Responses to Compound Objects in Monkey Inferotemporal Cortex: The Whole Is Equal to the Sum of the Discrete Parts

Arun P. Sripati¹ and Carl R. Olson^{1,2}

¹Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, and ²Department of Neuroscience, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

It is commonly thought that neurons in monkey inferotemporal cortex are conjunction selective—that a neuron will respond to an image if and only if it contains a required combination of parts. However, this view is based on the results of experiments manipulating closely adjacent or confluent parts. Neurons may have been sensitive not to the conjunction of parts as such but to the presence of a unique feature created where they abut. Here, we compare responses to two sets of images, one composed of spatially separate and the other of abutting parts. We show that the influences of spatially separate parts combine, to a very close approximation, according to a linear rule. Nonlinearities are more prominent—although still weak—in responses to images composed of abutting parts.

Introduction

Inferotemporal cortex (IT) is the terminus of the ventral stream of visual areas in primates and as such plays a critical role in visual object recognition. The contribution of IT to object recognition is thought to hinge critically on its containing neurons that respond selectively to natural images. The mechanism of image selectivity is widely thought to be nonlinear on the basis of the classic observation that removing any part from a neuron's preferred image can cause a dramatic reduction in the response (Perrett et al., 1982; Desimone et al., 1984; Tanaka et al., 1991; Kobatake and Tanaka, 1994; Ito et al., 1995; Yamane et al., 2006). This phenomenon is commonly assumed to arise from a process whereby the conjunction of parts exerts a much stronger influence on neuronal firing than predicted from linear summation of the influences of the individual parts. However, it could also arise if the neuron were selective for some simple juxtapositional feature created where the parts abut. Nonlinear interactions between parts on closed surfaces may likewise indicate sensitivity to the smooth merger between the parts (Brincat and Connor, 2004, 2006; Yamane et al., 2008).

A few recent studies have circumvented this problem by measuring neuronal responses to objects composed of spatially separate parts (Baker et al., 2002; De Baene et al., 2007). These have revealed statistically significant nonlinear part-part interactions in ~10% of IT neurons (Baker et al., 2002). However, this number is difficult to interpret. On the one hand, it may overestimate the contribution of nonlinear part-part interactions because the effects, although statistically significant, could be very weak. On

the other hand, it may underestimate the rate of incidence of nonlinear interactions because it was based on testing each neuron with only a small set of parts. It is possible that the observed nonlinear interactions occurred when parts were especially ineffective or especially effective at eliciting neuronal activity. In fact, superadditive nonlinear interactions between ineffective and subadditive nonlinear interactions between effective parts occur in a common model of neuronal activation based on passing linearly summed inputs through a sigmoidal activation function (Koch, 1999).

To resolve this issue, we have assessed neuronal responses to objects composed of parts that cover a broad range in their individual ability to modulate neuronal activity and have analyzed the results using a measure of signal strength that allows for unbiased comparison between linear effects (dependent simply on which parts are present) and nonlinear effects (dependent on the particular combination of parts that is present). For objects containing spatially separate parts, we find that nonlinear effects occur and that when present they nudge neurons slightly toward conjunction selectivity. However, they are extremely weak, accounting for only around a twentieth of the modulation in firing rate induced by the stimulus, and develop only late in the response. For objects composed of confluent parts, nonlinear interactions are stronger but still not prominent, accounting for only around a quarter of the stimulus-induced modulation in neuronal firing rate.

Materials and Methods

Neurophysiological methods. Two rhesus macaque monkeys, one male and one female (laboratory designations Je and Ec), were used. All experimental procedures were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee and were in compliance with the guidelines set forth in the United States Public Health Service Guide for the Care and Use of Laboratory Animals. All aspects of the behavioral experiment (stimulus presentation, eye position monitoring, and reward delivery) were under control of a computer running Cortex software (NIMH Cortex). Eye position was monitored by means of a scleral search coil system (Riverbend Instruments). At the beginning of

Received Jan. 3, 2010; revised March 3, 2010; accepted March 26, 2010.

Research was supported by National Institutes of Health (NIH) Grants R01 EY018620 and P50 MH084053 and the Pennsylvania Department of Health through the Commonwealth Universal Research Enhancement Program. Technical support came from NIH Grant P30 EY08098. Collection of MR images was supported by NIH Grant P41 RR03631. We thank Karen McCracken for technical assistance.

Correspondence should be addressed to Arun P. Sripati, at his present address: Center for Neuroscience, Indian Institute of Science, Bangalore 560012, India. E-mail: sparun@cns.iisc.ernet.in.

DOI:10.1523/JNEUROSCI.0016-10.2010

Copyright © 2010 the authors 0270-6474/10/307948-13\$15.00/0

each day's session, a varnish-coated tungsten microelectrode with an initial impedance of $\sim 1.0\text{ M}\Omega$ at 1 kHz (FHC) was introduced into the temporal lobe through a transdural guide tube advanced such that its tip was approximately 10 mm above IT. The electrode was then advanced by use of a micromanipulator until phasic visual responses were observed. Action potentials of single neurons were isolated from the multineuronal trace using a commercially available spike-sorting system (Plexon Inc.). All waveforms were recorded during the experiments and spike sorting was performed offline using commercially available spike-sorting software (Plexon Inc.). At the end of the data collection period, recording sites were established by structural MRI to occupy the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus (at levels ranging from A14–22 mm in Je and A9–18 mm in Ec relative to the interaural plane). The recording locations fell well within the range investigated in previous studies of shape selectivity in IT (Tanaka et al., 1991; Kobatake and Tanaka, 1994; Ito et al., 1995; Brincat and Connor, 2004; Yamane et al., 2008).

Baton stimuli. Upon identifying and isolating a neuron, we performed preliminary testing by presenting (for 500 ms each at fixation) a series of items drawn from a 21-image library of parts (supplemental Fig. S1, available at www.jneurosci.org as supplemental material). We selected for use in systematic tests seven parts that seemed to elicit responses from the neuron covering a considerable range in strength and an eighth "blank" part: the unadorned end of the stem. The full stimulus set for each neuron consisted of 64 baton stimuli obtained by placing these eight parts in all possible combinations at the top and bottom of a vertical stem (Fig. 1). Shapes placed at the top and bottom were identical short of a 180° rotation. Each baton, consisting of a stem and two appended shapes, was $\sim 2^\circ$ wide and 6° tall. Batons were presented with the center of the rectangular stem at fixation. The responses of a total of 104 neurons were recorded in this manner.

