

From orientations to objects: Configural processing in the ventral stream

Hugh R. Wilson

Centre for Vision Research, York University,
Toronto, Ontario, Canada



Frances Wilkinson

Centre for Vision Research, York University,
Toronto, Ontario, Canada



The ventral or form vision hierarchy comprises a sequence of cortical areas in which successively more complex visual attributes are extracted, beginning with contour orientations in V1 and culminating in face and object representations at the highest levels. In addition, ventral areas exhibit increasing receptive field diameter by a factor of approximately three from area to area, and conversely neuron density decreases. We argue here that this is consistent with configural combination of adjacent orientations to form curves or angles, followed by combination of these to form descriptions of object shapes. Substantial data from psychophysics, functional magnetic resonance imaging (fMRI), and neurophysiology support this organization, and computational models consistent with it have also been proposed. We further argue that a key to the role of the ventral stream is dimensionality reduction in object representations.

etc. Skipping connections also incorporate skipping feedback (Nakamura, Gattass, Desimone, & Ungerleider, 1993; VanEssen et al., 1992).

This architecture naturally leads to the question: Why are there multiple cortical areas in the ventral pathway rather than just one or two? Furthermore, why are there different modes of connection and feedback among them? To explore current thinking about these questions, we shall first review anatomical data on the ventral pathway and then attempt to link these to relevant functional data from psychophysics, functional magnetic resonance imaging, (fMRI), and neurophysiology.

Introduction

The ventral, form vision hierarchy comprises a sequence of approximately five cortical areas in monkeys: V1, V2, V4, TEO (occipital-temporal cortex), and TE (temporal cortex) (VanEssen, Anderson, & Felleman, 1992). In humans, the higher levels of this pathway also include the fusiform face area (FFA) (Kanwisher, McDermott, & Chun, 1997) and the lateral occipital complex (LOC) (Haxby, Gobbini, Furey, Ishai, & Pietrini, 2001), for a total of approximately 10 ventral visual areas identified at present (Wang, Mruczek, Acaro, & Kastner, 2014). Each of these hierarchical areas provides input directly to the area above and receives feedback from that area. In addition to these direct connections, “skipping connections” also exist, in which an area provides input to a layer two levels above, such as V1 to V4, V4 to TE,

Anatomy of the ventral pathway

Each retina contains approximately 1.25 million ganglion cells (Rodieck, 1998), most of which project to the lateral geniculate and thence to V1. Estimates of the total number of neurons in V1 fall in the range of 1.4×10^8 (Leuba & Kraftsik, 1994) up to about 6.75×10^8 (Miller, Balaram, Young, & Kaas, 2014). Thus, there are between 100 and 500 times as many V1 neurons as there are ganglion cell axons. This, of course, reflects the presence of around 12 orientation selective neurons centered on each ganglion cell input, as well as at least six different peak spatial frequency channels (Wilson, McFarlane, & Phillips, 1983), three chromatic opponent channels, and a range of different disparities and motion direction selectivities. Multiplying these figures easily accounts for the factor of 100–500 increase of V1 neurons over retinal ganglion cells: V1 apparently explicitly represents the biologically most relevant local image features, such as contours, by using an over-complete code (Olshausen & Field, 1996). But what would happen if even a hundredfold neuron increase

Citation: Wilson, H. R., & Wilkinson, F. (2015). From orientations to objects: Configural processing in the ventral stream. *Journal of Vision*, 15(7):4, 1–10, <http://www.journalofvision.org/content/15/7/4>, doi:10.1167/15.7.4.

