Shape Selectivity in V4 Using Novel Time Dynamic and Three Dimensionalized Stimuli

Alex Rockhill, Wyeth Bair, Anitha Pasupathy

August 6, 2016

Abstract

The study of intermediate brain areas in the object recognition pathway, such as the fourth visual cortex (V4), is important in understanding the mechanisms of the visual processing system. The function of V4 neurons has been characterized by Pasupathy and Connor for two dimensional (2D) shapes by an angular position and curvature (APC) model where objects are represented by contour components at relative angular positions on the object. This model was extended to three dimensionalized (3D) objects by modeling neuronal firing rate as a function of zenith and azimuthal angle position for two curvatures; a “bowl” and a “protrusion.” The goodness-of-fit of this model was consistent with that of the Pasupathy and Connor APC model for 2D shapes. Hence, our data supported the extension of the APC model to 3D shapes.

Area V4 is a cortical region that processes visual information in order to contribute to our ability to characterize objects. It integrates information about boundary and surface features from the primary visual cortex and sends signals downstream to the inferior temporal cortex, which supports complex object recognition in a pathway called the ventral stream of visual processing (Kandel 2013). Pasupathy and Connor have previously developed an APC model, which characterized the function of V4 neurons as selecting for a particular boundary curvature of a 2D object at a particular orientation relative to the center of the object (2001). Our investigation sought to determine whether this model applied to more realistic 3D objects as shown in *Figure 1*. The 3D image space is much larger than the 2D image space. There is one dimension of angular position and curvature for 2D shapes compared to two dimensions of both angular position and curvature for 3D shapes, and 3D shapes must be illuminated by a light source, with that light source having any location in 3D space. Thus, in order to investigate the applicability of the APC model to 3D objects, simplifications had to be made. In this experiment we chose to focus on the two dimensions of angular position by modeling the response of a neuron in relation to the zenith and azimuthal of a salient feature as described in depth in the *Methods* and *Model* sections. In addition, we are currently investigating of the temporal resolution of V4 neurons so that we can show images for the optimum length of time and increase the efficiency of our data collection.

We hypothesize that V4 neurons may have a distinct temporal integration time on the order of tens of milliseconds, which is required to recognize angular and positional features of rapidly moving or changing realistic shapes. If our hypothesis is correct, then we should be able to characterize the angular and positional preference and temporal resolution of these neurons using fast streams of realistic, dynamic stimuli. Our first experiment testing this used the “bowl” and “protrusion” 3D shapes. These 3D stimuli were shown in random order interspersed with a subset of Pasupathy and Connor’s 2D shapes in their previous experimental paradigm; stimuli are shown for relatively long intervals of 300 milliseconds, interspersed with blank inter-stimulus intervals (2001). Our results suggested that the previous APC model would be able to be extended to more realistic, 3D shapes.

In order to test this hypothesis further, we are currently increasing the efficiency of our experimental paradigm by testing the temporal resolution of V4 neurons. To do this, we created 2D stimuli that change their boundary features rapidly by randomly showing multiple shapes in quick temporal progression. This data will allow is to compare the ability of neurons to characterize 2D boundary curvature preference for different durations of stimulus presentation. This will allow us to find the optimum amount of time necessary to show stimuli to V4 neurons such that we are able to reproduce previous results, thus increasing the efficiency with which our data is collected.

Further experiments will test whether 3D objects shown nearer to the boundary of the temporal resolution of V4 neurons will allow us to determine their APC preference at greater detail; using a larger range of positions and curvatures, and with greater confidence; having increased statistical power. Better understanding of the function of V4 neurons will contribute to the optimization of deep neural networks, the development of visual therapeutics and increase our basic scientific understanding of perceptual phenomena as explained in applications.