Hourglass stimuli. We selected a neuron for recording if it appeared to respond to at least one of the hourglass stimuli. This was the only step of selection. Each stimulus in the hourglass set consisted of two parts, each of which was a subset of line segments of an equilateral triangle. Each segment of the triangle was 2.71° long and 0.22° thick. For testing in every neuron, we selected the exact same set of eight parts, namely the eight possible subsets of segments which together make up the triangle. Parts placed at the top and bottom were identical short of a 180° rotation. This resulted in a total of 64 stimuli obtained by combining the eight top and eight bottom parts in a combinatorial fashion. Stimuli were presented with the junction of the two triangles at fixation. The responses of 142 neurons were recorded in this manner. Among these, a subset of 58 neurons had responses to both the baton and hourglass stimulus sets.

Task and trial design. We monitored neuronal activity during passive fixation. A trial began with onset of a small red fixation cross (measuring 0.4°) and attainment of fixation. After a delay of 200 ms, eight stimuli were presented in succession, each with a duration of 200 ms followed by an interval of 200 ms during which the fixation cross was again visible. After completion of the sequence, the display vanished and the monkey was rewarded with ~ 0.1 cc of water. Although the fixation window was large (4.2°), we found on *post hoc* analysis that the gaze remained closely centered throughout the duration of each trial. The average across sessions of the SD of horizontal and vertical gaze angle was 0.2° .

The trials conformed to 16 conditions. Each of the 16 conditions was identified with a set of eight stimuli. The 64 stimuli were divided into nonoverlapping arbitrary groups of eight across conditions 1–8. On each trial conforming to a given condition, the stimuli associated with it were presented once each in random order. The fact that no stimulus was repeated within a trial minimized the potential for repetition suppression. The 64 stimuli were parceled out again across conditions 9–16 subject to the constraint that the eight stimuli in each condition had to have been drawn one each from conditions 1–8. Thus any two images presented in the same trial under conditions 1–8 were presented in separate trials under conditions 9–16. This minimized the potential for image-image interactions to influence the results. The 16 conditions were imposed in random order subject to the constraint that one trial conforming to each condition had to be completed successfully before a new

block began. The fact that each stimulus was presented only twice within each block of 16 trials minimized the potential for repetition suppression. A complete session consisted of four blocks of 16 trials over the course of which each of the 64 stimuli was presented eight times.

Noise-free estimate of variance explained by main and interaction effects. The aim of this step of analysis was to compare main and interaction effects with respect to their ability to explain the variance of a neuron's firing rate across stimuli. To do this accurately required bypassing a distortion inherent in measures of explained variance produced by an ANOVA. In an ANOVA assessing the dependence of firing rate on the identity of eight top and eight bottom parts, the underlying model possesses 64 degrees of freedom (one for baseline firing rate, seven for main effects of top, seven for main effects of bottom and 49 for interaction effects). Since the number of degrees of freedom in the data (responses to 64 stimuli) equals the number of degrees of freedom in the model, the model can always account for 100% of the explainable variance (i.e., the cross-stimulus variance) in firing rate. However, the variance explained contains terms that reflect both signal and noise: $M = M_s + M_n$, where M is variance explained by main effects, M_s is the moiety due to signal, and M_n is the moiety due to noise; and likewise, for interaction effects, $I = I_s + I_n$. The presence of the noise terms acts to move the observed ratio between main and interaction effects (M/I) closer to one than the true ratio (M_s/I_s).

To bypass this problem, we adopted an alternative approach guaranteeing that the contribution of noise to each estimate of explained variance will fluctuate around zero. For each neuron's odd-numbered trials, we performed an ANOVA with top and bottom part identity as factors and with firing rate in a window 50–350 ms after stimulus onset as the dependent variable. The underlying model contained eight coefficients reflecting the contributions of top parts, eight coefficients reflecting the contributions of bottom parts, and 64 coefficients reflecting the contributions of specific top-bottom combinations. From these coefficients, we generated three eight by eight matrices: a "top" matrix in which each row contained a coefficient (identical across columns) representing the contribution of the corresponding top, a "bottom" matrix in which each column contained one coefficient (identical across rows) representing the contribution of the corresponding bottom, and an interaction matrix in which the 64 interaction coefficients represented the contributions of the corresponding top-bottom pair. We analyzed the correlation of each of these matrices with the matrix of mean firing rates elicited by the 64 stimuli 50–350 ms after stimulus onset on even-numbered trials. We computed the fraction of variance across stimuli explained by each class of effect as $\text{sign}(r)^2 r^2$, where r was the corresponding correlation coefficient.

The impact of this correction was nontrivial. In the case of pure noise, with the firing rate on odd- and even-numbered trials unrelated to stimulus identity, the estimate of variance explained by main and interaction effects would be 22% and 78% respectively without correction whereas, with correction, it would be zero on average. For the neuron of Figure 1, the uncorrected estimates of variance explained by main and interaction effects were, respectively, 86% and 14%, whereas the corrected estimates were 77% and 2%. On the basis of the uncorrected measures, we would say that the main effects explained six times as much variance as the interaction effects. On the basis of the corrected measures, we would say 40 times as much.

Linear model. In the linear model, the response R_L to a baton containing parts P_i at the top and P_j at the bottom is given by the following:

$$R_L(P_i, P_j) = a_i + b_j + c, \quad (1)$$

where a_i and b_j represent the contributions of P_i and P_j , respectively, and c is a baseline constant. Because the presence of the eighth top (or bottom) part implies the absence of the other seven, only seven degrees of freedom are associated with each part. This is reflected in the constraint that $a_1, a_2, \dots, a_8$ and $b_1, b_2, \dots, b_8$ must sum to zero. Thus, the model possesses 15 degrees of freedom (seven for the top part, seven for the bottom part and one for the baseline firing rate). These parameters, derivable by linear regression, were obtained by means of a built-in Matlab function (`anovan`, Matlab 7.0, MathWorks).

Figure 1. Responses of an IT neuron to a set of 64 baton stimuli. Batons in the same row have the same top part (rows progress downward in order of part efficacy) and batons along columns have the same bottom part (columns progress leftward in order of part efficacy). The response to each baton is displayed directly below it, with black bars representing the average firing rate in each time bin and tick marks indicating the occurrence of action potentials on individual trials. Stimuli were displayed for 200 ms with onset corresponding to the left edge of the window. The most effective stimulus (marked with an asterisk) was the one containing the preferred top and preferred bottom.