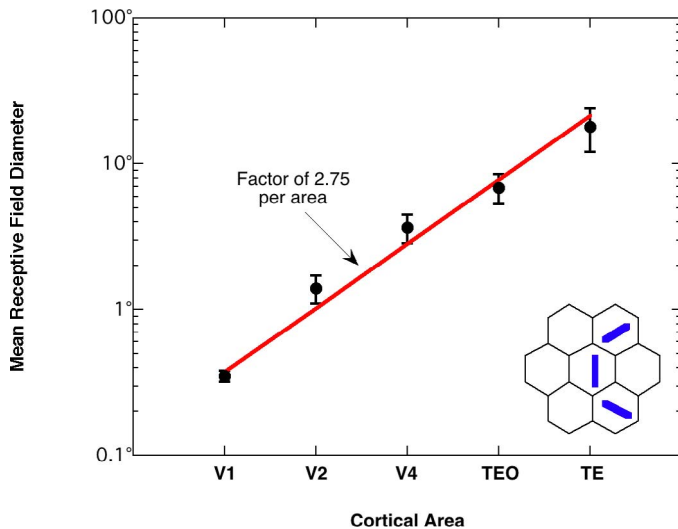


Figure 1. Mean data from four studies on receptive field diameter in successive cortical areas in the ventral pathway. Error bars represent the range of means across the four studies. The red line shows that the data can be fit by constant diameter increase of $2.75\times$ from area to area. Inset in lower right shows an example of a curved contour represented by a combination of adjacent, oriented (orientations in blue) V1 receptive fields (individual hexagons) in a $3.0\times$ larger diameter V2 receptive field.

were replicated in V2, then V4, then TEO, then TE? The result would be quite striking: The ventral visual pathway would have to contain more than 10^{16} neurons! However, there are only about 10^{11} neurons in the entire neocortex, so this combinatoric explosion cannot be embodied in the image analysis computations of the ventral pathway. Parenthetically, if there were 10^{16} neurons in TE, then the “grandmother cell” concept would indeed be viable, but it clearly is not.

As a first step toward understanding the transformations actually present in the ventral pathway, it is instructive to examine both receptive field size and cortical neuron density along the pathway. Average macaque receptive field diameters within 5° of the fovea are plotted in Figure 1, where data are averages from several studies (Boussaoud, Desimone, & Ungerleider, 1991; Elston & Rosa, 1998; Kobatake & Tanaka, 1994; Op de Beeck & Vogels, 2000). See Wilson and Wilkinson (2014) for a graph with data from the individual studies. The data are extremely well fit by a straight line on this semilog plot, and the best fitting line indicates that receptive field diameter increases by a factor of 2.75 from area to area. As the 95% confidence interval for the slope lies between 2.27 and 3.24, it does not differ significantly from a factor of 3.0. Thus, the V1 mean diameter of 0.36° increases to a diameter of $21\text{--}29^\circ$ in TE, a factor of 57–81. The diameter increase of about 3.0 is consistent with receptive fields one area higher in the hierarchy being constructed from combi-

nations of nearest neighbors in its input area. Combination of nearest neighbors is dependent on a retinotopic map, so reduction in retinotopy might reduce the nearest neighbor constraint in higher areas. This arrangement is illustrated schematically in the figure inset in which a hexagonal array three times the V1 receptive field diameter combines elements to represent a curved arc. Angles and contour intersections (e.g., T-junctions) can also be constructed in this way. All of these configurations have been reported to occur in V2 (Anzai, Peng, & VanEssen, 2007; Hegd e & Van Essen, 2003). It is important to note that curves, angles, etc., cannot be constructed from orientations on a grid much smaller than 3×3 nearest neighbors.

Given that receptive fields cover approximately 9.0 times the area in each successive ventral cortical area, it is mathematically possible to exactly represent all spatial information from the input area by subsampling to $1/9$ as many spatial locations with 9.0 times as many types of receptive field configurations per location. Thus, with 12 orientations in V1, V2 could represent $12 \times 9 = 108$ different curves, angles, etc. (108 is the number of linearly independent configurations; all other configurations can be described by linear combinations of these 108 dimensions.) If subsampling is all that is going on, each area in the ventral pathway should contain about the same number of neurons. However, recent data show that this possibility is dramatically wrong. Measurements of neuron density per mm^3 in several primates across all cortical areas show that density decreases exponentially along an axis from posterior medial to anterior lateral cortex, with the highest density being in V1 (Cahalane, Charvet, & Finlay, 2012). Orthogonal to this exponential decay axis, cortical neuron densities are roughly constant. The best fitting surface to the baboon data is shown in Figure 2. Along the axis of exponential decrease, neuron density drops by roughly a factor of 6.0 from V1 to lateral prefrontal cortex. Most of this density decrease occurs from V1 roughly along the ventral visual pathway, so it can be concluded with reasonable certainty that neuron density is at least a factor of 5.0 times lower in TE than it is in V1. Furthermore, the surface areas of V4, TEO, and TE are smaller than those of V1 and V2 (VanEssen et al., 1992), so the absolute number of neurons in TE must be somewhat less than one fifth that in V1.