Methods

We recorded from single, well-isolated neurons in brain area V4 of awake, behaving macaques while they fixated on a spot. Receptive fields were hand-mapped to determine their approximate size and location before our novel stimuli were shown. Our 3D shapes were rendered using the commercial software package *Blender* as shown in *Figure 2*. These shapes were created by modifying a mesh of 1026 points on a sphere (two points at each pole with 16 equally spaced horizontal rings and 64 points on each ring) connected by flat, three dimensional planes which were smoothed using Blender's smoothing algorithm. Pixel values for the image were then created by calculating the amount of light that reflected from a light source off of these mesh planes and were detected by a camera. The "bowl" and "protrusion" shapes were created by moving one point on a sphere one diameter towards or away from the center of the sphere with the interpolation feature in Blender enabled, which moved the surrounding points proportionally. The camera was aligned so that the object filled the space. The light sources were positioned in a plane three radii's length toward the camera so that they were behind the camera for visual appearance and clarity. The three light sources were (in this plane) both three radii upwards and to the left, both three radii upwards and to the right and just three radii downward. Images of 150 shapes (75 "bowls" and 75 "protrusions" where 25 orientations of each were shown at three different light angles) were then rendered in bitmap file format, as were 227 of the 366 Pasupathy-Connor shapes. These were hand-picked by Bair with the intent of preserving coverage of the APC image space while reducing the total number of shapes that were shown. These bitmap images were then shown on a cathode ray tube monitor, calibrated so that the images were grayscale (the intensity values of red, green and blue were equal), using Pype software developed by James A. Mazer and Pasupathy in 2005-2008. Neurons were simultaneously recorded from using Plexon software. Data were analyzed in MATLAB and Python with code contributed by Dean Pospisil for APC analysis of Pasupathy-Connor shapes that was consistent with previous analysis. Shapes for the 2D temporal resolution experiment were created by fitting a cubic spline to 16 points around a circle with one point moved half a radius inward to create a dented circle shape, and use the same display, recording and analysis software.

Model

The APC model from the Pasupathy and Connor experiments fitted normal distributions to both the angular position in radians around the boundary of a shape starting from a particular point and the curvature of the boundary relative to the center of the shape. Previous research about vision and neuroscience has shown that the most computationally efficient coordinate system is the system most prevalent in biology. For example, Batista et al found that reach planning was eye-centered; subjects that reached toward a target in a particular direction caused a set of neurons to fire that was dependent on the relative direction to where they were looking (1999). This hypothesized object-centered coordinate system is also computationally efficient; objects can be separated, recognized separately, and then rejoined, and this coordinate system is supported by data.

In order to extend this model to 3D objects, the parameter space must be made smaller; the curvature was only manipulated by having “bowls” and “protrusions” and there were three light source locations which were each modeled separately. Preliminary analysis suggested that orientations of “bowls” and “protrusions” were favored regardless of the light source location. Hence, the 3D APC model fit normal distributions to the angle of the salient feature in both positional directions; the zenith and azimuthal angles. The best fits were determined by finding the minimum error between actual responses and predicted responses for a two dimensional Gaussian model with a discrete sampling the possible expected values and variances. Variances were capped on the low end to prevent trivial Gaussian fits to the maximum response value. Also, error was calculated using only responses above and below one standard deviation from the mean so as not to cause the disproportionate number of responses to non-preferred objects to be the dominate source of error, which was observed to cause too large of variances. This conclusion is ideologically consistent with the analysis of Pasupathy and Connor, where they used only a discrete number of hand selected points for their APC model so as not to cause the model to be fit predominantly to less salient features of a shape, which are more prevalent (2001). For example, in *Figure 1c*, nearly half of the boundary area is the broad, flat section to the lower left. Each of the points on this section would be given equal predominance with every other point, including the points to the upper right which constitute a much more salient feature but are much fewer in number.

Results

Neurons in area V4 have been shown to encode the boundary curvature of simple, flat, 2D shapes in an object-centered coordinate system. For example, a particular neuron may prefer a wide variety of stimuli that have a protrusion (convexity) to the upper right. The subset of 229 of the shapes—used by Pasupathy and Connor— that were used in this experiment were fit to the APC model. The median R^2 value of the 16 cells studied was 0.57, which was almost exactly the same as the 0.57 median *r* value found by Pasupathy and Connor. We found that our results replicated those of Pasupathy and Connor (2001), suggesting that the neurons being studied in this experiment had the same response characteristics as those studied previously.

