REVIEW

THE NEURAL BASIS OF IMAGE SEGMENTATION IN THE PRIMATE BRAIN

A. PASUPATHY *

Department of Biological Structure, Washington National Primate Research Center, University of Washington, Seattle, WA 98195, USA

Abstract—Image segmentation is a fundamental aspect of vision and a critical part of scene understanding. Our visual system rapidly and effortlessly segments scenes into component objects but the underlying neural basis is unknown. We studied single neurons in area V4 while monkeys discriminated partially occluded shapes. We found that many neurons tuned to boundary curvature maintained their shape selectivity over a large range of occlusion levels as compared to neurons that are not tuned to boundary curvature. This lends support to the hypothesis that segmentation in the face of occlusion may be solved by contour grouping.

This article is part of a Special Issue entitled: Contributions From Different Model Organisms to Brain Research. © 2014 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: object recognition, shape representation, monkey, ventral pathway.

Contents

INTRODUCTION

The visual world that reaches our eyes is encoded as local contrast values in the activity patterns of retinal ganglion cells. This representation is isomorphic to the visual stimulus and continuous in that there are no demarcations for where one object ends and another begins. We nevertheless perceive the world not as a uniform pixelated representation, but as a meaningful arrangement of objects and regions. This is achieved by a process called image segmentation which takes as its input the continuous retinal representation and parses it into components that ultimately underlie the percept that is the brain's best guess for the current state of the outside world. Image segmentation facilitates scene understanding and makes our interactions with the world around us more effective. It has been shown to improve stimulus discrimination (Croner and Albright, 1999) and provides structure for deploying visual attention (Qiu et al., 2007). While we understand a great deal about how isolated stimuli are encoded in various stages of the visual processing hierarchy, very little is known about how, where, and when images are parsed into components. How scenes are segmented is one of the most important unanswered questions in vision and discovering the underlying principles will constitute a major advance in the field and could lead to better artificial vision systems. Furthermore, while it is universally accepted that feedback and recurrent processes contribute to complex brain function, the underlying mechanisms and circuitry in the visual cortex are largely unknown. In fact, there are essentially no examples of neurophysiological manipulations that can be used to control cortical feedback with the precision with which feedforward signals, driven from sensory input, can be manipulated and used to modulate neuronal responses. Because image segmentation is thought to engage feedback and recurrent processes (Kosai et al., 2014), it provides a relatively untapped opportunity to understand and manipulate cortical feedback, possibly by changing stimulus and task conditions. This could have major implications for a deeper understanding of cortical processing in general.

The approach

Segmentation is computationally challenging—even the most cutting edge machine vision systems are unable to replicate the segmentation abilities of the human visual system. To understand the neural basis of segmentation, it would be tempting to try to decode the visual cortical representations of a wide-variety of stimuli with extensive clutter and occlusions, stimulus characteristics that make segmentation a hard problem. But currently, this turns out to be an impractical strategy because the space of complex images is too large, the

---

*Address: Department of Biological Structure, University of Washington, 1959 N.E. Pacific Street, Box 357420, HSB G520, Seattle, WA 98195, USA. Tel: +1-206-685-0768.
E-mail address: pasupat@uw.edu
Abbreviations: RF, receptive field; ROC, receiver operating characteristic.

time available to record any given neuron in the lab is limited to brief periods (few hours) due to experimental constraints, and neuronal responses of most visual cortical neurons are nonlinear functions of visual stimuli, and we do not have a good understanding of the underlying nonlinearities or the bases of representation. These constraints make it extremely difficult, if not impossible, to analytically evaluate the neuronal dynamics associated with segmentation on the basis of responses to an arbitrary set of stimuli. A more fruitful approach, in our experience, has been one of targeted hypothesis testing: we identify plausible hypotheses based on shape theory and human psychophysical literature and then focus on designing well-balanced, customized stimuli that can directly address those hypotheses. In this case, the stimulus design targets a localized region of shape space relevant to the hypotheses being tested and facilitates systematic and controlled tests that can reveal the underlying nonlinearities and representational bases. Below, we review our recent experiments (Kosai et al., 2014) to test one longstanding psychophysical hypothesis that image segmentation and subsequent recognition of partially occluded objects are achieved by contour grouping (Wertheimer, 1938).

Contour-based segmentation and primate V4

Gestalt psychologists have hypothesized that visual scenes are perceptually grouped into objects and that the component objects are detected and recognized by first grouping contours based on principles of similarity, proximity, continuity, common fate, symmetry, convexity, etc. (Wertheimer, 1938, see Wagemans et al., 2012, for review). This strategy of applying Gestalt principles to contours has been a popular tool for segmentation in computer vision (Leung and Malik, 1998). This stands in contrast to region-based segmentation, where the image is partitioned into pixel sets with coherent image properties such as brightness, color and texture (Leung and Malik, 1998)—an approach more commonly used in traditional computer vision algorithms. Depending on the specific task design, psychophysical studies lend support to contour-based strategies (Jolicoeur et al., 1986; Ben-Av et al., 1992; Houtkamp et al., 2003), region-based strategies (Fine et al., 2003) or a combination (Mumford et al., 1987).

