. We recently reported that neurons in posterior IT vary in the way they signal the presence of multiple contour parts (Brincat and Connor, 2004). For present purposes, we define "parts" to be contour fragments that can be described geometrically with single values for position, orientation, and curvature; e.g., the backward facing concavity at the trailing edge of a dolphin's dorsal fin. We define "multipart configurations" to be combinations of such contour fragments; e.g., the trailing edge concavity plus the leading edge convexity of a dolphin's dorsal fin. Thus, "configurations" would always contain more than one distinct position value and in most cases distinct orientation and curvature values as well. Obviously, the terms "parts" and "configurations" could be used in other senses; our conclusions are limited to the specific definitions given above.

Some posterior IT cells show independent sensitivity to separate contour parts, i.e., the overall response is a linear function of the response components associated with each of those parts. Other neurons exhibit nonlinear selectivity for specific multipart configurations, i.e., the overall response is dominated by an interaction effect associated with the combined presence of two or more contour parts at specific positions. In the analyses presented here, we asked how these linear and nonlinear signals evolved through time after stimulus onset. We found that linear parts information emerged rapidly, whereas nonlinear signals for multipart configurations developed gradually, on a time course consistent with a recurrent network process for comparing parts information across neurons.

Results

We analyzed a sample of 89 posterior IT neurons recorded from two macaque monkeys. The monkeys

*Correspondence: connor@jhu.edu
[2] Present address: Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

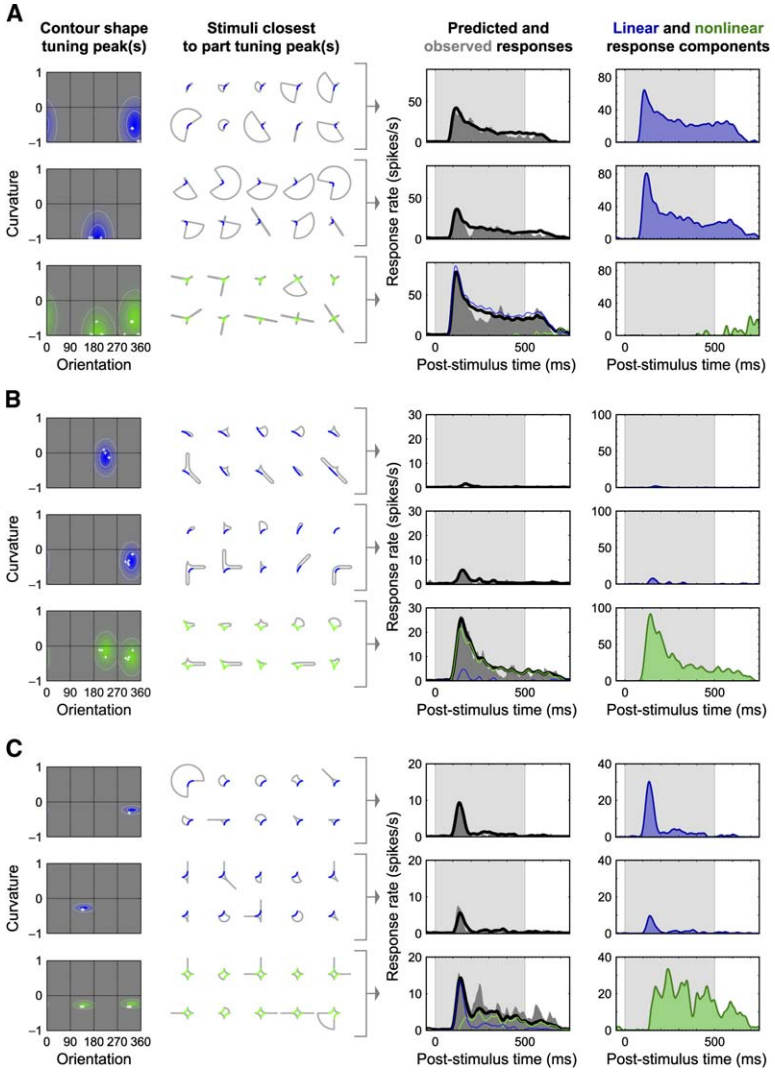

Figure 1. Temporal Dynamics of Linear Single-Part and Nonlinear Multipart Configuration Tuning for Three Individual Posterior IT Neurons

(A) Neuron with linear tuning for two types of contour parts. Left column: contour shape tuning functions corresponding to linear, single-part (blue) and nonlinear, multipart (green) response components. (The full tuning function also includes position dimensions not shown here; see Figure S1.) White crosses mark locations in shape tuning space of contour parts near the tuning peaks for a subset of ten shape stimuli (second column) that exhibited the closest match to the tuning peaks. Third column: averages across example stimuli of observed responses (gray histogram), predicted responses (black line), and single-part (blue line) and multipart (green line) components of the predicted responses. Light gray shading denotes stimulus presentation period. Right column: temporal profile of linear single-part and nonlinear multipart response components (based on all stimuli used to study this cell).

(B) Neuron with nonlinear tuning for a specific multipart configuration.