Neurons such as those in *Figure 3* responded preferentially to 3D shapes with curvature at a certain orientation (in this example, a bowl pointing upward and tilted towards the viewer). The median R^2 value for the Gaussian fits for zenith and azimuthal angles to responses greater than one standard deviation greater than the mean was 0.51, with fits ranging from 0.26 to 0.84. This inexact correlation between APC and firing rate is consistent with previous findings such as Britten et al who found that large populations of neurons can have a high safety factor between a stimulus and response if the neurons are generally correlated regardless of strength of correlation. This evidence supports the characterization of these neurons with a 3D extension of the APC model.

These data suggest that 3D shapes cause cells to fire at a greater rate. *Figure 4* shows the response of cells to 3D shapes plotted against responses for 2D shapes. Eleven cells had responses with p-values less than 0.05 for a two-sample t-test in the direction of 3D responses being greater, compared to five with p-values less than 0.05 in the direction of 2D responses being greater.

Discussion

Our results support the extension of the APC model for cells in area V4 into more dimensions, which better approximates naturalistic stimuli. We found that cells being studied in this experiment could be characterized by the APC model for 2D shapes with the same goodness-of-fit as was found by Pasupathy and Connor (2001). This suggests that our setup was correctly displaying images and recording from neurons in V4. We found that the goodness-of-fit of the two dimensional Gaussian model for neurons being shown 3D “bowls” and “protrusions” was similar to that found by Pasupathy and Connor. The most parsimonious explanation for our ability to model responses to 2D and 3D shapes with the same method is that V4 neurons operate using an object-based system. Neurons in this system would respond to a feature at a particular orientation relative to the object in their receptive field. This system would yield the same results for both 2D and 3D objects. If the response of V4 neurons was not object-based but rather image-based, two such dissimilar objects as the 2D and 3D shapes (*Figure 1*) would likely not be able to drive the same neuron. Many important conditions are vastly different, such as color, luminance, texture and contrast. Therefore, our data support that V4 neurons use an object-based object recognition system that depends on curvature at angular positions.

Three dimensionalized objects are much more similar to the three dimensional objects like food, predators, shelter and mates which are relevant to our evolution. The data from this experiment supports our hypothesis that 3D shapes are more salient stimuli to V4 neurons than 2D stimuli, which is consistent with our intuition that these shapes are more realistic and therefore more salient.

Applications

A better understanding of the role of V4 neurons in the object recognition system of our brains can improve analogous visual object recognition computer programs, and contribute to designing visual prosthetics and the explanation of visual phenomena. Deep neural networks work analogously to our visual processing system, wherein hierarchical units process information in order to affect downstream units (Krizhevsky 2012). In this way many light-dark boundaries become a curve, many curves become a shape and many shapes become a gestalt. These stages of processing are roughly analogous to the primary visual cortex (V1), V4, the inferior temporal cortex (IT) and the prefrontal cortex (Kandel 2013). Artificial deep neural networks often experience systemic failure by having high confidence predictions for random or unrecognizable images (Nguyen et al 2015), while biological networks often do not fail in this way. By better understanding the role of each area in the brain, the networks based on this concept can be better debugged.

Incremental increases in our understanding of the visual processing system furthers us along the path of reproducing it biologically as well as artificially. Neural prosthetics have been developed to bypass nonfunctional parts of the motor system such as using signals from the motor cortex to affect neurons in the spinal cord, bypassing a spinal cord injury (Mondello et al 2014). Analogically, signals from neurons in V1 could be used to affect neurons in IT if a computer was able to process and relay this information. In this way, understanding of the function of V4 could lead to the development of a neural prosthetic. The effector, signals to IT, need not be controlled by signals from V1; using artificial signals, mimicked from V1 signals, an image could be conjured in a subject's perception. There are a myriad of ways to apply artificially conjuring a visual image, such as telepathic communication.

Many of the greatest discoveries that will be enabled by better basic scientific understanding of the visual processing system, especially V4, are unknown. Understanding the systemic function of V1 led Bressloff et al to hypothesize that uniform increases in the firing rate of neurons in this area may cause the geometric pattern of the visual hallucinations caused by psychoactive drugs (2001). Connections such as this were not predicted by those studying V1. It was not until the dissemination of the findings on this brain area that multidisciplinary collaborations were able to hypothesize explanations for seemingly unrelated phenomena. The scientific implications of better understanding of the function of V4 in the visual system are yet to be fully comprehended.