One possible locus for contour-based segmentation in the primate brain is area V4, an intermediate stage in the ventral (i.e., form processing) pathway, where many neurons encode shape in terms of their boundary characteristics (Pasupathy and Connor, 2001). For example, a V4 neuron may respond strongly to shapes that include a sharp convexity to the lower right and weakly to shapes that do not (Fig. 1). A second neuron may respond preferentially to a set of shapes that include a concavity to the left. We have shown that a population of such neurons can provide a complete and accurate representation of two-dimensional shapes on the basis of their boundary characteristics (Pasupathy and Connor, 2002). These curvature-tuned neurons would be an ideal neural substrate for contour-based segmentation; but, because most shape tuning characterizations are conducted with isolated stimuli, we do not know whether or how these neurons contribute to segmentation. We therefore studied the responses of curvature-tuned V4 neurons as animals discriminated partially occluded shapes to determine how they might contribute to the segmentation of occluded objects.

Non-human primate model

To understand the neural basis of image segmentation, we conducted single unit studies in macaque monkeys as they performed a shape discrimination task. Our choice of animal model is informed by several factors. First, macaque monkeys are highly visual animals. Their lives in their natural habitat suggest high visual acuity and hand–eye coordination. Their visual system is comparable to that of humans in terms of visual acuity (Cavonius and Robbins, 1973) and in the manner in which they explore their environment. Monkeys and humans can easily discriminate complex images and objects that are only 2° in diameter at central fixation (e.g., Asaad et al., 1998). Monkeys are very similar to humans in their exploration of high-interest targets in scenes (Berg et al., 2009). Voluntary eye movements are qualitatively similar in humans and monkeys (Fuchs, 1967); monkeys like humans, have coordinated eye movements important for maintaining stereopsis (Schor and Tyler, 1981). Several behavioral studies in monkeys suggest that they segment visual scenes into objects and regions the way humans do (Munakata et al., 2001). Theories of segmentation, based on human psychophysics are consistent with neurophysiological studies in monkeys. Specifically, shape theory and human psychophysics suggest that T-junctions are highly informative about occlusion and that segmentation of occluded objects may originate at T-junctions
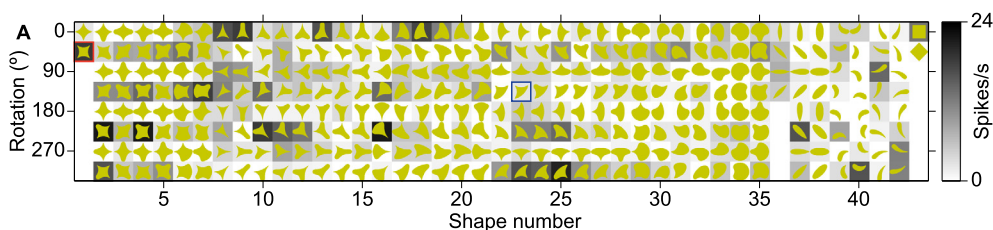

Fig. 1. Responses of a V4 neuron tuned to boundary curvature. Shape preference was characterized using a set of 43 shapes (columns) presented at 8 rotations (rows) in a passive fixation task. Some shapes (1, 36 and 43) were shown at fewer rotations due to rotational symmetry. The background intensity of each icon depicts the average response to that shape. Responses were strongest for shapes containing a sharp convexity to the lower right. Shapes highlighted by red (preferred) and blue (non-preferred) squares were chosen as the discrimination stimuli for the behavioral task (see Fig. 4). Previously published in Kosai et al. (2014).
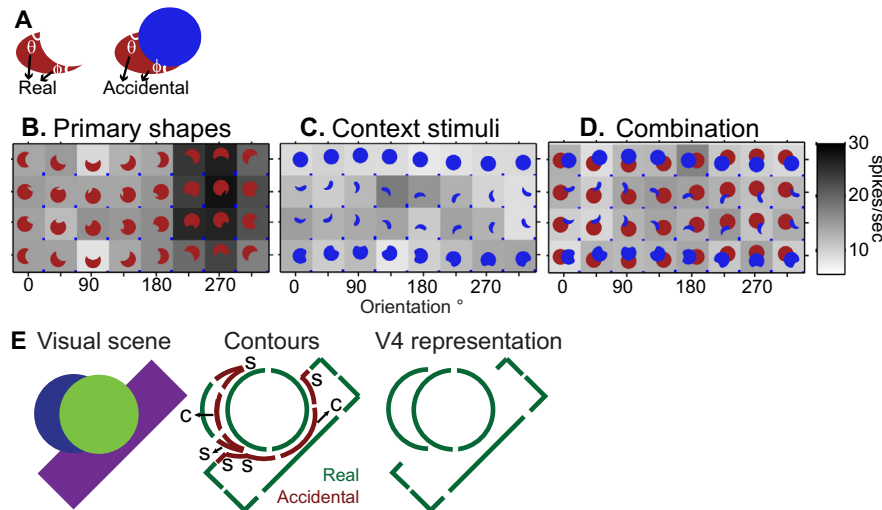
Fig. 2. Suppression of preferred responses under partial occlusion context. (A) Angles θ and φ are real contours for the crescent in isolation (left); when the crescent is adjoined by a contextual stimulus (right), these angles are interpreted as accidental contour features formed at the T-junctions between the occluding (blue) and occluded (red) shapes and are perceptually less salient. (B–D) Example V4 neuron that exhibits suppressed encoding of accidental contours. Average responses of an example neuron to: four primary shapes (B), context stimuli (C) and combination stimuli (D) presented at 8 orientations (columns) are shown in grayscale. Blue bars in the lower right corner of each icon indicate standard errors of the mean (SEM). (B) Primary shapes with a sharp convexity at the bottom of the shape (225–315°) evoked strong responses from this cell. (C) Contextual stimuli presented in the non-preferred color evoked weak responses. (D) Preferred primary shape responses (B: 225–315°) were strongly suppressed in the presence of corresponding contextual stimuli. (E) Schematic of how a visual scene is encoded in area V4. The left panel shows an example visual scene with partially occluded objects. All boundaries in the image are shown in the middle panel. Real contours are shown in green. Accidental sharp convexities at T-junctions (labeled *s*) and accidental concavities between the T-junctions (labeled *c*) are shown in red. In area V4 only the real contours (green) are strongly encoded. This may serve as the first step of segmentation in the primate brain. Adapted from Bushnell et al. (2011).