(C) Neuron with mixed tuning showing early responses to single parts and delayed responses to the multipart configuration.

performed a fixation task while stimuli were flashed in random order at the receptive field center. Example results are presented for a cell showing linear summation of information about separate parts (Figure 1A), a cell with nonlinear selectivity for multipart configurations (Figure 1B), and a cell with mixed linear/nonlinear response properties (Figure 1C). In each case, spiking responses to a large number of abstract shape stimuli were fitted with a model that characterized the cell's tuning for contour shape through time. The model was a combination of temporal response components representing either linear sensitivity to individual parts (blue histograms, right column) or nonlinear selectivity for multipart configurations (green histograms). These temporal response components reflected the evolving weights for Gaussian functions and combinations of Gaussian functions used to characterize smooth tuning for contour part curvature, orientation, and position (left column and Figures S1–S3). Example stimuli containing individual contour parts (blue) or multipart configurations (green) near the corresponding Gaussian tuning peaks are shown in the second column. The predicted (black line) and observed (gray histogram) average responses to these example stimulus sets are shown in the third column.

The linear cell (Figure 1A) responded strongly to shapes containing either sharp (angle-like) concave curvature oriented toward the right (top row) or sharp concave curvature oriented toward the left (middle row). The combined effect of these two concavities (bottom row) was approximately additive (gray histograms, third column). Correspondingly, this cell's linear, single-part model components (blue histograms, right column) were large, and there was little additional weight in the nonlinear, multipart model component (green histogram). This kind of linear summation of information about separate parts produces inherently ambiguous outputs because a given spike rate can correspond to various combinations of single or multiple contour parts (see Figure S1).

In contrast, the nonlinear cell (Figure 1B) was strongly selective for shapes containing a specific configuration of two parts: a broad concavity oriented near 200° and a broad concavity oriented near 315° (bottom row). Responses to shapes containing just one of these concavities were weak (top and middle rows). Responses to shapes containing both concavities were much stronger

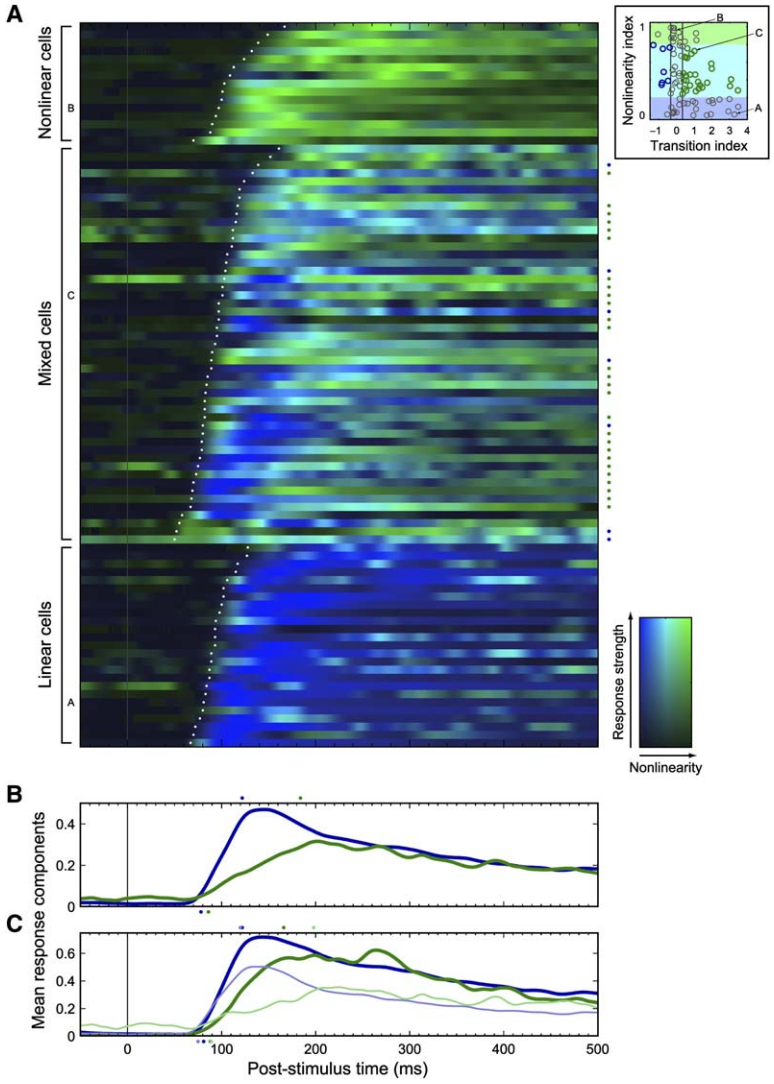

Figure 2. Population-Level Dynamics of Linear Single-Part and Nonlinear Multipart Responses in Posterior IT