Sigmoidal model. The response of the sigmoidal model R_S simply transforms the output of the linear model R_L using a sigmoidal function given by the following:

$$R_S = A \int_{-\infty}^{R_L} N(\mu, \sigma), \quad (2)$$

where A is the amplitude of the sigmoid, and $N(\mu, \sigma)$ is the Gaussian distribution with mean μ and SD σ . The sigmoidal function is monotonic with a maximum of A and minimum of zero. The point of half-maximum of the sigmoid and its rate of rise are controlled by the mean and SD of the Gaussian respectively. Thus, there are three additional parameters (A , μ , and σ) in the sigmoidal model compared with the linear model. We estimated the parameters of the sigmoidal model by minimizing the sum-of-squared error between the observed data and the model's response using built-in functions in Matlab (lsqcurvefit, Matlab 7.0, MathWorks).

Statistical comparison of the linear and sigmoidal models. For any dataset, fits obtained using the sigmoidal model must be as good as or better than the fits from the linear model simply because the sigmoidal model has three additional parameters. We performed a partial F test to determine whether the improvement obtained with the sigmoidal model was significantly in excess of the improvement expected from the extra degrees of freedom in themselves.

Impact of interaction effects on shape selectivity. To quantify neuronal shape tuning, we computed rank at half-height of the curve obtained by plotting firing rate (normalized to the response to the best baton) against rank (from best to worst baton). To characterize the effect of removing the interaction effects, we recalculated the tuning curve using firing rates predicted by main effects alone. To characterize the effect of randomizing the relation between interaction effects and main effects, we recalculated the tuning curve after shuffling the interaction effects. Before shuffling, the neuron's interaction effects formed an 8×8 matrix with rows and columns corresponding to tops and bottoms and with entries corresponding to the interaction coefficients from the ANOVA. The sum of the eight coefficients in any given row or column was zero by the definition of interaction effects as orthogonal to main effects. To maintain this constraint while disrupting the relationship between interaction and main effects, we randomly shuffled the values within each row and then within each column.

Comparison of "part" and "part $\times$ location" models. The model implicit in the ANOVA performed on data from individual neurons is a "part $\times$ location" model in the sense that completely independent sets of coefficients embody shape selectivity at the top and bottom locations. This model, given by Equation 1, has 15 free parameters. We compared this to a simpler "part" model based on the assumption that a neuron exhibits identical part selectivity with only a difference in gain at the top and bottom locations. The response of the part model to a baton with parts P_i at the top and P_j at the bottom is given by the following:

$$R_{\text{part}}(P_i, P_j) = a_i + k a_j + c, \quad (3)$$

where c is a baseline constant, a_i is the contribution of P_i , a_j is the contribution of P_j , and k is a constant scaling factor capturing the difference in gain between the top and bottom locations. This model has 9 degrees of freedom (7 for part identity and 2 for the constants k and c). We adjusted the parameters of the part model to obtain an optimal fit to each neuron's data by minimizing the sum-of-squared errors between the predicted and observed responses (lsqcurvefit, Matlab 7.0, MathWorks). We then used a partial F test to determine whether the improvement in fit attained by use of the part $\times$ location model compared with the part model was significantly greater than expected simply from the increase in the number of free parameters.

Time course of main and interaction effects. The aim of this step of analysis was to calculate the strength of main and interaction effects as a function of time elapsed since stimulus onset. To do this accurately required correcting for a systematic inflation of signal strength occurring when an ANOVA is applied to the full body of raw data. To achieve this

correction, we first performed for each neuron an ANOVA on data from odd-numbered trials, with firing rate in a window 50–200 ms after stimulus onset as the dependent variable and with the identity of the top part and the identity of the bottom part as factors. We then passed the firing rates in successive small bins on even-numbered trials through a filter based on the coefficients of the linear model generated by the ANOVA, using an approach such that the estimate of signal strength should hover around zero in the case of pure noise. The procedure was based on the reasonable assumption that stimulus selectivity, if present instantaneously, conformed to the pattern present on average across the 50–200 ms window.

The underlying model contained eight coefficients reflecting the contributions of top parts, eight coefficients reflecting the contributions of bottom parts, and 64 coefficients reflecting the contributions of specific top-bottom combinations. If any top part, bottom part or combination tended to increase the firing rate above (or decrease it below) the mean, the corresponding coefficient was positive (or negative). We converted each coefficient by a sign operation to +1 or –1. Coefficients of +1 (and –1) thus marked "preferred" (and "non-preferred") parts or combinations of parts. We used the resulting arrays of 8, 8, and 64 signed values as templates representing, respectively, the neuron's pattern of top-part, bottom-part, and combination selectivity during the 50–200 ms sampling interval.

For each 10 ms sampling period on even-numbered trials, we then performed an ANOVA with firing rate as the dependent variable and with top and bottom part identity as factors. This yielded eight coefficients reflecting the contributions of top parts, eight coefficients reflecting the contributions of bottom parts, and 64 coefficients reflecting the contributions of specific top-bottom combinations. We multiplied each coefficient by the corresponding signed value in the template. Then we computed the average of the resulting top-part terms, the average of the resulting bottom-part terms and the average of the resulting combination terms. Each average represented the degree by which the firing rate elicited by preferred stimuli exceeded the firing rate elicited by non-preferred stimuli. These three values can be thought of as reflecting how strongly on average across stimuli the influence of the top part in isolation, the influence of the bottom part in isolation and the influence of the particular combination pushed neuronal activity toward the pattern expected from the analysis performed on odd-numbered trials. The main effect strength was calculated as the sum of the top-part and bottom-part averages.

Latency of main and interaction effects across the population. To calculate the latency of main and interaction effects for the entire population of neurons, we took the distribution of the respective effect strengths across cells for each time bin, and identified the time bin in which it first deviated significantly from a mean of zero using a t test ($\alpha = 0.01$). A similar procedure was used to calculate the time of onset of the response itself as distinct from stimulus-selective activity.

Results

We monitored the activity of neurons in the left inferotemporal cortex of two monkeys during passive viewing of images from two stimulus sets. Each stimulus set consisted of 64 stimuli created by combining eight top parts and eight bottom parts. Shapes used at the bottom were identical to those used at the top short of a 180° rotation. In the baton stimulus set, each image consisted of two parts joined by a vertical stem (Fig. 1). One of the eight parts was the unadorned end of the stem. The remaining seven were selected from a library of 21 available parts (supplemental Fig. S1, available at www.jneurosci.org as supplemental material) on the basis that, when presented in isolation during preliminary testing, they appeared to elicit a response from the neuron under study. In the hourglass stimulus set, the parts were the eight shapes (one a blank) obtained by hiding or displaying the three line segments of an equilateral triangle (Fig. 2). These sets differ with regard to the way in which an interaction effect (reflecting neuronal sensitivity to a particular combination of top and bot-

Figure 2. Responses of an IT neuron to the set of 64 hourglass stimuli. Conventions are as in Figure 1.