References

Batista, AP., CA Buneo, LH Snyder, RA Anderson. Reach plans in eye-centered coordinates. Science. 1999 Jul 9; 285(5425):257-60.

Bressloff, Paul C., Jack D. Cowan, Martin Golubitsky, Peter J. Thomas, and Matthew C. Wiener (2001). What geometric visual hallucinations tell us about the visual cortex. Unpublished.

Britten, K.H., W.T. Newsome, M.N. Shadlen, S. Celebrini, J. A. Movshon (1996). A relationship between behavioral choice and the visual responses of neurons in the macaque MT. Visual Neuroscience (1996), 13, 87-100.

Kandel, Eric, James H. Schwartz, Thomas M. Jessell, Steven A. Siegelbaum, A. J. Hudspeth (2013). “Principles of Neural Science.” The McGraw Hill Companies Inc. New York. Print. 396, 623.

Krizhevsky, Alex, Ilya Sutskever, Geoffrey E. Hinton (2012). ImageNet Classification with Deep Convolutional Neural Networks NIPS; 1:9.

Mondello, Sarah E., Michael R. Kasten, Philip J. Horner, and Chet T. Moritz (2014). Therapeutic intraspinal stimulation to generate activity and promote long-term recovery. Front Neurosci. 2014; 8: 21.

Nguyen A, Yosinski J, Clune J. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In Computer Vision and Pattern Recognition (CVPR '15), IEEE, 2015.

Pasupathy A, Connor CE (2001). Shape representation in area V4: position-specific tuning for boundary conformation. J Neurophysiol. 2001 Nov; 86(5):2505-19.

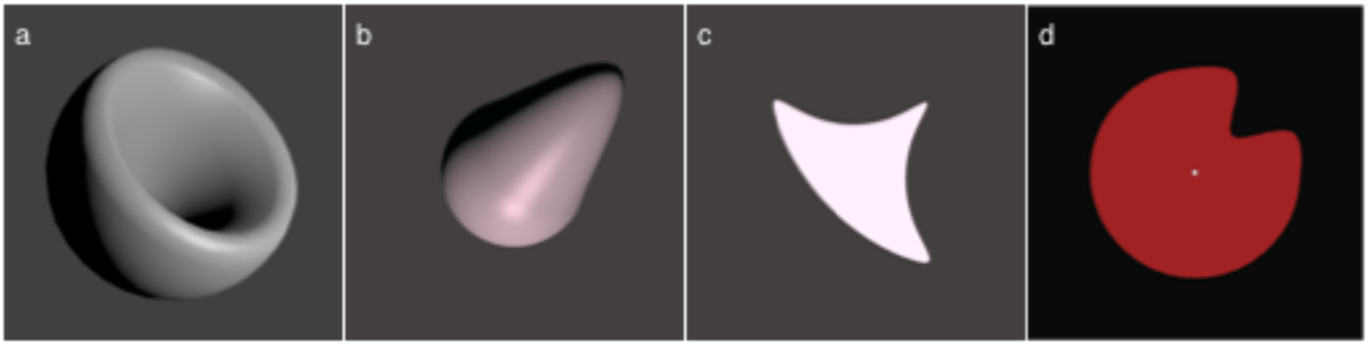


Figure 1a. A three dimensionalized bowl stimulus. b. A three dimensionalized protrusion stimulus. c. A Pasupathy Connor two dimensional stimulus. d. A two dimensional temporal resolution stimulus.