(A) Temporal profiles of linear and nonlinear responses for 89 neurons. Each neuron is represented by a horizontal stripe in which brightness signifies total response strength, and color signifies relative weighting of linear (cf. Figure 1, blue histograms) and nonlinear (Figure 1, green histograms) response components (see color scale bar). Neurons are segregated into categories according to overall degree of nonlinearity (linear, nonlinear, and mixed) and ordered within categories by response onset latency (white dots). Within the group of cells showing a mixture of linear and nonlinear responses, many neurons exhibited a temporal transition from early linear → later nonlinear responses (green dots to the right of single-unit traces), whereas very few transitioned from nonlinear → linear responses (blue dots). Traces corresponding to example neurons displayed in Figures 1A–1C are labeled at left. Inset: scatter plot showing distribution of nonlinearity and temporal transition index values used to classify neurons. Blue, aqua, and green regions demarcate classification boundaries for linear, mixed, and nonlinear cell groups. Vertical lines indicate absolute bounds for classifying neurons as transitional; green and blue symbols indicate neurons transitioning from linear → nonlinear and nonlinear → linear responses, respectively. Arrows indicate example neurons from Figures 1A–1C.

(B) Population average temporal profiles of linear and nonlinear response components across all 89 neurons. Thick blue curve: population mean linear response component. Thick green curve: population mean nonlinear component. Colored dots indicate onset and peak times for corresponding curves.

(C) Average temporal profiles of linear and nonlinear response components within cell categories. Thick blue curve: mean linear response component across 25 linear cells. Thin blue curve: mean linear component across 34 transitional cells. Thick green curve: mean nonlinear component across 15 nonlinear cells. Thin green curve: mean nonlinear component across transitional cells.

than the sum of the single-part responses. Correspondingly, the linear response components (blue histograms) were small, whereas the nonlinear component (green histogram) was large. Visual inspection of Figure 1B and the full stimulus set in Figure S2 reveals that strong responses were uniquely associated with this combination of concavities and not with any other single feature or combination of features. The strong nonlinear component associated with this configuration in the fitted model provides quantitative confirmation of this observation. Nonlinear responses like these convey explicit information about multipart shape configurations.

The mixed response pattern in Figure 1C consisted initially of linear sensitivity to two types of contour parts but transitioned to nonlinear selectivity for the conjunction of those two parts. Thus, shapes containing either a broad concavity oriented near 315° (top row) or a broad concavity oriented near 135° (middle row) evoked transient responses. The response to shapes containing both types of concavities (bottom rows) included an early phase (100–200 ms poststimulus onset), approximately equal to the linear sum of these single-part responses, but also a later, sustained phase not predicted by the single-part responses. Correspondingly, the linear response components (blue histograms) consisted of transient peaks, whereas the nonlinear component (green histogram) had a delayed rise and a sustained profile. Neurons like these convey information that is initially ambiguous but then resolves into an explicit signal for a multipart configuration (see Figure S3).

To visualize how linear single-part and nonlinear multipart information developed in time across the population, we separately summed each neuron's linear and nonlinear response components and plotted the results against time relative to stimulus onset (Figure 2A). Each cell is represented by a horizontal stripe in which

brightness signifies total response strength and color signifies degree of nonlinearity (blue, linear; green, nonlinear; aqua, intermediate; see color scale bar). To clarify distinct trends in the population, we segregated cells into three categories according to whether their responses were predominantly linear (25 cells), predominantly nonlinear (15 cells), or mixed (49 cells; see y axis of inset scatter plot and Experimental Procedures). This classification is not meant to imply the existence of qualitatively distinct neural populations. Rather, our results are consistent with a continuous underlying distribution of response linearity and timing (see neural network simulation below). Within each category, cells are ordered according to response onset latency (white dots; see Experimental Procedures). Predominantly nonlinear cells responded at significantly ($p = 0.008$; randomization two-sample t test) longer latencies (mean ± SD: 112 ± 26 ms) than linear cells (93 ± 16 ms) and had delayed peak response times (181 ± 58 versus 146 ± 61 ms), although this trend fell slightly short of significance ($p = 0.08$). For mixed cells, linear responses predominated at earlier time points, and nonlinear responses were strongest at later time points: out of 34 mixed cells with a substantial change in response type over time (see x axis of inset scatter plot and Experimental Procedures), 79% (27/34; $p = 0.0004$; binomial test) had longer peak latencies for nonlinear as compared to linear response components, and across all 49 mixed cells, nonlinear response components were delayed by an average of 63 ms relative to linear components ($p = 0.007$; randomization paired t test).

These trends at the single-cell level produced strikingly different temporal profiles for linear single-part and nonlinear multipart information averaged across the population (Figure 2B). Linear response strength, averaged across all 89 neurons (thick blue line), grew rapidly over the course of about 40 ms, reaching 90% maximum at 122 ms after stimulus onset. Nonlinear response strength (thick green line) grew more slowly over the course of about 100 ms, reaching 90% maximum at 184 ms after stimulus onset. The 62 ms peak offset between linear and nonlinear response components was highly significant ($p < 0.0001$; randomization test). A similar temporal difference was observed between mean linear and nonlinear response components (Figure 2C, thin blue and green lines, respectively) averaged across only those neurons that exhibited substantial transitions in response type (linear → nonlinear or nonlinear → linear responses) over time (cf. Figure 2A, blue and green symbols). This temporal difference was also present, but smaller in magnitude, between responses of predominantly linear and nonlinear neurons (Figure 2C, thick blue and green lines, respectively). These results suggest that both within- and between-cell temporal differences contribute to the population-level dynamics, but within-cell transitions from early linear to later nonlinear responses make a larger contribution.

