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 101
The approach 101
Contour-based segmentation and primate V4 102
Non-human primate model 102
Behavioral task 103
Results 104
Acknowledgments 108
References 108

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
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. 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).

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
characteristics, e.g. overall stimulus orientation and shape stimuli may be dictated by the surface-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 sub-group 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 \(0^\circ\) (right of center) to \(360^\circ\) in a counter-clockwise direction. A 2-D Gaussian function with a peak at curvature \(+1.0\) (sharp convex) and angular position \(322^\circ\) (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...
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. 4. Results from an example curvature-tuned neuron during behavior. Same neuron as in Fig. 1. (A, B) Response PSTHs (r = 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).

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.
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 $>0.5$; $N = 24/61$), 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 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 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.

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.

Acknowledgments—This work was funded by NEI grant R01EYO11839 to A. Pasupathy, Vision Core grant P30EYO17330 to the University of Washington, and NIH ORIP grant OD010425 to the Washington National Primate Research Center. I thank Wyeth Bair for helpful discussions and comments on the manuscript. Technical support was provided by Bioengineering group at the Washington National Primate Research Center.

REFERENCES


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


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