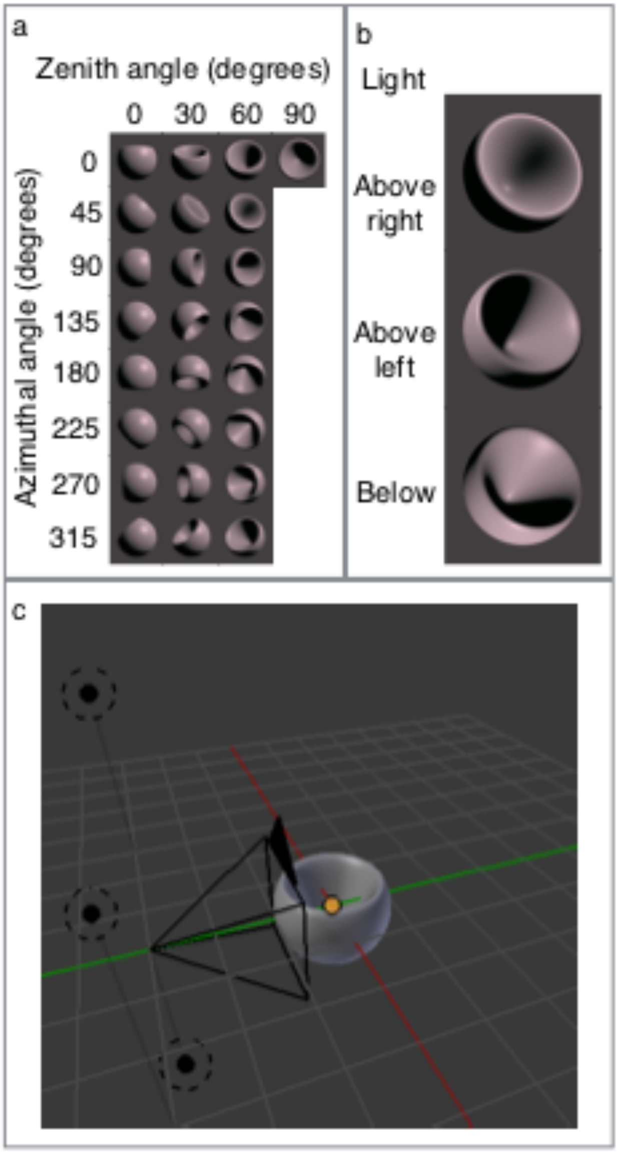


Figure 2a. *The positional paradigm for 3D shapes, varying zenith and azimuthal angles.*
b. The lighting paradigm for the 3D shapes. Using light sources from above and to the left, above and to the right and below at the center.
c. A view of the shape rendering setup in Blender.

Responses to Shapes

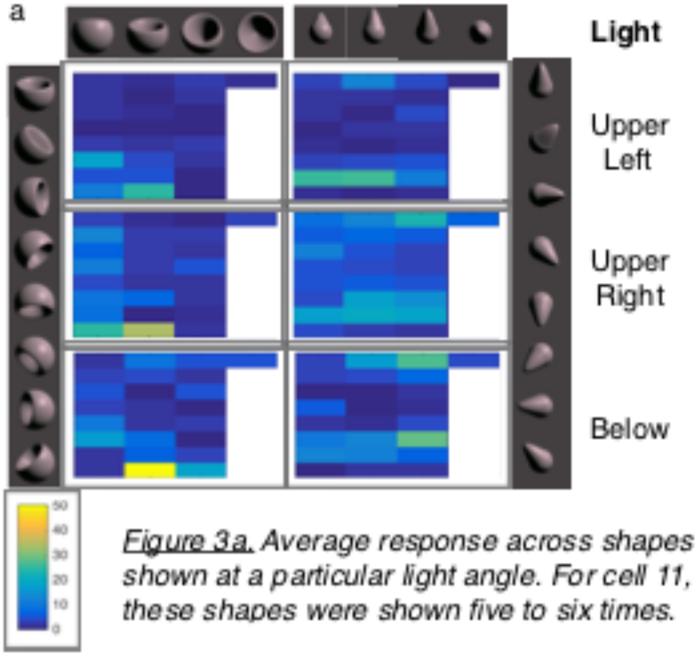
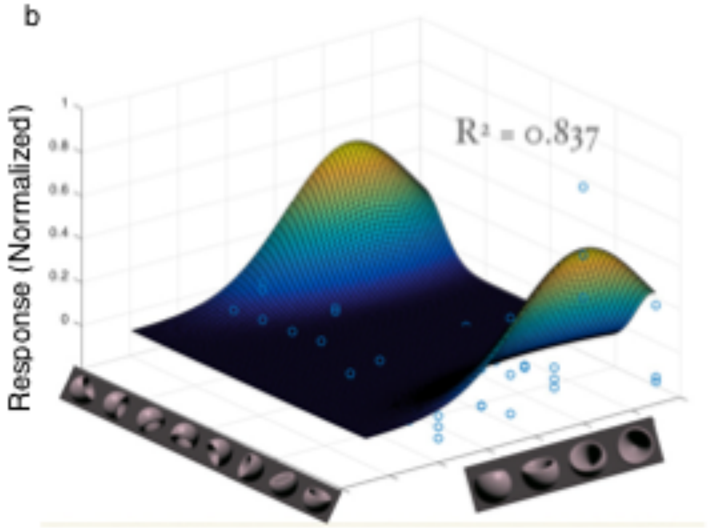