These anatomical observations indicate that the visual system is progressively reducing the amount of information encoded at higher levels of the form vision hierarchy. It is appropriate to think of each output neuron in an area as representing an independent dimension along which shape representations may vary. From this perspective, one major role of intermediate level form vision is projecting image information into a very low dimensional subspace

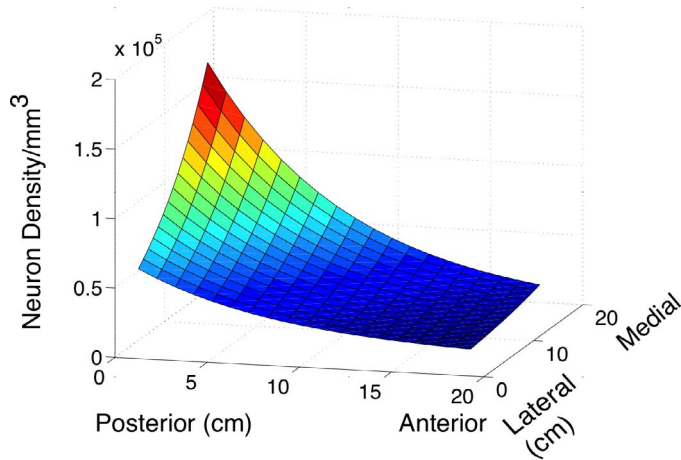


Figure 2. Best fitting two-dimensional surface describing neuron density in baboon neocortex, with density represented from highest to lowest on a red-yellow-green-blue scale. The axis of maximum density change runs from posterior medial to anterior lateral cortex and represents an exponential decrease by a factor of 6.0 across the surface. Posterior-anterior and lateral-medial axis dimensions are cm. Plotted from equation provided by Cahalane et al. 2012.

relative to representations in V1, analogous to the compression in jpeg images. Dramatic evidence supporting this interpretation has just been published by Lehky and colleagues (Lehky, Kiani, Esteky, & Tanaka, 2014). These authors compared the dimensionality of a large class of visual stimuli to the estimated dimensionality of their neural representations in macaque TE. Color images (125×125 pixels) of 806 common objects (faces, dogs, chairs, flowers, etc.) were used to stimulate each of 647 macaque TE neurons. Using principal component analysis, the dimensionality of the stimulus space was calculated to be approximately 507. However, the population code among the 647 neurons was dramatically reduced to just 93 ± 11 dimensions. This is in striking agreement with the estimate above of a five-fold reduction based on neocortical anatomy (Cahalane et al., 2012), and it supports the hypothesis of dimensionality reduction playing a major role in the ventral pathway.

Dimensionality reduction generally implies loss of information, so the natural question is how this can be accomplished without a major compromise in object recognition. There are two plausible ways in which dimensionality reduction can still be effective in retaining critical information about biologically relevant objects. The first is that significant portions of many visual patterns are textures: fields of grass, foliage in a forest, a stucco wall, etc. Extensive research has shown that humans encode textures using a small number of statistical variables, such as mean luminance, contrast, an orientation histogram, and a mechanism sensitive to the darkest texture elements

(Chubb, Landy, & Econopouly, 2004; Landy, 2014). This statistical description represents an enormous dimensionality reduction for large areas of many images.

A second major aspect of dimensionality reduction results from the fact that there are many correlations among image components in natural shapes. The classic way of effectively using image correlation to effectively reduce dimensionality is principal component analysis (PCA), which can readily be implemented by neural networks using Hebb synapses (Diamantaras & Kung, 1996). Indeed, there is recent evidence that adults implicitly learn both the mean (or prototype) plus at least several principal components when they attempt to memorize a group of faces (Gao & Wilson, 2014). Unpublished measurements from our laboratory show that 90% of the variance in the geometric shape of faces in both frontal and partial side view can be captured by just one fourth of the principal components. Similar results obtain for other representations such as independent components. The implication is that far fewer dimensions are needed for accurate representation of natural images than are present in the image data. It has also been demonstrated that a few principal components can effectively predict visual object category (face, shoe, chair, etc.) from the fMRI BOLD signal using cross-validation (O'Toole, Jiang, Abdi, & Haxby, 2005).

Finally, important evidence has been reported regarding the number of synapses per cortical neuron throughout the ventral visual pathway. As neuron density becomes lower in more anterior lateral areas, the number of spine synapses onto layer III pyramidal neurons increases dramatically (Elston, 2002; Elston & Rosa, 1998). In fact, the increase is approximately exponential and represents roughly a seven-fold increase from V1 to TE. This represents a comparable factor to the decrease in neuron density depicted in Figure 2. Thus, the rule of thumb for the visual system and most likely the entire brain is that decreasing neuron density progressing towards prefrontal cortex is complemented by major increases in neuronal connectivity. In short, neurons in TE are more sparse, but much larger, and intercommunicate to a much greater extent.