(Helmholtz, 1909; Guzman, 1968; Huffman, 1971; Clowes, 1971; Waltz, 1975; Elder and Zucker, 1998; Rubin, 2001). Consistent with this, we demonstrated that the encoding of accidental contours that are formed at the junction between occluded and occluding objects is suppressed in area V4 (Fig. 2B–D; Bushnell et al., 2011). This suppression emerges soon after stimulus onset suggesting that this may be the first step of segmentation; this supports the hypothesis that segmentation of occluded objects may begin at T-junctions with the suppression of accidental contours (Fig 2E).

Decades of experiments in monkeys suggest that their visual cortex is similar to humans in terms of fractional extent, in terms of its organization into dorsal and ventral streams, and in terms of the receptive field (RF) size and tuning properties for many sub-regions of the visual cortex (van Essen, 2004; Orban et al., 2004). The anatomy is also strikingly similar from the structure of the retina, to the lateral geniculate nucleus, to the layering and organization of the primary visual cortex. Lesions of different portions of the ventral visual pathway, important for form vision, produce similar deficits in humans and monkeys (for example, cf Gross, 1973; Merigan, 1996; Gallant et al., 2000; Farah, 1990; Mishkin and Pribram, 1954). It is no surprise then, that much of what we know today about the neural basis of object recognition—for example, how sensory signals are encoded along the ventral visual pathway, how these representations change as a function of learning and experience, and how these signals underlie behavior—come from investigations in the monkey (for review, see Kourtzi and Connor, 2011).

With regard to experimental methods, a wide variety of techniques have been successfully implemented in the monkey to address systems-level questions. The monkey has been a remarkably successful preparation in both anesthetized and awake preparations, the former because of stable maintenance under anesthesia for several days and anesthetic regimens that retain robust neuronal responses in many brain areas, and the latter because of robust tolerance of skull implants for years. While single and multi-electrode extracellular recordings have been the most widely used experimental technique, imaging methods, including optical imaging and functional magnetic resonance imaging (fMRI) have been successfully implemented and effectively used to address questions at larger spatial scales. 2-photon calcium imaging, a method well-established in smaller animals and one that allows the visualization and physiological characterization of many neurons simultaneously, and neuronal sub-compartments, has more recently been successfully implemented in the anesthetized monkey (e.g., Nauhaus et al., 2012) and is currently being adapted for the awake preparation. To discover causal links between activity and behavior, pharmacological, electrical and thermal perturbation techniques have been extensively developed and applied, while primate optogenetics is currently under development.

Behavioral task

Our goal was to design a behavioral task that engages segmentation. We chose a shape discrimination task
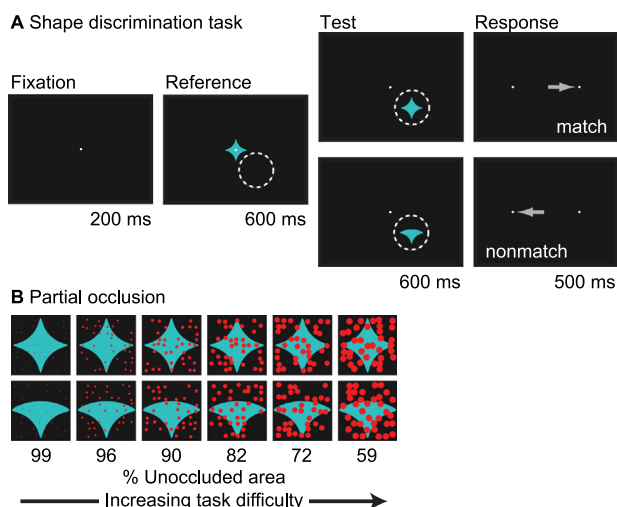
Fig. 3. Behavioral task. (A) Sequence and duration of trial events. After acquiring fixation, monkeys viewed two shapes: the reference stimulus at the center of gaze followed by the test stimulus in the neuron's RF. The fixation point was then extinguished and two peripheral choice targets appeared. Animals reported whether the reference and test stimuli were the same or different (match/nonmatch) with rightward and leftward saccades, respectively. New discrimination stimuli were chosen each day based on the shape preferences of the neuron recorded (a preferred and non-preferred shape). (B) The test stimulus was partially occluded by a field of 36 dots randomly positioned in a 9×9 square grid within the neuron's RF. Task difficulty was titrated by varying the diameter of occluding dots; occlusion level was parameterized as the percentage of the RF area that remained unoccluded (% unoccluded area). Previously published in Kosai et al. (2014).

under partial occlusion, since occlusions are a major challenge for segmentation. We trained animals to report whether two shapes, presented in sequence were the same or different (Fig. 3). The first, called the reference, was presented in isolation at center of gaze, while the second, called test, was superimposed by a set of occluding dots and was presented in the center of the RF of the cell under study. Animals could perform this task by first segmenting the test shape from the surrounding occluders to determine whether it was the same or different from the reference shape; alternatively, they could simply compare the visible portion of the test shape with that of the reference without actively segmenting the former, i.e. performing a partial template match. This captures the situation in natural vision where partially occluded objects can be recognized either by first segmenting them or just on the basis of partial visible information. In fact, there has been a longstanding debate about whether object recognition facilitates segmentation or vice versa. Regardless of how the animals solved the task, and whether segmentation precedes or follows recognition, our goal was to evaluate whether V4 responses reflect the segmented image and our results described below suggest that they do.