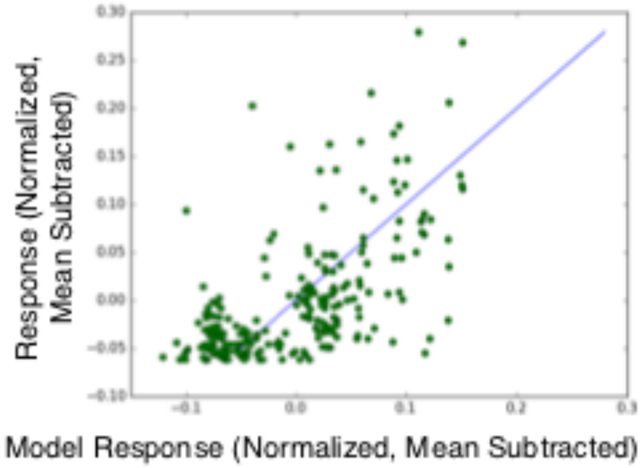
Figure 3a. Average response across shapes shown at a particular light angle. For cell 11, these shapes were shown five to six times.

Gaussian Fit for "Bowls" Shown to Cell 11



b. A graph of a two dimensional Gaussian fit of responses greater than one standard deviation greater than the mean as a function of zenith and azimuthal angles for "bowls" for one particular cell. The surface is the predicted response: color (cold being 0 and warm being 1) is redundant with the z-axis and is only used for visual clarity. The actual responses are plotted with hollow points.

APC Fit for 2D Shapes Shown to Cell 11



c. The modeled response for the 229 2D Pasupathy-Connor shapes that were also shown is plotted on the x-axis and the actual response of the cell is plotted on the y-axis. A line of unity slope is plotted for visual aid. Points farther from the unity line are predicted with less accuracy than those closer.

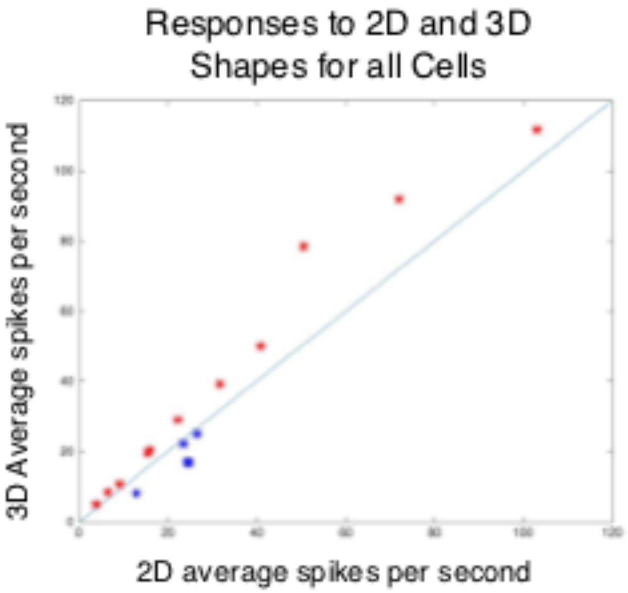


Figure 4. *Responses for 2D shapes plotted in relation to responses for 3D shapes. Points are plotted in red if a one directional two-sample t-test yielded a p-value of less than 0.05 with 3D responses having a greater mean (11) and in blue if the p-value was less than 0.05 and responses to 2D shapes were greater. No cells had non-significant t-tests in both directions.*