Neurophysiology and fMRI

Neurophysiology of intermediate ventral pathway areas was pioneered by Van Essen's group (Gallant, Braun, & VanEssen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996). These seminal studies showed that many V4 neurons are selectively responsive to concentric, radial, or hyperbolic gratings and are

less responsive to conventional sinusoidal gratings. This pioneering work was developed enormously by Pasupathy and Connor in an elegant series of papers (Pasupathy & Connor, 1999, 2001, 2002). Briefly, they showed that many V4 neurons were selectively sensitive to curvature extrema (usually convex) when the extremum was located at a particular position relative to the center of a closed curved object. Furthermore, they showed that the collection of neurons from which they recorded was capable of producing a population code for the shape of closed curved contours (Pasupathy & Connor, 2002).

Further support for the processing of curved shapes in V4 derives from human fMRI (Wilkinson et al., 2000). A first study compared the BOLD responses in V4 for conventional sinusoidal gratings to both concentric and radial gratings of the type first used by Gallant et al. (1993). Although all three grating types produced statistically indistinguishable activation in V1, concentric and radial gratings produced significantly stronger BOLD signals than sinusoidal gratings in V4. In addition, concentric gratings were the only ones to generate a significant BOLD signal in the fusiform face area (FFA), leading us to propose that circular or ellipsoidal shapes extracted in V4 might provide the basis for representing head shapes in FFA (Wilkinson et al., 2000). Subsequent electrophysiology supported this by showing that many neurons in macaque face patches will respond to circular shapes in addition to faces (Tsao, Freiwald, Tootell, & Livingstone, 2006).

Additional evidence for the representation of ellipsoidal shapes in FFA was provided by an fMRI study that compared BOLD responses to head shapes, internal features, and full faces (Nichols, Betts, & Wilson, 2010). Using multivoxel pattern analysis along with cross validation, the study showed that all three stimulus categories could be predicted at levels significantly above chance. This is consistent with configural pooling in V4 playing a role in the representation of head shapes in FFA. Primate neurophysiology supports this result by showing that some neurons in monkey face patches are selectively tuned for head aspect ratio (Freiwald, Tsao, & Livingstone, 2009).

Psychophysics and neural modeling

Key psychophysical results on intermediate level form processing have come from two lines of research in our laboratory, both triggered by primate V4 neurophysiology (Gallant et al., 1993). These are research on Glass patterns (Glass, 1969) and on radial frequency (RF) patterns (Wilkinson, Wilson, & Habak,

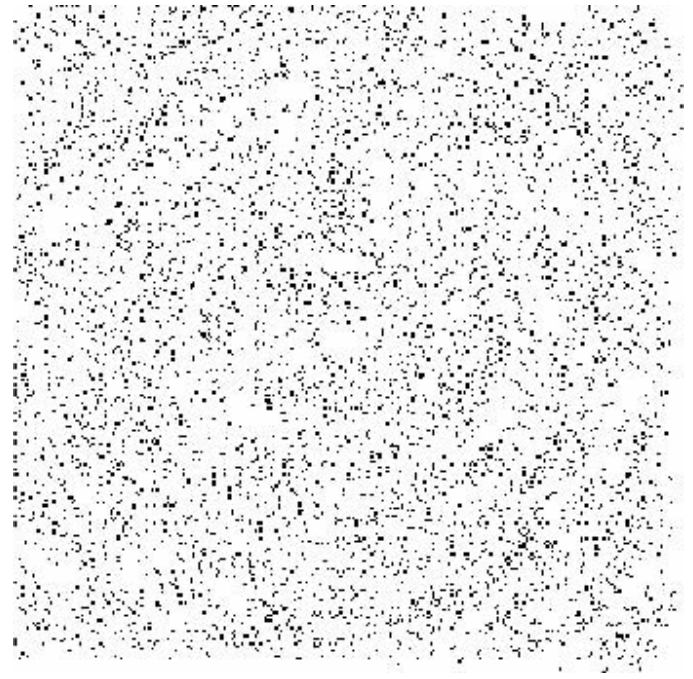


Figure 3. Concentric Glass pattern. Each dot is paired with a second dot that falls along a tangent to the underlying (invisible) concentric circle pattern.