Each day, we isolated a single neuron and characterized its RF location, color preferences and shape tuning properties. Specifically, V4 responses to shape stimuli may be dictated by the surface-based characteristics, e.g. overall stimulus orientation and

spatial frequency content, or the contour-based characteristics captured by tuning for boundary curvature. Neurons tuned to either attribute can be shape-selective, i.e. respond strongly to some stimuli but not others, but only the latter group of neurons may be suitable to underlie contour-based segmentation. So, for each recorded neuron we first determined whether the neuron's responses were dictated by the boundary characteristics of shape stimuli. This identified the subgroup of neurons that could contribute to contour-based segmentation. For all shape-selective neurons, both for neurons deemed to be curvature-tuned and for others, we chose two discrimination stimuli, one that evoked strong responses and another that evoked weak responses from the cell in question. As the animal performed the behavioral task with the chosen stimuli, we asked how neuronal responses were modulated by occlusion level to infer how the different subgroups of neurons could contribute to the segmentation of objects in the face of occlusion.

Results

For each neuron that we studied, we first carried out an initial screen of a large and systematic set of shapes. The results of this screen for an example curvature-tuned neuron are shown in Fig. 1. This neuron responded preferentially to shapes with a sharp convex projection to the lower right and poorly to shapes without this feature; all of the shapes that evoked a strong response from this cell included this preferred feature. We have previously shown that neurons selective for contour features can be well described by a simple descriptive model: a two-dimensional Gaussian function in a shape space defined by curvature × angular position (Pasupathy and Connor, 2001). Here curvature ranges from −1 (concavities) to +1 (convexities) and angular position ranges from 0° (right of center) to 360° in a counter-clockwise direction. A 2-D Gaussian function with a peak at curvature +1.0 (sharp convex) and angular position 322° (down and to the right), consistent with our subjective interpretation of the responses, provided a good fit to the observed data: coefficient of correlation between observed and predicted responses was 0.7. On the basis of this initial shape screening, we chose two discrimination stimuli, one with the identified critical feature (sharp convex to the lower right), and another without, to be used in the subsequent occlusion experiments.

As the animal performed the shape discrimination task with the chosen discrimination stimuli, we continued to study neuronal responses. Fig. 4A shows PSTHs when the preferred shape was presented as the test stimulus within the RF either in isolation (black line) or superimposed by occluding dots during the behavioral task. Overall, responses were strongest when the shape was unoccluded, and declined gradually with increasing levels of occlusion. Responses to the nonpreferred stimuli increased slightly with occlusion (Fig. 4B, compare black line with other colors). To ask how the change in responses as a function of occlusion compares to change in animal behavioral performance, we constructed psychometric and neurometric curves
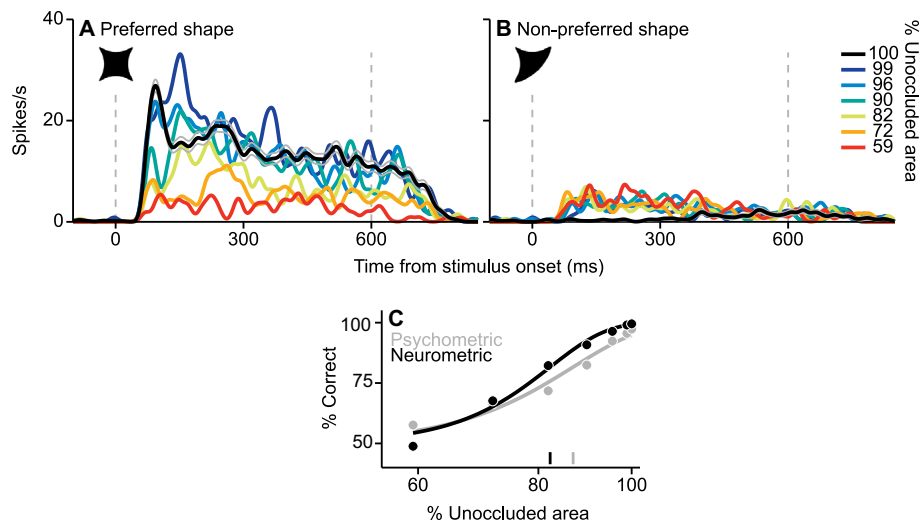
Fig. 4. Results from an example curvature-tuned neuron during behavior. Same neuron as in Fig. 1. (A, B) Response PSTHs ($\sigma = 10$ ms) for the preferred (A) and non-preferred shapes (B) at different occlusion levels (colored lines) when presented as test stimuli. Responses to the preferred shape were strong when it was unoccluded (black; thin lines show SEM) and decreased with increasing occlusion level; the opposite occurred for responses to the non-preferred shape. (C) Comparison of behavioral (gray) and neuronal (black) performance across occlusion level. Symbols indicate % correct performance at each occlusion level; lines are descriptive fits to the data. Neurometric curves were constructed based on responses in 50–350 ms counting window from test stimulus onset. Tick marks along the abscissa mark neurometric and psychometric thresholds (black and gray, respectively). Previously published in Kosai et al. (2014).