This large temporal lag between linear and nonlinear signals suggests a time-consuming network process that transforms feedforward signals to generate explicit tuning for multipart configurations. To illustrate how such a process might work, we simulated a neural network with a simple recurrent architecture based on previous models of nonlinear cortical processing (Chance et al., 1999; Salinas and Abbott, 1996). Each model unit received two feedforward inputs corresponding to a pair of contour part orientations (Figure 3A, blue arrows and tuning curves). The entire space of orientation pairs is represented by the 2D grid in Figure 3A. For simplicity, we simulated tuning for orientation only, though in reality, posterior IT neurons are tuned across multiple contour-related dimensions (Brincat and Connor, 2004). In addition, each model unit received recurrent connections from other units (green arrows and tuning curves). The recurrent connection weights had a difference-of-Gaussians (Mexican hat) pattern, such that each unit received excitatory inputs from units with similar tuning and inhibitory inputs from units with dissimilar tuning (Figure 3A, left). Each point on the 2D orientation-pair grid contained units ranging continuously from predominantly feedforward to predominantly recurrent connectivity (Figure 3A, right) (Chance et al., 1999).

The time course of inputs to a model unit at the center of the tuning grid (i.e., tuned for concavities at 0° and 180°) with strong recurrent connectivity illustrates how nonlinear tuning for multipart configurations could arise and why it would be delayed (Figure 3B). The temporal profile of the feedforward input strength (blue arrows) was matched to the time course of linear responses observed in linear posterior IT neurons (Figure 2B, thick blue line). When stimulated by an optimal multipart configuration (Figure 3B, top row), excitatory recurrent input (bright green) to the example unit is initially weak (125 ms) but increases gradually to a peak near 225 ms because of mutually excitatory interactions between units near the center of the tuning grid. Thus, the delayed activity of these highly recurrent units represents an inference about the presence of multipart configurations based on implicit comparison of input information across neurons. In contrast, when these units are stimulated by a single part (Figure 3B, bottom row), excitatory recurrent inputs are recruited only weakly, and activity is suppressed by inhibitory recurrent inputs (dark green) from units with different tuning profiles. In this way, the information conveyed by recurrent neurons remains largely uncontaminated by single-part information.

For these units with strong recurrent inputs (Figure 3B), the temporal response profile (Figure 3C) is thus dominated by nonlinear selectivity for multipart configurations (green), and this nonlinear response component has a slow rise and a delayed peak (near 200 ms). In contrast, model units with mixed recurrent/feedforward inputs exhibit a rapidly emerging linear sensitivity to individual parts (Figure 3D, blue) combined with a slowly rising nonlinear component (green). Model units with mainly feedforward inputs exhibit a rapidly emerging linear response component (Figure 3E, blue) with little nonlinear selectivity (green). These model response types parallel the neural response types exemplified in Figures 1A–1C.

We separately summed linear and nonlinear response components across model units with the same procedures used in the neural analysis. This produced temporal profiles for linear and nonlinear information (Figure 3F) that paralleled the neural population activity (Figures 2B and 2C). The agreement between model and neural temporal profiles was robust across a reasonable range of model parameters (see Experimental Procedures).