1998). In both series of experiments a key focus was on how the visual system detects and discriminates patterns that are either circular or deviate from circularity by modest amounts. The focus on circularity derived from the observation that many natural biological forms approximate such structure, including human faces, many fruits, foliage of deciduous trees, eroded rocks, etc. Research on both types of patterns has led to similar conclusions, although there are differences which will require explanation.

Let us first consider Glass patterns, which are produced by positioning a pair of dots of fixed separation at a random position within the stimulus. To define a pattern, each dot pair is oriented on the tangent to an invisible contour defining the pattern. For a concentric Glass pattern, these are arcs of circles centered on the pattern origin (see Figure 3). For a parallel vertical pattern these would be parallel lines. Pattern detection thresholds are measured by determining how many dot pairs (signal) are required among a group of random pairs to discriminate the pattern from random noise, in which all dot pairs fall at random orientations. Two major results emerged from our studies. First, concentric Glass patterns have the lowest detection thresholds, whereas parallel patterns have the highest. Second, the data supported linear summation of orientation information along the circular contours of concentric patterns, but no analogous summation was found for parallel patterns (Wilson, Wilkinson, & Asaad, 1997). This difference in

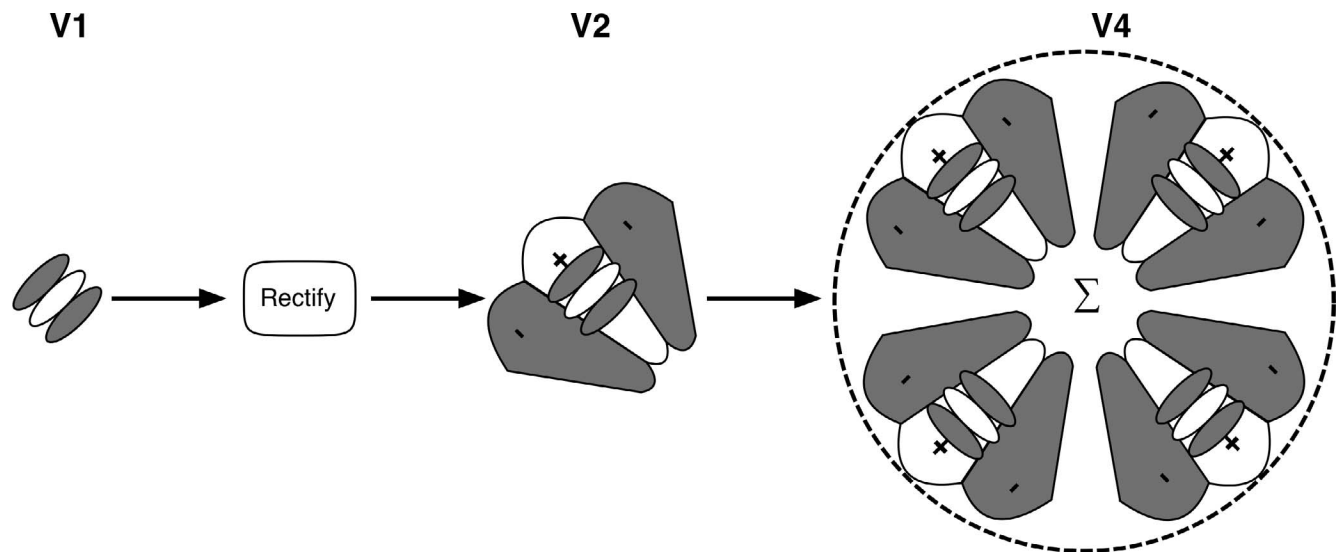


Figure 4. Neural model of V1, V2, and V4 processing to produce units sensitive to concentric structure in Glass patterns. Excitatory regions of receptive fields are plotted in white and inhibitory surrounds in gray. For more details, see text.

summation explained the difference in thresholds. This work was subsequently extended to radial Glass patterns as well with similar results (Wilson & Wilkinson, 1998). It is worth emphasis that the superior performance for concentric patterns has been corroborated by a number of psychophysical (Kelly, Bischof, Wong-Wylie, & Spetch, 2001; Kurki & Saarinen, 2004; Lestou, Lam, Humphreys, Kourtzi, & Humphreys, 2014; Seu & Ferrera, 2001), fMRI (Ostwald, Lam, Li, & Kourtzi, 2008), and visual evoked potential (Pei, Pettet, Vildaviski, & Norcia, 2005) studies. Even the use of oriented Gabor functions instead of dot pairs, which eliminates all orientation ambiguity at the first stage of Glass pattern processing, has provided evidence for superior performance with concentric patterns (Achtman, Hess, & Wang, 2003).