for each session (Britten et al., 1992). To characterize behavioral performance, we calculated the proportion of correct behavioral responses at each occlusion level and then fit a cumulative Weibull distribution function to these psychometric data using a least squares method and extracted estimates of the psychometric threshold, defined as the level of occlusion corresponding to 82% correct performance (Britten et al., 1992). For the neurometric curve, we first counted spikes in the window 50–350 ms after stimulus onset; the lower cutoff of 50 ms was chosen to account for visual response latency in V4. Then, for each occlusion level, we quantified neurometric performance by calculating the area under the receiver operating characteristic (ROC) curve derived from the spike count distributions for preferred and non-preferred stimuli. To this neurometric curve we fit a cumulative Weibull distribution function and extracted estimates of the neurometric threshold. Fig. 4C shows psychometric (gray) and neurometric (black) performance plotted against the percentage of the test stimulus area that was unoccluded by dots. For unoccluded stimuli (100% unoccluded area), psychometric and neurometric performance was almost perfect (97% and 99% correct, respectively). As occlusion level increased, both showed a very similar declining trend, approaching chance levels at the highest occlusion level tested. As a result, psychometric and neurometric thresholds were similar (tick marks along the abscissa; 87% and 84%, respectively) and threshold ratio was just below 1 (0.96).

A contrasting example of a cell that was not curvature-tuned is shown in Fig. 5. This neuron's responses were also strongly shape selective; only a few stimuli evoked strong responses whereas most evoked weak or no responses (Fig. 5A). However, unlike in the previous example, it was difficult to explain this neuron's responses in terms of selectivity for specific contour features. The best-fitting Gaussian function had a peak curvature $+0.22$ (shallow convex) and angular position $105°$ (up and to the left), but this model failed to explain the data well; correlation between observed and predicted data was 0.23. This could, in principle, arise from high variability in the neuron's responses rather than a true lack of tuning for boundary curvature but we confirmed that this was not the case (Kosai et al., 2014). We, nevertheless, chose two discrimination stimuli (see red and blue squares) that evoked very different responses when unoccluded (compare black lines in Fig. 5B, C). Psychophysical performance (Fig. 5D, gray) for unoccluded shapes was high and declined gradually with increasing occlusion: the psychometric threshold was 68% indicating better performance than the previous example. Unlike in that example, however, this neuron's responses to the preferred stimulus declined rapidly even at the weakest occlusion levels (Fig. 5B, compare responses at 100% and 90% unoccluded area). As a result, neurometric performance declined rapidly with increasing occlusion (Fig. 5D) and resulted in a high neurometric threshold (90%) and threshold ratio (1.32), reflecting the poor neuronal sensitivity to shape information under occlusion. Thus, this neuron was similar to the previous in terms of its strength of selectivity for unoccluded stimuli but was strikingly different in terms of its susceptibility to partial occlusion.

It is possible that differences in rate of decline of preferred responses in Figs. 4 and 5 are simply due to differential modulation by the nonpreferred color of the occluding dots: the neuron in Fig. 5 may be more strongly suppressed by the nonpreferred color, thus resulting in a higher neurometric threshold. This appears not to be the case because responses to nonpreferred stimuli increase with occlusion for both neurons, suggesting that the nonpreferred color was not universally suppressive.
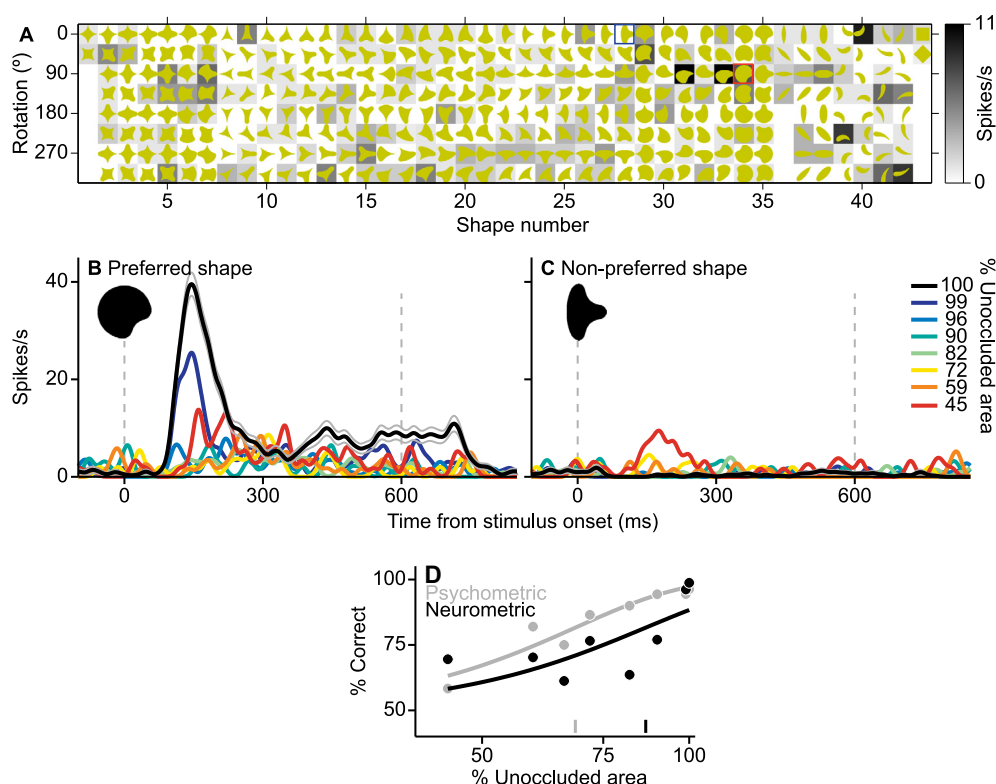
Fig. 5. Results from a contrasting non-curvature-tuned neuron. (A) Same format as Fig. 1. The neuron responded strongly to a few shapes, but its shape preferences were hard to describe in terms of local contour features. (B, C) Responses to the preferred shape were strong when it was unoccluded and decreased rapidly with increasing occlusion level; responses to occluded shapes were weak, even for low occlusion levels (compare responses at 99 and 96% unoccluded area). Responses to the non-preferred shape were generally weak. (D) Psychometric and neurometric curves were largely unmatched; behavioral performance was superior at most occlusion levels. Previously published in Kosai et al. (2014).
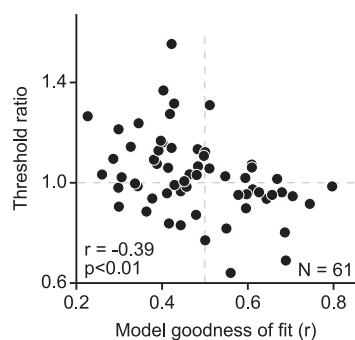