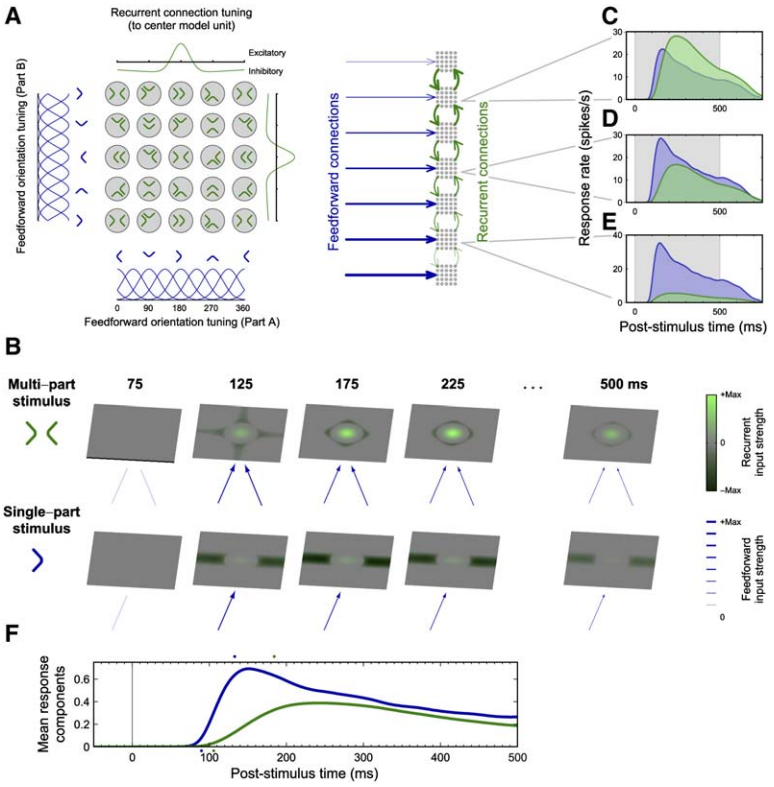

Figure 3. A Recurrent Network Model Explains Observed IT Response Dynamics

(A) Illustration of network architecture. Left: each model unit receives Gaussian-tuned feedforward inputs (blue curves) 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. Green illustrations show peak part combinations for a subsampling of model units. Each model unit also receives recurrent inputs from other units. The green curves represent strength of recurrent inputs to a unit at the center of the 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 (blue) and recurrent (green) connections vary continuously across the population of model units.

(B) Dynamic evolution of feedforward and recurrent inputs to a model unit at the center of the tuning grid. Top: an optimal multipart combination elicits an early feedforward response (blue arrows), followed by gradual recruitment of excitatory recurrent inputs from units with similar tuning (bright green patch at center of surface plot). Bottom: a single-part stimulus elicits a weaker feedforward input that is largely suppressed by recurrent inputs from units with dissimilar tuning (dark green patches in surface plot).

(C–E) Linear (blue) and nonlinear (green) response components of three example model units with strong (C), moderate (D), and weak (E) recurrent connectivity strengths.

(F) Average temporal profiles of linear and nonlinear response components across the population of model units. Blue curve: mean linear response component. Green curve: mean nonlinear response component.

Thus, a relatively simple network architecture with continuous variation in relative feedforward/recurrent connection strength can explain the existence of delayed, nonlinear tuning for object part combinations in posterior IT. We conclude that explicit information about object part combinations is generated by recurrent processing of initially linear signals for individual parts.

Discussion

Our results show that in posterior IT, after stimulus onset, there is a gradual transformation from linear information about single object parts (simple contour fragments) to nonlinear information about multipart configurations. This transformation is due to a combination of tuning changes within cells (from linear to nonlinear) and gradual appearance of activity in cells with predominantly nonlinear tuning. The end result is explicit representation of multipart configurations and, thus, a more compact and efficient neural code for shape (Brincat and Connor, 2004). Our conclusions are limited, in a strict sense, to what we defined here as parts: simple contour fragments that could be described with single values for position, orientation, and curvature. However, the important principle is that information about simpler components appears rapidly, whereas information about configurations of multiple components evolves gradually. One could speculate that similar dynamic transformations occur for simpler parts at earlier processing stages and for more complex parts at later processing stages in the ventral pathway. Further transformation of this kind could produce the global shape sensitivity and extreme sparseness observed in anterior IT (Baker et al., 2002; Freedman et al., 2003; Fujita et al., 1992; Gross et al., 1972; Logothetis et al., 1995; Tsunoda et al., 2001; Young and Yamane, 1992). A sparse, explicit representation of complex shape configurations is easier to decode for purposes of object recognition and would maximize efficiency of memory storage and associative learning (Fiete et al., 2004; Rolls and Treves, 1990).

The gradual emergence of configural information and gradual changes in single-neuron tuning reflect dynamic network processing of shape information, as opposed to a single feedforward pass through the ventral pathway. Recurrent processing could occur locally or could involve feedback from higher areas. However, many neurons exhibited nonlinear tuning at response onset, suggesting that feedback connections from higher-level areas—which are thought to modulate rather than establish neural selectivity (Bullier et al., 2001; Felleman and Van Essen, 1991)—do not play a primary role. Our network simulations demonstrated that local recurrent processing within IT could explain the observed time courses of linear and nonlinear shape information. Through recurrent processing, the network implicitly compares part information across neurons, in the sense

that active neurons with similar tuning reinforce each other. Over the course of about 60 ms, this comparison process yields increasingly strong inferences about multipart configurations, represented by the growing activity of highly selective nonlinear neurons. The potential role of recurrent processing is also supported by other simulation studies showing that recurrent networks exhibit complex dynamics (Pugh et al., 2000) and can produce nonlinear selectivity for conjunctions of inputs that are initially combined additively (Salinas and Abbott, 1996).

Fast reaction times in basic-level object categorization tasks (Thorpe et al., 1996) have been interpreted as evidence that object recognition relies primarily on feedforward processing. However, reaction times are virtually unaffected by manipulations known to degrade configural processing, suggesting that performance of these tasks depends mainly on detection of diagnostic parts (Rousselet et al., 2003). Rapidly emerging signals for individual parts (Figure 2B, blue lines) constitute a preliminary neural representation that could support this kind of early, nonconfigural categorization. In contrast, when object discrimination depends on specific part configurations, processing time is increased (by ~30–50 ms relative to tasks involving differences in simple part shape) (Arguin and Saumier, 2000; Wolfe and Bennett, 1997). Studies of illusory contour completion also indicate fast local processing of parts and slower integration of parts into global figures (Ringach and Shapley, 1996). Gradually emerging signals for multipart configurations (Figure 2B, green lines) constitute a more evolved neural representation that could support these slower configuration-based perceptions.