Figure 3. Results of analyses of variance performed on data from 104 neurons tested with the baton stimulus set and 142 neurons tested with the hourglass stimulus set. In each case, top-part identity and bottom-part identity were the factors and firing rate was the dependent variable. **A**, Distribution of neurons with respect to main effects of top part, main effects of bottom part and interaction effects (each number is a percentage). **B**, Variance explained by interaction effects plotted against variance explained by main effects for neurons exhibiting at least one significant effect ($n = 95$). The average variance explained by main and interaction effects is indicated by the arrows. **C**, **D**, Analogous plots for the hourglass dataset.

tom parts) could arise. In the case of the baton set, because the parts are small relative to the object and are physically separate, any interaction effect would almost certainly arise from neuronal sensitivity to the combination of parts as such. In the case of the hourglass set, because the parts are large relative to the object and are contiguous, an interaction effect might also arise from neuronal sensitivity to a local feature (for instance a vertex) or a coarse shape (for instance a Z) created by the juxtaposition of two parts.

Do part influences combine linearly or nonlinearly?

The response of an IT neuron to a full set of baton stimuli is shown in Figure 1. This neuron clearly preferred some top parts over others (stimuli are arranged in order from those containing the least effective top part—row 1—to those containing the most effective top part—row 8). Likewise, it preferred some bottom parts over others (stimuli are arranged in order from those containing the most effective bottom part—column 1—to those containing the least effective bottom part—column 8). The most effective of the 64 stimuli was the baton containing the preferred top and the preferred bottom—as

expected if the influences of the parts summed linearly. However, this does not rule out the occurrence of subtle nonlinear effects. To characterize linear and nonlinear effects, we performed an ANOVA ($\alpha = 0.05$) with firing rate (in an interval from 50 to 350 ms) as the dependent variable and with the shape at the top (eight levels) and the shape at the bottom (eight levels) as factors. In this analysis, main and interaction effects represent respectively linear and nonlinear interactions between parts. The neuron of Figure 1 exhibited significant main effects of top part and bottom part ($p = 0$) as well as a significant interaction effect ($p = 0.004$). However, the main effects were preponderant, accounting for 77% of the explainable variance in the firing rate compared with 2% for the interaction effect.

The distribution across 104 neurons of significant effects elicited by baton stimuli is summarized in Figure 3A. A majority of neurons (91%) exhibited a main effect of top and/or bottom. A subset of these neurons (33%) also exhibited an interaction effect. To compare the strengths of main and interaction effects, we split the data into halves, and calculated how much of the variance in one half was explained by main or interaction effects measured in the other half. This procedure guarantees that variance explained in the case of pure noise will fluctuate around zero and thus provides a level playing field on which to compare main and interaction effects. Across neurons that exhibited at least one significant effect, either main or interaction ($n = 95$), main effects ac-

counted for a markedly larger fraction of variance (36% on average) than did interaction effect (2% on average). This pattern was highly consistent across neurons (Fig. 3B) and was statistically significant across data from both monkeys (paired t test, $p = 2.0 \times 10^{-26}$). It persisted when consideration was restricted to 32 neurons demonstrably sensitive to both top and bottom as indicated by the presence of two main effects and an interaction effect (main and interaction effects accounted for 51% and 6% of variance respectively; this difference was significant: paired t test, $p = 2.4 \times 10^{-14}$). Finally, it persisted and was statistically significant upon analysis of data from individual monkeys (variance accounted for by main and interaction effects: 36% and 2%, significantly different at $p = 1.6 \times 10^{-13}$ for 48 neurons in monkey Je; 36% and 3%, significantly different at $p = 3.5 \times 10^{-14}$ across 47 neurons in monkey Ec).

The response of an individual IT neuron to the full set of hourglass stimuli is shown in Figure 2. This neuron clearly preferred some top parts over others and some bottom parts over others. However, the most effective stimulus was not the one obtained by combining the best top part and best bottom part,

contrary to the expectation based on simple linear summation. An ANOVA performed on data from this neuron revealed significant main effects of top part and bottom part and a significant interaction effect ($p = 0$). The main effects accounted for 79% of the explainable variance in firing rate compared with 11% for the interaction effect.

The distribution across 142 neurons of significant effects elicited by hourglass stimuli is summarized in Figure 3C. A majority of neurons (92%) exhibited a main effect of top and/or bottom. A smaller but substantial fraction (54%) exhibited an interaction effect. To compare the strengths of main and interaction effects, we analyzed data from neurons that exhibited at least one significant effect ($n = 132$). Main effects accounted for a larger fraction of variance (22% on average) than interaction effects (7% on average). This pattern was reasonably consistent across neurons (Fig. 3D) and was statistically significant (paired t test, $p = 1.0 \times 10^{-18}$). It persisted when consideration was restricted to 64 neurons demonstrably sensitive to both top and bottom as indicated by the presence of two main effects and an interaction effect (main and interaction effects accounted for 32% and 11% of variance respectively and this difference was significant: paired t test, $p = 4.7 \times 10^{-12}$). Finally, it persisted and was statistically significant upon analysis of data from individual monkeys (variance accounted for by main and interaction effects: 22% and 8%, significantly different at $p = 4.5 \times 10^{-8}$ for 65 neurons in monkey Je; 22% and 6%, significantly different at $p = 1.5 \times 10^{-12}$ across 67 neurons in monkey Ec).

The results described above indicate that neuronal responses to stimuli in both sets were dominated by linear effects, reflecting independent sensitivity to the identity of the top part and the identity of the bottom part, in contrast to nonlinear effects reflecting sensitivity to specific combinations of parts. The hourglass set differed from the baton set in that interaction effects were more frequently significant (χ^2 test, $p = 0.0012$) and were stronger on average (t test on fraction of firing-rate variance explained, $p = 0.00007$). This pattern persisted and was statistically significant upon analysis of data from individual monkeys (percentage variance accounted for by interactions in baton and hourglass: 2% and 8%, significantly different at $p = 0.001$ for neurons from monkey Je; 3% and 6%, significantly different at $p = 0.02$ for neurons from monkey Ec). The greater incidence of interaction effects in hourglass stimuli may be attributable to cases in which neurons were sensitive to local features and coarse forms created by the juxtaposition of specific pairs of hourglass parts. That interaction effects were relatively weak even though they could have arisen by such a mechanism adds weight to the overall conclusion that linear effects predominate.