Fig. 6. Relationship between model goodness of fit and threshold ratio. (A) Threshold ratios versus the curvature model's goodness of fit. For each neuron, we identified (using nonlinear least squares methods) the 2-D Gaussian function in a shape space defined by angular position and boundary curvature, that best predicted neuronal responses to shape stimuli during passive fixation; the correlation between observed and predicted responses provided a measure of the goodness of fit. Neurons that were best fit by the model (curvature-tuned) had the lowest threshold ratios. Previously published in Kosai et al. (2014).

The trend exemplified by the representative neurons in Figs. 4 and 5—greater sensitivity of curvature-tuned neurons for shape information under occlusion—held across our population of 61 neurons in two monkeys. We found a significant negative correlation between the Fisher r-to-Z-transformed values of goodness of fit of the curvature model and the threshold ratios (Fig. 6,

$r = -0.39$, $p < 0.005$). In other words, neurons better fit by the curvature model were associated with lower threshold ratios. Among neurons that were well-described by the curvature model (goodness of fit $\geqslant 0.5$; $N = 24/61$), neurons typically had threshold ratios near or below 1, indicating that they were just as sensitive, or more sensitive than behavior. In contrast, among neurons that were poorly described by the curvature model (goodness of fit < 0.5; $N = 37/61$), many had threshold ratios above 1, indicating that they were less sensitive than behavior. The lower threshold ratios of curvature-tuned neurons resulted largely because their responses yielded neurometric thresholds (median = 86%) that were significantly lower (T-test, $p < 0.01$) than for the non-curvature-tuned neurons (median = 94%). We also verified that these differences cannot be attributed to difference in shape selectivity, peak firing rates or response variability (Kosai et al., 2014). One final possibility is that our choice of the discriminants engineered this effect: we may have chosen shapes that were hard to discriminate under occlusion for curvature-tuned neurons but easier ones for non-curvature-tuned neurons (compare shapes in Figs. 4 and 5), simply because it was feasible to identify a critical feature to manipulate for the former group but not the latter. This would lead to higher psychometric thresholds for the former than the latter and thus lower threshold ratios. This however, was not the case in our data: there was no difference in the psychometric thresholds between

the sessions with curvature-tuned versus non-curvature-tuned neurons (*T* test, *p* > 0.5; also see example cell in Fig. 3 and scatter in Fig. 5 in Kosai et al., 2014). This is because we always strived to find two shapes with a localized difference in their contour regardless of whether the neuron was curvature-tuned or not. In summary, the results above suggest that curvature-tuned V4 neurons demonstrate a more gradual decline in responses in the presence of occlusion as compared to non-curvature-tuned neurons.

The current best circuit models of V4 that capture selectivity for isolated shapes in both the curvature-tuned and non-curvature-tuned populations (Cadieu et al., 2007) behave more like the non-curvature-tuned neurons, i.e. their responses decline rapidly with increasing occlusion. Our preliminary modeling efforts suggest that if we include an algorithm for segmenting the image into component objects prior to encoding, the behavior of curvature-tuned neurons, i.e. their shallower decline in responses with occlusion, can be captured (Nicholas et al., 2014). This supports the idea that curvature-tuned neurons encode the segmented version of the occluded image while the non-curvature-tuned neurons do not. These results are consistent with previous evidence from lesion studies and physiology that V4 may contribute to image segmentation in general and contour-based segmentation in particular. First, lesions in primate V4 profoundly disrupt form discrimination of objects that require segmentation from the background (Merigan, 1996) and V4 lesions in humans disrupt texture-based segmentation (Allen et al., 2009). Second, neurophysiological studies on figure-ground modulation and border ownership hypothesize that segmentation is achieved by contour-based strategies that likely originate in area V4. For example, border ownership signals in area V2 (Zhou et al., 2000), which indicate which of two overlapping objects a contour belongs to, are hypothesized to depend on feedback from hypothetical "grouping cells" in V4, which segment images based on Gestalt rules of continuity and convexity of the bounding contour (Craft et al., 2007). Our finding that curvature-tuned neurons are highly sensitive to shape information despite occlusions, and that their shape selectivity emerges well before behavioral decision-related signals in V4 (Kosai et al., 2014), confirms the importance of contour-based mechanisms in processing visual scenes with occlusion and provides direct neurophysiological evidence in favor of the psychophysical theory that image segmentation and grouping and the ensuing recognition and discrimination of shapes under occlusion are mediated by contour-based strategies.