The transition we observed from early representation of parts to later representation of multipart configurations may help to explain previously observed dynamics of face-related neural responses. Early responses of face-selective cells in anterior IT carry information sufficient for distinguishing faces from other object categories, but information about individual facial identity does not develop until ~50 ms later (Sugase et al., 1999). Similar dynamics have been described for face-related MEG responses in human occipitotemporal cortex (Liu et al., 2002). Categorization of faces versus nonfaces requires only detection of diagnostic parts and could conceivably be based on early linear signals for individual parts. In contrast, recognition of specific faces requires configural information (Tanaka and Farah, 1993) and might be delayed by the slower emergence of nonlinear signals for multipart configurations.

Recurrent processing to derive information about combinations of simpler elements may be a general feature of vision. Area MT is the first stage at which neurons demonstrate selectivity for the direction of pattern (e.g., plaid) motion, which requires integration of multiple signals for component (e.g., grating) motion directions (Movshon et al., 1985; Rodman and Albright, 1989). MT responses encode component motion initially and then gradually become selective for pattern motion, with a delay of ~60 ms (Pack et al., 2001; Smith et al., 2005). This delay is remarkably similar to what we observed in posterior IT, suggesting a common mechanism for integrating component signals to derive higher-order structural information. MT temporal profiles at the level of individual cells are a matter of debate. Some results suggest that most MT cells transition from component to pattern tuning (Pack et al., 2001), whereas other results suggest distinct subpopulations with stable component or pattern tuning but divergent response latencies (Smith et al., 2005). A complete transition toward pattern tuning under some circumstances could reflect a complete shift toward perception of pattern (plaid) motion, which conflicts with perception of component grating motion. For our shape stimuli, there was no such conflict between component and pattern perception, which could explain the persistent component tuning we observed in a subpopulation of posterior IT cells. Pattern responses are optimal for supporting global percepts, but persistent responses to simpler components could serve to make information about local structure continuously available.

Experimental Procedures

Behavioral and Neurophysiological Methods

We recorded spiking activity of well-isolated single neurons from two monkeys (*Macaca mulatta*) trained to maintain fixation for a juice reward. Neurons were sampled in an unbiased manner from posterior and central IT cortex (TEO/PIT and posterior TE/CIT) in the lower bank and lip of the superior temporal sulcus from stereotaxic A-P −5 to +8.5 mm. No statistically significant trends in response characteristics were observed in the A-P direction, so all neurons were grouped together in the analyses presented here.

Neurons were studied with parametric shape-stimulus sets in which simple parts (straight and curved contour fragments) were varied systematically and permuted into a large number of combinations (Figures S1–S3). For each cell, stimulus sets were optimized for color, orientation, local curvature, and size, based on preliminary tests. Stimuli were displayed as colored silhouettes against a blank gray background and were flashed in random order at the center of a cell's receptive field for 500 ms each, with a 250 ms interstimulus interval, during 4.5 s fixation trials (6 stimuli/trial). Post hoc tests were used to examine responses of each neuron to preferred and nonpreferred shapes across a range of absolute positions and sizes; the shape selectivity of almost all neurons was highly consistent across the ranges within which they responded (Brincat and Connor, 2004) (Figure S4).

Data Analysis

Spike trains were smoothed with an asymmetric Gaussian function (15 ms SD causal side, 5 ms SD acausal side) and averaged across all repetitions of each stimulus. This procedure yielded a robust estimate of instantaneous response rate that avoided backward bias in time by means of the primarily causal weighting in the smoothing filter (cf. Thompson et al. [1996]).

These responses were fitted with a model that characterized each neuron's tuning for contour shape (curvature and orientation) and position (x,y object-relative position and x,y absolute position) through time (relative to stimulus onset). The model described shape selectivity with a combination of multiple excitatory and inhibitory Gaussian tuning functions (representing linear sensitivity to single parts) and all higher-order products of same-sign tuning functions (representing nonlinear sensitivity to multipart configurations). (Results are only reported here for excitatory response components.) Each of these linear and nonlinear model components was fitted with a temporal weight function describing how its contribution to neural responses varied across time (in 5 ms steps). Model fitting was carried out in three steps: (1) shape and position tuning parameters for varying numbers (1–6) of Gaussian functions were fitted to time-averaged response rates with an iterative least-squares algorithm, (2) an optimal number of Gaussian functions was selected based on a stepwise regression procedure (Brincat and Connor, 2004), and (3) these parameters were held constant while temporal weighting functions for each component were fitted to instantaneous response rates.