Figure 4. Interaction effects produced a moderate sharpening of neuronal stimulus selectivity. **A**, Baton set. Average normalized firing rate as a function of stimulus rank for observed data (thick line), data modified to remove all interaction effects (thin line) and data modified by shuffling interaction coefficients across stimuli (broken line). **B**, Baton set. Rank at half-height of each neuron's observed tuning curve plotted against rank at half-height of the tuning curve resulting from removal of interaction effects. **C**, **D**, Analogous plots for the hourglass set.

Do nonlinear interactions sharpen or broaden shape selectivity?

Interaction effects, although relatively weak, could still exert a systematic impact on neuronal stimulus selectivity. In particular, they could either sharpen or broaden the shape tuning curve. To explore this issue, we constructed a tuning curve for each neuron, plotting firing rate against stimulus rank—which ranged from 1 for the most effective baton to 64 for the least effective, with firing rate normalized to the range from the worst to the best baton. We performed this procedure both on the observed firing rates and on firing rates obtained by taking into account main effects alone. We found (Fig. 4A) that the average of the tuning curves actually observed (thick line) was narrower than the average of the tuning curves obtained after interaction effects had been removed (thin line) or shuffled (broken line). To assess the consistency of this outcome across neurons, we computed, for each neuron, the stimulus rank at half-height in the observed data and in the data with interaction effects removed (Fig. 4B). The tendency for the observed tuning curve to be narrower than the tuning curve without interaction effects was highly significant (observed mean = 18.2; interaction-absent mean 23.0; $p = 4 \times 10^{-6}$, pairwise t test).

Figure 5. Neuronal response nonlinearities arose from some mechanism other than passing the summed inputs through a sigmoidal activation function. **A**, Baton set. Observed firing rate versus firing rate predicted by main effects across all neurons and stimuli (104 neurons $\times$ 64 batons = 6656 points). White circles indicate average values for successive percentiles of the predicted firing rate in 2.5% steps. Deviations from the identity line indicate a nonlinear component in the response. Processing the linear sum through a sigmoidal activation function would produce deviations above the identity line at low firing rates and deviations above it at high firing rates. No such trend is present. There is simply a strong correlation between the predicted linear sum and the observed firing rate ($r = 0.79$ across all neurons). **B**, Baton set. Cumulative distribution across neurons of the probability that the sigmoidal model accounts for the data no better than the linear model. The fraction of neurons in which the sigmoidal model was significantly better (gray shaded region) was no greater than expected by chance. **C**, **D**, Analogous plots for the hourglass set.

We performed an identical analysis for the hourglass set (Fig. 4C,D). The observed tuning curves (Fig. 4C, thick line) were narrower than the tuning curves obtained after interaction effects had been removed (thin line) or shuffled (broken line). The sharpness of tuning, as measured by rank at half-height, was narrower for the observed response compared with the response predicted from main effects (Fig. 4D; observed mean = 17.6, interaction-absent mean = 27.0; $p = 1 \times 10^{-20}$, pairwise t test).

We conclude that nonlinear interactions, although comparatively rare and weak, are systematic in that they consistently sharpen the pattern of selectivity arising from main effects. The sharpening was greater for the hourglass set compared with the baton set (average difference in rank at half-heights: 9.4 for hourglass, 4.7 for baton, $p = 0.0004$, t test) as might be expected in light of the greater frequency and strength of hourglass interaction effects.

Do nonlinear interactions arise from a sigmoidal activation function?

Nonlinear interactions might arise from a process in which the influences of the top and bottom parts are first summed linearly

and then processed through a nonlinear sigmoidal activation function. This would give rise to a systematic relation between the efficacy of the parts and the nature of the nonlinear effect. For weakly effective parts, a threshold effect would occur: the response to the combination would be greater than predicted by linear summation. This would result in narrower tuning than for a linear activation function. For strongly effective parts, a saturation effect would occur: the response to the combination would be less than predicted by linear summation. This would result in broader tuning than for a linear activation function.

To search for such a pattern, we plotted the observed responses to the 64 batons against the responses predicted from main effects alone. We did this both using the raw firing rate (Fig. 5A) and using the firing rate normalized, for each neuron, to the response elicited by its best baton (data not shown). In the resulting plots, there was no hint of the pattern expected from passing the summed input through a sigmoidal activation function. The large white symbols, representing the running average of the surrounding cloud of data points, cleave to the identity line throughout each plot, whereas, according to the sigmoidal hypothesis, they should have been above it to the left and below it to the right of each graph.

To assess whether the responses of at least a few neurons could be explained by such a mechanism, we fit the responses of each neuron to two models: (1) a linear model in which the response is that predicted by the main effects (15 parameters) and (2) a sigmoidal model in which the output of the linear model is passed through a sigmoidal activation function (15 linear parameters + 3 sigmoidal parameters). Then, for each neuron, we performed a partial F test to determine whether the improvement in fit afforded by the sigmoidal model was significantly greater than the improvement expected simply from adding three extra degrees of freedom. The cumulative distribution of the probability of the null hypothesis (sigmoidal model no better than linear model) is shown in Figure 5B. The fraction of neurons (4%) in which the sigmoidal model yielded a significantly better fit ($p < 0.05$) was no greater than that expected from type I errors (5%).

We performed an identical analysis on the hourglass data (Fig. 5C,D). A plot of the observed firing rate against the firing rate predicted by main effects did not reveal any hint of the pattern expected from passing inputs through a sigmoidal activation function (Fig. 5C). The fraction of neurons significantly better fit by a sigmoidal model than by a linear model (9%; Fig. 5D) was not significantly in excess of the fraction expected from type I errors (χ^2 test, $p = 0.25$).

We conclude that the relatively weak nonlinear interactions between top and bottom parts observed in neuronal responses to both stimulus sets arose from some mechanism other than passing the summed influences of the two parts through a sigmoidal activation function.

Do sigmoidal nonlinearities affect the sharpness of tuning?

In the previous two sections, we showed that nonlinearities (1) sharpen neuronal tuning curves and (2) do not commonly conform to the pattern expected from passing inputs through a sigmoidal activation function. Here we consider whether these observations are related. Would sigmoidal nonlinearities, if present, have had a consistent impact on the sharpness of tuning? To explore this issue, we examined the responses of model neurons to pairs of simulated inputs (representing the influences of the top and bottom parts of an image) with strengths selected at random from normal distributions. We found that the transformation from a linear to a sigmoidal activation function can have no effect on the sharpness of tuning (supplemental Fig. S2A–C, available at www.jneurosci.org as supplemental material) can enhance it (supplemental Fig. S2D–F, available at www.jneurosci.org as supplemental material) or can reduce it (supplemental Fig. S2G–I, available at www.jneurosci.org as supplemental material). The outcome is dependent on the distribution of input strengths relative to the inflection point of the sigmoid. When the inputs are biased to be especially weak, thereby largely impinging upon the compressive portion of the sigmoid (supplemental Fig. S2D–F, available at www.jneurosci.org as supplemental material), sigmoidal activation tends to sharpen the tuning curve. On the other hand, when inputs are biased to be especially strong, thereby impinging upon the expansive portion of the sigmoid (supplemental Fig. S2G–I, available at www.jneurosci.org as supplemental material), sigmoidal activation tends to broaden the tuning curve. Thus the question whether nonlinearities systematically affect the sharpness of tuning is orthogonal to the question whether the nonlinearities are of the form expected from a sigmoidal activation function.