Several psychophysical studies suggest that contour-based grouping can be time-consuming. The incremental grouping theory, a model for contour-based segmentation developed based on psychophysical and neurophysiological findings, hypothesizes a gradual spread of enhanced activity across the representation of an object in the visual cortex (Roelfsema, 2006). Consistent with this idea we find that shape selectivity emerges gradually in the presence of partial occlusion. Fig. 7 shows the time course of shape selectivity for the neuron in Fig. 4.
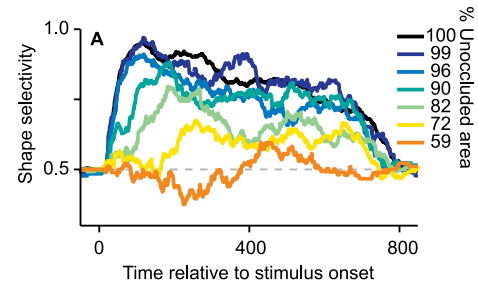

Fig. 7. Time course of shape selectivity under partial occlusion. (A) Data from an example neuron (same as in Figs. 1 and 4). (A) Time course of selectivity for unoccluded shapes (black) and for shapes under different occlusion levels (colored lines). Selectivity for unoccluded shapes was strong and peaked soon after stimulus onset; selectivity weakened with increasing occlusion level and peaked later. Previously published in Kosai et al. (2014).

To quantify shape selectivity as a function of time, we performed a sliding-window ROC analysis on the responses to the preferred and non-preferred stimuli. At each time point (1-ms increments), we counted spikes in a 100-ms centered window and assessed selectivity by computing the area under the ROC curve constructed from the spike count distributions for the preferred and non-preferred stimuli. Selectivity values ranged from 0.5 (nonselective) to 1.0 (very selective). For unoccluded stimuli shape selectivity was strong and emerged early. With increasing levels of occlusion, shape selectivity declined gradually and emerged progressively later. This protracted buildup of selectivity is a unique observation: while a variety of stimulus manipulations including contrast (Gawne, 2000), motion (Kawano et al., 1994), spatial frequency (Frazor et al., 2004) and distance from the RF center (Bringuier et al., 1999; Rossi et al., 2001), alter response latencies in the visual cortex, we are unaware of any that alter the latency of selectivity but not the latency of the response. The delayed onset of selectivity under occlusion is consistent with the idea that shape responses in V4 represent an inference signal about the presence of the preferred feature of the neuron in the visual scene and the computation of this inference takes time when confronted with the ambiguity imposed by occlusion.

In summary, we have shown that (i) responses of curvature-tuned V4 neurons are more robust under conditions of partial occlusion as compared to other shape-selective V4 neurons; and (ii) shape selectivity emerges gradually in the presence of partial occlusion. These findings are consistent with theory based on human psychophysics that hypothesizes that contour-based mechanisms underlie segmentation and object representation. There is still a lot of work to be done: current circuit level models of V4 that predict responses of curvature-tuned neurons (Cadieu et al., 2007) cannot predict the gradual build-up of shape selectivity under partial occlusion, so these models need to be updated. Our preliminary results suggest that the addition of a recurrent process that differentially weights edge elements in the visual stimulus based on whether or not they belong to the preferred contour could reproduce the observed delay in the presence of occlusion (Nicholas et al., 2014). We also do not know how responses of non-curvature-tuned

neurons are built and why these neurons are less resistant to occlusion. Nevertheless, these results highlight the strength of the monkey model for investigations into the neural bases of visual object recognition.

Understanding the neural basis of segmentation and object recognition is arguably one of the toughest challenges faced by Systems Neuroscience. To rise to this challenge, it is critical and essential that our experimental investigations are informed by knowledge gained from all approaches—shape theory, human psychophysics, animal lesion studies, human patients, etc. For example, our stimulus design can and should be guided by the more stringent cases of image segmentation used in human psychophysics. Because the monkey and human visual systems are remarkably similar, this can be readily done. As a field, it is also important to leverage our collective experimental results, from various labs, across brain areas, and diverse sets of stimuli, to constrain biologically-realistic computational models to gain deeper insight into the underlying circuitry and mechanisms.

REFERENCES

Allen HA, Humphreys GW, Colin J, Neumann H (2009) Ventral extrastriate cortical areas are required for human visual texture segmentation. J Vis 9(2):1–14.

Asaad WF, Rainer G, Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. Neuron 21:1399–1407.

Ben-Av MB, Sagi D, Braun J (1992) Visual attention and perceptual grouping. Percept Psychophys 52:277–294.

Berg DJ, Boehnke SE, Marino RA, Munoz DP, Itti L (2009) Free viewing of dynamic stimuli by humans and monkeys. J Vis 9. 19.1–15.

Bringuier V, Chavane F, Glaeser L, Fregnac Y (1999) Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. Science 283:695–699.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci 12:4745–4765.

Bushnell BN, Harding PJ, Kosai Y, Pasupathy A (2011) Partial occlusion modulates contour-based shape encoding in primate area V4. J Neurosci 31:4012–4024.

Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. J Neurophysiol 98:1733–1750.

Cavonius CR, Robbins DO (1973) Relationships between luminance and visual acuity in the rhesus monkey. J Physiol 232:239–246.

Clowes MB (1971) On seeing things. Artif Intell 17:79–116.

Craft E, Schutze H, Niebur E, von der Heydt R (2007) A neural model of figure-ground organization. J Neurophysiol 97:4310–4326.

Croner LJ, Albright TD (1999) Segmentation by color influences responses of motion-sensitive neurons in the cortical middle temporal visual area. J Neurosci 19:3935–3951.

Elder JH, Zucker SW (1998) Evidence for boundary-specific grouping. Vis Res 38:143–152.

Farah MJ (1990) Visual agnosia. Cambridge, MA: MIT Press.

Fine I, MacLeod DI, Boynton GM (2003) Surface segmentation based on the luminance and color statistics of natural scenes. J Opt Soc Am A Opt Image Sci Vis 20:1283–1291.

Frazor RA, Albrecht DG, Geisler WS, Crane AM (2004) Visual cortex neurons of monkeys and cats: temporal dynamics of the spatial frequency response function. J Neurophysiol 91:2607–2627.

Fuchs AF (1967) Saccadic and smooth pursuit eye movements in the monkey. J Physiol 191:609–631.

Gallant JL, Shoup, Mazer JA (2000) A human extrastriate area functionally homologous to macaque V4. Neuron 27:227–235.

Gawne TJ (2000) The simultaneous coding of orientation and contrast in the responses of V1 complex cells. Exp Brain Res 133:293–302.

Gross CG (1973) Visual functions of inferotemporal cortex. In: Jung R, editor. Handbook of sensory physiology, Vol. VII, Part 3B. Berlin: Springer-Verlag. p. 451–482.

Guzman A (1968) Decomposition of a visual scene into three-dimensional bodies. In: Fall joint computer conference. Arlington, VA: AFIPS Press. p. 291–304.

Helmholtz H (1909) Treatise on physiological optics. New : Dover.

Houtkamp R, Spekreijse H, Roelfsema PR (2003) A gradual spread of attention during mental curve tracing. Percept Psychophys 65:1136–1144.

Huffman DA (1971) Impossible objects as nonsense sentences. Mach Intell 5:295–323.

Jolicoeur P, Ullman S, Mackay M (1986) Curve tracing: a possible basic operation in the perception of spatial relations. Mem Cogn 14:129–140.

Kawano K, Shidara M, Watanabe Y, Yamane S (1994) Neural activity in cortical area MST of alert monkey during ocular following responses. J Neurophysiol 71:2305–2324.

Kosai Y, El-Shamayleh Y, Fyall A, Pasupathy (2014) The role of visual area V4 in the discrimination of partially occluded shapes. J Neurosci 34:8570–8584.

Kourtzi Z, Connor CE (2011) Neural representations for object perception: structure, category, and adaptive coding. Ann Rev Neurosci 34:45–67.

Leung T, Malik J (1998). Contour continuity in region-based segmentation. In: Proc. Euro. conf. computer vision, Vol. 1 (Burkhardt H, Neumann B, eds), pp. 544–559. Freiburg, Germany.

Merigan WH (1996) Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. Vis Neurosci 13:51–60.

Mishkin M, Pribram KH (1954) Visual discrimination performance following partial ablations of the temporal lobe: I. Ventral vs lateral. J Comp Physiol Psychol 47:14–20.

Mumford D, Kosslyn SM, Hillger LA, Herrnstein RJ (1987) Discriminating figure from ground: the role of edge detection and region growing. Proc Natl Acad Sci 84:7354–7358.

Munakata Y, Santos LR, Spelke ES, Hauser MD, O'Reilly RC (2001) Visual representation in the wild: how rhesus monkeys parse objects. J Cogn Neurosci 13:44–58.

Nauhaus I, Nielsen K, Disney AA, Callaway EM (2012) Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex. Nat Neurosci 15:1683–1690.

Nicholas E, Pasupathy A, Bair W (2014) Advancing models of shape selectivity in V4. SFN abstracts.

Orban GA, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. Trends Cogn Sci 8:315–324.

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

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

Qiu FT, Sugihara T, von der Heydt R (2007) Figure-ground mechanisms provide structure for selective attention. Nat Neurosci 10:1492–1499.

Roelfsema PR (2006) Cortical algorithms for perceptual grouping. Annu Rev Neurosci 29:203–227.

Rossi AF, Desimone R, Ungerleider LG (2001) Contextual modulation in primary visual cortex of macaques. J Neurosci 21:1698–1709.

Rubin N (2001) The role of junctions in surface completion and contour matching. Perception 30:339–366.

Schor CM, Tyler CW (1981) Spatio-temporal properties of Panum's fusional area. Vis Res 21:683–692.

Van Essen D (2004) Organization of visual areas in macaque and human cerebral cortex. In: Chalupa LM, Werner JS, editors. The visual neurosciences, Vol. 1. MIT Press. p. 507–521.

Waltz DL (1975) Understanding line drawings of scenes with shadows. In: Winston PH, editor. The psychology of computer vision. New York: McGraw Hill.

Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. Psychol Bull 138:1172–1217.

Wertheimer G (1938) Laws of organization in perceptual forms. In: Ellis ED, editor. A sourcebook of Gestalt psychology. London: Routledge and Kegan Paul.

Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20:6594–6611.