Response onset latencies were defined by the first time point at which instantaneous response rate exceeded baseline (mean

response from −50 to +50 ms relative to stimulus onset) by 10% of the baseline-to-peak response range and remained above that level for at least 25 ms (to avoid noise effects). Peak response latencies were defined by the first time point at which instantaneous response rate exceeded baseline by 90% of the response range. Reported latencies were calculated from the fitted total excitatory response strength; the raw neural responses yielded similar results. The index of nonlinearity represented by the color scale in Figure 2A (and plotted in the inset scatter plot) is defined as: $NL/(L + NL)$, in which L and NL represent a neuron's summed linear and nonlinear response components. This index ranges from 0 (purely linear) to 1 (purely nonlinear). The same index, integrated across time from response onset to stimulus offset, was used to classify neurons as predominantly linear (0–0.2), predominantly nonlinear (0.8–1), or mixed (0.2–0.8). A temporal transition index (Figure 2A, inset) was used to measure the degree to which the response type of single neurons with both meaningful linear and nonlinear response components changed over time. This index was the signed difference between the peak latency of a cell's nonlinear and linear response components, expressed as a proportion of the smaller (earlier) of the two latency values. Negative transition index values indicate a transition from nonlinear → linear responses, a value of 0 corresponds to no component latency difference, and positive values indicate a linear → nonlinear transition; values of ±1 correspond to the situation in which the temporal difference between response components is equal to the absolute latency of the earlier component. Absolute values of the transition index were used to classify neurons as transitional (>0.33) or nontransitional (<0.33).

All reported significance levels are based on two-tailed randomization tests of the associated statistic, unless otherwise noted. For comparisons based on population-mean response functions (e.g., comparison of linear and nonlinear peak latencies in Figure 2B), we used a randomization test in which the observed difference was compared to a null-hypothesis distribution of difference values based on randomly reassigning single-unit response components to linear and nonlinear categories.

We studied 109 shape-selective neurons in posterior IT. Of these, 89 cells were best described by a model with excitatory tuning for more than one type of part and were, therefore, included in the analyses in this report. 20 cells with excitatory tuning for only a single type of part were excluded because the current report focuses on integration of multiple parts. Further methodological details can be found in Brincat and Connor (2004).

Recurrent Network Simulation

We simulated a neural network with recurrent circuitry similar to previously published models of response selectivity in parietal and striate cortex (Chance et al., 1999; Salinas and Abbott, 1996). The intracellular voltage response of each model neuron, $V(t)$, was determined by the sum of feedforward and recurrent inputs through the standard rate-model equation (Dayan and Abbott, 2001)

$$\tau\frac{dV}{dt} + V(t) = g^{ff}V^{ff}(t) + g^{rcr}V^{rcr}(t) \quad (1)$$

in which $V^{ff}(t)$ and $V^{rcr}(t)$ are the feedforward and recurrent input voltages, g^{ff} and $g^{rcr} = (1 - g^{ff})$ are the strengths of feedforward and recurrent inputs, and τ is the membrane time constant (15 ms). Values of g^{ff} were systematically and evenly varied across model neurons, from 1.0 (i.e., purely feedforward input) to g^{ff_min} (cf. Chance et al. [1999]); a g^{ff_min} value of 0.25 yielded a close approximation to the observed data. The feedforward connections confer additive selectivity for two Gaussian tuning ranges along separate stimulus dimensions (cf. Salinas and Abbott [1996]), which here could be thought of as local contour orientation for two spatially separate object parts, A and B

$$V_i^{ff}(\theta_A, \theta_B, t) = \left[e^{\frac{(\theta_A - \mu_{i,A})^2}{-2\sigma_{ff}^2}} + e^{\frac{(\theta_B - \mu_{i,B})^2}{-2\sigma_{ff}^2}} \right] \cdot R_{linear}(t) \quad (2)$$

In which $\mu_{i,A}$ and $\mu_{i,B}$ are the ith model unit's tuning peaks for parts A and B (evenly distributed across all pairs of part orientations), σ_{ff} is the Gaussian tuning width (30°; the empirically observed population mean value for orientation), and $R_{linear}(t)$ is the empirically observed mean linear response component of predominantly linear neurons (Figure 2B, thick blue line). The recurrent input had a difference-of-Gaussians (Mexican hat) profile with excitatory connections between neurons with similar tuning for parts A and B and inhibitory connections for neurons with dissimilar tuning

$$V_i^{rcr}(t) = \sum_j \left[\left(w_{EXC}c_{EXC}e^{\frac{(\Delta\mu_{ij})^2}{-2\sigma_{EXC}^2}} - w_{INH}c_{INH}e^{\frac{(\Delta\mu_{ij})^2}{-2\sigma_{INH}^2}} \right) \cdot R_j(t-1) \right] \quad (3)$$