Did parts vary broadly in efficacy?

To be sure that we had selected parts that varied broadly in efficacy, as required for the preceding analysis, we ranked the top and bottom parts used in testing each neuron from best to worst. The ranking was based on the average of the firing rates elicited by the eight images that contained a given part minus the prestimulus baseline firing rate. For baton stimuli, the response to the best part was on average around three and a half times as strong as the response to the worst part (Fig. 6A,B). There was no significant difference between neuronal responses to the best top part and the best bottom part ($p = 0.82$, t test). Likewise neuronal selectivity for parts, as judged by the difference between the response to the best and worst parts, did not differ significantly between tops and bottoms ($p = 0.83$, t test).

Top and bottom parts in the hourglass set elicited weaker responses overall. This effect may have arisen because the top and bottom parts were fixed in the hourglass set whereas in the baton set, parts were selected to drive neuronal responses. Nonetheless, neurons were highly selective for part identity, as evidenced by the fact that the response to the best part was on average five times as strong as the response to the worst part (Fig. 6C,D). Neurons tested with the hourglass set responded slightly more strongly to the preferred bottom part than to the preferred top part ($p = 0.004$, t test) and were slightly more

Figure 6. The top and bottom parts selected for testing each neuron covered a broad range with regard to their efficacy in driving neuronal activity. **A**, Mean across neurons of the raw firing rate deviation from prestimulus baseline elicited by the top with the indicated rank in the baton set. **B**, Same for bottom parts. **C**, **D**, Analogous plots for the hourglass set.

selective for bottom identity than for top identity (best–worst response, $p = 0.0008$, t test).

Each stimulus set contained images in which the top or bottom was blank. In some batons, the top or bottom of the vertical bar was unadorned. In some hourglass stimuli, the top or bottom was altogether absent. To determine whether the response to the blank part was less than to other parts, we ranked the top and bottom parts for each neuron from 1 (strongest response) to 8 (weakest response). If the blank part tended to elicit weaker responses than the other parts, then the average value of its rank would be greater than 4.5. For the baton set, the average rank of the blank part (4.6) was slightly but not significantly (t test, $p = 0.42$) greater than 4.5. In the hourglass set, the average rank of the blank part (4.8) significantly exceeded 4.5 (t test, $p = 0.01$). Thus there was a common tendency for the blank part to elicit a weak response.

Does selectivity for part shape depend on part location?

Because we used the same set of parts (short of a 180° rotation) to form the tops and bottoms of each stimulus set, we were able to compare neuronal shape preferences at the two locations. Even if neurons exhibited similar patterns of selectivity for tops and bottoms, they still might produce more deeply modulated responses at one of the locations. To factor out this effect before carrying out a correlation analysis, we calculated the normalized response to each top part as $N_i = (R_i - R_{\text{mean}})/S$, where R_i is the average response to shapes with the i^{th} top, and R_{mean} is the average of all top responses, and S is the SD of the average responses to the eight tops. We calculated the normalized response to each bottom part in a similar fashion. To determine whether neurons had match-

ing shape preferences at top and bottom, we plotted, for all shapes in all neurons, the normalized response when the shape was at the top against the normalized response when the same shape was at the bottom (supplemental Fig. S3A, available at www.jneurosci.org as supplemental material). The values were positively ($r = 0.41$) and highly significantly ($p = 4.2 \times 10^{-35}$) correlated. Thus neurons exhibited reasonably consistent shape selectivity at the top and bottom locations.

The above result raises the possibility that neurons were sensitive simply to the presence of a shape and not to its location. To test this idea, we performed for each neuron a test comparing two models relating the firing rate to the display: a simple part model (with 9 parameters, see Materials and Methods) in which firing depended on which parts were present and a complex part $\times$ location model in which firing depended on which part was at which location (with 15 parameters, see Materials and Methods). We then performed a partial F test to determine whether the improvement in fit afforded by the part $\times$ location model was significantly greater than expected simply from its possessing six additional degrees of freedom. A cumulative distribution of the probability that the two models are equivalent is shown in supplemental Figure S3B (available at www.jneurosci.org as supplemental material). For baton stimuli, the part $\times$ location model was significantly better ($p < 0.05$) in 45% of neurons (shaded region in supplemental Fig. S3B, available at www.jneurosci.org as supplemental material). Thus around half of the neurons were sensitive to which part was at which location or alternatively and—in this experiment—indissociably to the orientations of the parts.

The same analysis performed on the hourglass data revealed a positive ($r = 0.35$) and significant ($p = 2.8 \times 10^{-31}$) correlation between selectivity for tops and bottoms (supplemental Fig. S3C, available at www.jneurosci.org as supplemental material). Approximately 28% of neurons were sensitive to which part was at which location in the hourglass set (supplemental Fig. S3D, available at www.jneurosci.org as supplemental material).

Do linear and nonlinear signals differ in timing?

To explore the timing of main and interaction effects, we measured their strength as a function of time following stimulus onset in each neuron and then averaged the results (see Materials and Methods). For the baton stimulus set, curves representing the strength of the population signal (Fig. 7A) make clear that the onset of the visual response was followed at a short delay by the onset of the main-effect signal and at a longer delay by the onset of the interaction-effect signal. We calculated the population latency of each effect by finding the first time bin in which the mean of the distribution of the effect strength across neurons was significantly different from zero (t test; $\alpha = 0.01$). The time following stimulus onset at which the average firing rate began to increase (75 ms) was indeed earlier than the main-effect latency (95 ms) and the interaction-effect latency (135 ms). This pattern persisted upon analysis of data from individual monkeys (latency of main and interaction effects: 85 and 115 ms for 54 neurons from monkey Je; 105 and 125 ms for 50 neurons from monkey Ec). These observations were qualitatively unaltered by restricting consideration to 31 neurons demonstrably sensitive to both top and bottom as indicated by the presence of two main effects and an interaction effect. To determine whether the temporal offset between the main-effect signal and the interaction-effect signal was statistically significant, we performed the following analysis. First, we normalized the time-varying signal of each neuron to the maximal average value of that signal 50–150 ms after stim-

Figure 7. Timing of main and interaction effect signals. **A**, Baton set. Average population signal strength of main effects (thick line) and interaction effects (thin line) as a function of time following stimulus onset (vertical scale at left). The average population firing rate (gray) is overlaid for comparison (vertical scale at right). Horizontal black bar indicates times at which the main and interaction effects differed significantly in strength (t test, $p < 0.05$). **B**, Analogous plot for the hourglass set.

ulus onset as computed across all tested neurons. This was necessary to avoid confounding differences in timing with differences in net strength of the two signals. Next, we assessed whether there was any 10 ms bin in which the distribution across neurons of the strength of the main-effect signal was significantly greater than the strength of the interaction-effect signal. There were two bins (centered at 95 and 105 ms) for which this was true ($p = 0.03$, t test). This analysis, when repeated on data from each monkey, yielded no significant time bins in the data from one monkey (Je) and one significant time bin (centered at 105 ms) in the data from the second monkey (Ec)—a pattern that is likely due to the lower numbers of neurons used for statistical comparisons. We conclude that the main-effect signal developed significantly earlier than the interaction-effect signal.

During testing with hourglass stimuli (Fig. 7B), the onset of the visual response and of the main-effect signal occurred simultaneously (85 ms) slightly earlier than the onset of the interaction-effect signal (95 ms). These observations were qualitatively unaltered by restricting consideration to 45 neurons sensitive to both top and bottom as indicated by the presence of two

main effects and an interaction effect. The onset times of the main-effect and interaction-effect signals were not significantly different ($p = 0.3$, t test). These results were consistent between data obtained from both monkeys (latency of main and interaction effects: 85 and 95 ms for 69 neurons from monkey Je; 95 and 95 ms for 73 neurons from monkey Ec; these differences failed to reach statistical significance).

The key conclusion of this analysis is that interaction-effect signals developed later than main-effect signals for both stimulus sets although the phenomenon was large and statistically significant only for batons. To assess the significance of the difference in timing between the two interaction-effect signals, we performed an analysis analogous to the one used above for comparing main and interaction effects. This revealed that the hourglass interaction-effect signal was significantly greater than the baton interaction-effect signal in a 10 ms bin centered at 95 ms after stimulus onset ($p = 0.04$, t test). We conclude that nonlinear effects developed earlier for hourglass than for baton stimuli.

Are responses to hourglass stimuli related to the responses to individual parts?

Because the hourglass set (unlike the baton set) consisted of individual top and bottom parts presented in isolation as well as whole objects constructed from pairs of parts, we were able to ask how the response to the whole object was related to the responses evoked by the individual parts in isolation. To answer this question we plotted the response to each whole object against the average of the responses evoked by the two parts (supplemental Fig. S3, available at www.jneurosci.org as supplemental material). According to the principle of divisive normalization, the two values should be equal (Reynolds et al., 1999). There was a strong and significant positive correlation. The response to the whole object was slightly but significantly greater than the average of the individual part responses (mean response to whole = 11.2 spikes/s, mean of the average of the responses to the two parts = 10.7 spikes/s, $p = 1 \times 10^{-13}$, paired t test). This enhancement—a slight net deviation from divisive normalization—was additive (the intercept of the best-fit line was 1.11) rather than multiplicative (the slope of the best-fit line was 0.94). We conclude that neuronal responses to hourglass shapes are roughly consistent with the pattern expected from divisive normalization.

Discussion

The aim of this study was to characterize how the influences of parts within an object combine in driving the visual responses of IT neurons. Three features of the approach were critical. First, in common with a few previous studies (Baker et al., 2002; De Baene et al., 2007), we made use of baton stimuli containing widely separated discrete parts that could be independently manipulated without risk of creating juxtapositional features. Second, in departure from previous studies, we used numerous parts covering a wide range with regard to their individual efficacy in eliciting neuronal activity. This was required for testing a model in which nonlinear interactions arise from passing the summed inputs through a sigmoidal activation function. Third, we compared results obtained with the baton stimuli to results obtained with a stimulus set in which parts were confluent, with the consequence that apparent nonlinearities could arise from neuronal sensitivity to attributes created where two parts abut. We discuss the main conclusions below, first with reference to baton stimuli and then to the hourglass stimuli.

The influences of discrete parts combine approximately linearly

This is the key finding of the study. It stands in sharp contrast to the idea that IT neurons are nonlinearly selective for particular conjunctions of features (Perrett et al., 1982; Desimone et al., 1984; Tanaka et al., 1991; Kobatake and Tanaka, 1994; Ito et al., 1995; Yamane et al., 2006). However, it is concordant with the finding that neuronal responses to face parts sum linearly with few nonlinear interactions (Freiwald et al., 2009). It also resonates with the observation that responses to multiple objects are governed by divisive normalization. In visual areas including IT (Zoccolan et al., 2005), MT (Britten and Heuer, 1999; Majaj et al., 2007), V4 (Reynolds et al., 1999; Ghose and Maunsell, 2008), V2 (Reynolds et al., 1999) and V1 (Heeger, 1993; Carandini et al., 1997), the response to a multielement display is approximately equal to the average of the responses elicited by the individual elements. This raises the question: could the monkeys have been processing the baton stimuli as multiple objects and could this account for our observations? We think not for two reasons. First—although this hardly constitutes objective evidence—the batons seem compellingly unitary to a human observer. Second, it has been shown that requiring monkeys to process batons as wholes—which might be thought to enhance perceiving them as unitary objects—does not affect neuronal responses to them in IT, which remain the same as during passive viewing (Baker et al., 2002). This is consonant with observations indicating that discrimination training affects neuronal responses to stimuli (Sigala and Logothetis, 2002) but active performance of the discrimination does not (Suzuki et al., 2006). We conclude that an essentially identical linear rule determines responses to displays consisting of multiple objects and objects consisting of multiple parts.

Nonlinear interactions between discrete parts are rare and weak

We found that statistically significant main effects of top and bottom identity were 2.8 times more common than statistically significant interaction effects. This is in good agreement with the results of a few other studies using stimuli with discrete features [Baker et al. (2002) report a ratio of 3.2; De Baene et al. (2007) describe linear summation as the dominant pattern; McMahon and Olson (2009) report a ratio of 6.0]. The minor nature of nonlinear effects was even more evident in a measure based on signal strength, which revealed a difference of an order of magnitude between main effects based on individual part identity and interaction effects based on the combination of parts (Figs. 3B, 7A). The measure of signal strength that we used is worth particular note because it eliminates contributions from noise that have tended in previous studies (McMahon and Olson, 2009) to inflate the relative strength of interaction effects.