in which $\Delta\mu_{ij}$ is the Euclidean distance between pairs of neural tuning peaks in the 2D space defined by tuning dimensions A and B, c_{EXC} and c_{INH} are constants that make the unweighted excitatory and inhibitory Gaussian connectivity functions sum to one, w_{EXC} (1.0) and w_{INH} (2.0) are weights on the normalized excitatory and inhibitory functions, σ_{EXC} (30°) and σ_{INH} (300°) represent the spread of excitatory and inhibitory recurrent connections (cf. values in Salinas and Abbott [1996]), $R_j(t-1)$ represents the firing rate of the jth neuron in the previous 1 ms time step, and the summation is over all 22,500 model neurons (50 × 50 distinct tuning peaks × 9 distinct g^{ff} values). Finally, intracellular responses were converted to firing rates at each time step through a standard power-law function $R(t) = \alpha \max(V(t), 0)^{\gamma}$, with gain factor α (6.0) set to produce stable propagation of response rates similar to the observed values, and exponent γ (1.1) set to give an overall degree of nonlinearity across units similar to the observed value. The simulation results in Figure 3 are qualitatively robust to reasonable variations in all parameter values. Networks with only feedforward connectivity ($g^{ff\_min} = 1$) failed to show the temporal dynamics observed in the data and in the recurrent network.

Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/49/1/17/DC1/.

Acknowledgments

We thank L. Ziehm, W. Nash, W. Quinlan, and B. Sorensen for technical assistance and A.J. Bastian, K. Zhang, and L. Tran for helpful comments and discussions. This work was supported by the National Eye Institute and by the Pew Scholars Program in the Biomedical Sciences.

Received: June 6, 2005
Revised: September 13, 2005
Accepted: November 4, 2005
Published: January 4, 2006

References

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

Baker, C.I., Behrmann, M., and Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat. Neurosci. *5*, 1210–1216.

Barlow, H. (2001). Redundancy reduction revisited. Network *12*, 241–253.

Brincat, S.L., and Connor, C.E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat. Neurosci. *7*, 880–886.

Bullier, J., Hupe, J.M., James, A.C., and Girard, P. (2001). The role of feedback connections in shaping the responses of visual cortical neurons. Prog. Brain Res. *134*, 193–204.

Chance, F.S., Nelson, S.B., and Abbott, L.F. (1999). Complex cells as cortically amplified simple cells. Nat. Neurosci. *2*, 277–282.

Dayan, P., and Abbott, L.F. (2001). Theoretical Neuroscience. (Cambridge, MA: MIT Press).

Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A., and Suarez, H.H. (1995). Recurrent excitation in neocortical circuits. Science *269*, 981–985.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Fiete, I.R., Hahnloser, R.H., Fee, M.S., and Seung, H.S. (2004). Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. J. Neurophysiol. *92*, 2274–2282.

Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. *23*, 5235–5246.

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. Nature *360*, 343–346.

Gallant, J.L., Braun, J., and Van Essen, D.C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science *259*, 100–103.

Gross, C.G., Rocha-Miranda, C.E., and Bender, D.B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J. Neurophysiol. *35*, 96–111.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.

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

Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. Nat. Neurosci. *5*, 910–916.

Logothetis, N.K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Curr. Biol. *5*, 552–563.

Movshon, J.A., Adelson, E.H., Gizzi, M.S., and Newsome, W.T. (1985). The analysis of visual moving patterns. In Pattern Recognition Mechanisms, C. Chagas, R. Gattass, and C. Gross, eds. (New York: Springer), pp. 117–151.

Pack, C.C., Berezovskii, V.K., and Born, R.T. (2001). Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. Nature *414*, 905–908.

Pasupathy, A., and Connor, C.E. (1999). Responses to contour features in macaque area V4. J. Neurophysiol. *82*, 2490–2502.

Pasupathy, A., and Connor, C.E. (2002). Population coding of shape in area V4. Nat. Neurosci. *5*, 1332–1338.

Pugh, M.C., Ringach, D.L., Shapley, R., and Shelley, M.J. (2000). Computational modeling of orientation tuning dynamics in monkey primary visual cortex. J. Comput. Neurosci. *8*, 143–159.

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

Ringach, D.L., and Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. Vision Res. *36*, 3037–3050.

Rodman, H.R., and Albright, T.D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). Exp. Brain Res. *75*, 53–64.

Rolls, E.T., and Treves, A. (1990). The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. Network *1*, 407–421.

Rousselet, G.A., Mace, M.J., and Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. J. Vis. *3*, 440–455.

Salinas, E., and Abbott, L.F. (1996). A model of multiplicative neural responses in parietal cortex. Proc. Natl. Acad. Sci. USA *93*, 11956–11961.

Smith, M.A., Majaj, N.J., and Movshon, J.A. (2005). Dynamics of motion signaling by neurons in macaque area MT. Nat. Neurosci. *8*, 220–228.

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

Tanaka, J.W., and Farah, M.J. (1993). Parts and wholes in face recognition. Q. J. Exp. Psychol. A *46*, 225–245.

Thompson, K.G., Hanes, D.P., Bichot, N.P., and Schall, J.D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. J. Neurophysiol. *76*, 4040–4055.

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

Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat. Neurosci. *4*, 832–838.

Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D.G. Ingle, M.A. Goodale, and R.J.Q. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–586.

Wolfe, J.M., and Bennett, S.C. (1997). Preattentive object files: shapeless bundles of basic features. Vision Res. *37*, 25–43.

Young, M.P., and Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. Science *256*, 1327–1331.