Nonlinear interactions between discrete parts develop late

The finding that interactions between spatially separate parts of an object develop ~ 40 ms later than main effects (Fig. 7A) is concordant with an earlier finding obtained using shapes constructed from contour fragments (Brincat and Connor, 2006) but is at odds with a recent report indicating that nonlinear interactions between color and shape develop simultaneously with main effects (McMahon and Olson, 2009). Perhaps nonlinear interactions between spatially separate elements (discrete parts of an object) involve a time-consuming process whereas interactions between spatially coextensive elements (a shape and its color) do not.

Nonlinear interactions between discrete parts tend to sharpen image selectivity

Nonlinear interactions, when present, did induce a slight increase of the tendency toward conjunction selectivity. This finding supports earlier observations that nonlinear effects, although rare and weak, do tend systematically to enhance neuronal selectivity, rendering the representation of images in IT sparser than it otherwise would be (Baker et al., 2002; McMahon and Olson, 2009).

Nonlinear interactions between discrete parts result from a mechanism other than a sigmoidal activation function

Despite our best attempts to drive neurons across a wide dynamic range, we did not observe any threshold effects or response saturation. If nonlinearities are not the consequence of passing summed inputs through a sigmoidal activation function, then how do they arise? Synaptic inputs are thought to sum linearly when they impinge on separate dendritic branches but to sum nonlinearly when they impinge on the same branch (London and Häusser, 2005). Perhaps nonlinear interactions occur under the relatively rare condition when synapses conveying the influences of discrete parts of an object converge on the same dendritic branch. This alone cannot, however, account for the fact that interaction effects are delayed by several tens of ms relative to main effects.

Nonlinear interactions are more common between confluent parts than between discrete parts

Results obtained with hourglass stimuli differed from those obtained with baton stimuli in one key respect: nonlinear effects reflecting neuronal sensitivity to the particular combination of top part and bottom part were more prominent. This was evident in counts of statistically significant interaction effects (Fig. 3*A* vs *C*), in measures of interaction-effect strength in individual neurons (Fig. 3*B* vs *D*) and in the strength of the population interaction-effect signal (Fig. 7*A* vs *B*). Hourglass stimuli also differed from baton stimuli with regard to two aspects of response timing. (1) The onset of main effects reflecting sensitivity to part identity was simultaneous with the onset of the visual response for hourglass stimuli whereas it was delayed for baton stimuli. This may have been due to the fact that features differentiating one part from another were defined on a relatively coarse scale for hourglass stimuli and on a relatively fine scale for baton stimuli. Information about image identity is represented in a coarse-to-fine sequence in IT (Sripati and Olson, 2009). (2) The interaction-effect signal developed earlier for hourglass than for baton stimuli (Fig. 7*A* vs *B*). This may have been due to the fact that some interaction effects reflected sensitivity to a local feature created by the juxtaposition of two parts rather than to the combination of parts as such. Signals representing a particular combination of distant parts appear to develop more slowly than signals reflecting the identity of an individual part (Brincat and Connor, 2006).

How can we explain classic cases of conjunction selectivity?

The classic view that IT neurons are selective for combinations of features as distinct from individual features (Perrett et al., 1982; Desimone et al., 1984; Tanaka et al., 1991) is founded on the fact that removing any part from a preferred image can cause a drastic reduction in the response (Kobatake and Tanaka, 1994; Ito et al., 1995). However, removing either part from a preferred baton does not commonly cause a dramatic reduction. There must be a difference between this manipulation and the manipulations used in the classic studies. What might the difference be? We can

offer two suggestions. (1) Juxtapositional features. If the parts of an object are not physically separate then combining them might give rise to a new local feature where they adjoin (such as the X junction at the center of an hourglass image formed from a top triangle and a bottom triangle). A neuron selective for the X junction would give the appearance of being selective for the combination of two triangles. This would not have happened with baton stimuli because of the physical separation between the top and bottom parts. (2) Coarse footprint. The parts of an object might when combined create a coarse shape not implicit in either part alone (such as the Z of an hourglass image formed from two chevrons). We have recently found that IT neurons are selective for global shape up to a level of detail of around three cycles per object (Sripati and Olson, 2010). A neuron selective for a coarse footprint in the form of a Z would give the appearance of being selective for the conjunction of the two chevrons. This would not have happened with baton stimuli because the details differentiating one part from another are on such a fine scale as not appreciably to affect the coarse footprint. We surmise that IT neurons exhibit robust conjunction selectivity only when the conjunction of parts creates a preferred local juxtapositional feature or a preferred global coarse footprint. This view is consonant with the finding that nonlinear effects occur at a significantly higher rate for hourglass than for baton stimuli.

References

- Baker CI, Behrmann M, Olson CR (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* 5:1210–1216.
- 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.
- Britten KH, Heuer HW (1999) Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19:5074–5084.
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644.
- De Baene W, Premereur E, Vogels R (2007) Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *J Neurophysiol* 97:2900–2916.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062.
- Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187–1196.
- Ghose GM, Maunsell JH (2008) Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. *J Neurosci* 28:5115–5126.
- Heeger DJ (1993) Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J Neurophysiol* 70:1885–1898.
- Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73:218–226.
- 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.
- Koch C (1999) *Biophysics of computation*. Cambridge, MA: MIT.
- London M, Häusser M (2005) Dendritic computation. *Annu Rev Neurosci* 28:503–532.
- Majaj NJ, Carandini M, Movshon JA (2007) Motion integration by neurons in macaque MT is local, not global. *J Neurosci* 27:366–370.
- McMahon DB, Olson CR (2009) Linearly additive shape and color signals in monkey inferotemporal cortex. *J Neurophysiol* 101:1867–1875.
- Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47:329–342.
- Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753.

- Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in primate temporal cortex. *Nature* 415:318–320.
- Sripati AP, Olson CR (2009) Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. *J Neurosci* 29:7788–7796.
- Sripati AP, Olson CR (2010) Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency. *J Neurosci* 30:1258–1269.
- Suzuki W, Matsumoto K, Tanaka K (2006) Neuronal responses to object images in the macaque inferotemporal cortex at different stimulus discrimination levels. *J Neurosci* 26:10524–10535.
- Tanaka K, Saito H, Fukada Y, Morioka M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
- Wolfe JM, Bennett SC (1997) Preattentive object files: shapeless bundles of basic features. *Vision Res* 37:25–43.
- Yamane Y, Tsunoda K, Matsumoto M, Phillips AN, Tanifuji M (2006) Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J Neurophysiol* 96:3147–3156.
- 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.
- Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25:8150–8164.