A neural model that accounts for the linear summation of orientation information to produce the low thresholds for concentric Glass patterns is depicted in Figure 4 (Wilson et al., 1997). This model employs processing by oriented V1 receptive fields at 12 different orientations followed by rectification and subsequent filtering by orthogonal receptive fields. This filter-rectify-filter process has been shown to effectively encode contour curvature over a considerable range (Dobbins, Zucker, & Cynader, 1987; Wilson, 1999). Specifically, the orthogonal orientation of the second filter creates an end-stopped mechanism that will not respond to elongated linear contours, but it responds excellently to curves with the appropriate tangent orientation. More complex, multiplicative second stage units have also been used in this model (Poirier & Wilson, 2006). As neurophysiological evidence cited above suggests that local contour curvature is encoded in V2 (Anzai et al., 2007; Hegdé & Van Essen, 2003), it

was proposed that this second model stage represents V2 processing. The final stage sums responses of V2 curvature units which are tangent to the pattern center (with thresholds so that negative responses are not included) to produce receptive fields that account for the data on detection of concentric Glass patterns. From the evidence cited above, this final pooling stage is consistent with V4 neurophysiology: Because these model V4 receptive fields are large, they exhibit substantially increased position invariance in their responses relative to earlier stages.

Although this is an initial model and hardly a final one of intermediate level processing in the ventral pathway, nevertheless it encapsulates much of the anatomy and physiology enumerated above. First, note that the receptive field size in V2 of the model must have approximately three times the diameter of the V1 oriented receptive field to effectively encode curvature (see inset in Figure 1). Second, the model V4 receptive fields must be roughly three times larger in diameter than their V2 inputs to effectively encode concentric structure: Circles don't exist at a point. Additionally, the larger receptive fields in model V2 and V4 can be effectively subsampled spatially into much smaller arrays consistent with the changes in neuron density in Figure 2 above. Finally, these larger receptive fields engender increased position invariance.

In parallel with this research on Glass patterns, we also introduced a category of smoothly curved closed shapes known as radial frequency or RF patterns (Wilkinson et al., 1998). These are defined in polar coordinates by a radius R that is a sinusoidal function of the polar angle θ :

$$R(\theta) = R_0[1 + A\sin(\omega\theta)]$$

where R_0 is the mean radius, ω is the integer radial frequency in cycles per 360° , and A is the amplitude of the deviation from circularity ($A = 0$ for a circle). Examples are shown in the accompanying article by Loffler (2015). Detection and recognition of these patterns are both in the hyperacuity range (Wilkinson et al., 1998). Subsequent work has shown that these patterns are processed globally for radial frequencies below about six cycles (Loffler, 2008; Loffler, Wilson, & Wilkinson, 2003). In addition, study of a patient with V4 damage has documented an enormous deficit in the ability to discriminate RF patterns from circles in his damaged V4 quadrant but not in intact quadrants (Gallant, Shoup, & Mazer, 2000).

Psychophysical data on RF patterns have been successfully modeled using an embellishment of the model illustrated in Figure 3 (Poirier & Wilson, 2006; Wilson & Wilkinson, 2014). In particular, RF masking studies (Habak, Wilkinson, Zakher, & Wilson, 2004), RF adaptation (Anderson, Habak, Wilkinson, & Wilson, 2007), and subthreshold summation between different RFs (Bell & Badcock, 2009) all support the presence of higher level channels each tuned to a different shape in the range RF2–RF6. These channels can be accommodated using a population code of V4 units sinusoidally weighted and pooled at a higher cortical level, perhaps TE0 or the human lateral occipital complex (Bell, Wilkinson, Wilson, Loffler, & Badcock, 2009; Wilson & Wilkinson, 2014).

Recent research suggests that the model in Figure 3 must be modified in order to accurately describe RF analysis by the visual system (Kempgens, Loffler, & Orbach, 2013; Schmidtmann, Gordon, Bennett, & Loffler, 2013). Using a 2D array of oriented Gabors, the Schmidtmann et al. (2013) study showed that thresholds for circles up to about RF4 or RF5 were defined by a constant number of signal elements so long as all signal elements were confined to a single radius or annulus. Further experiments showed that neither a large area Glass pattern detector like Figure 3 nor an association field model (Field, Hayes, & Hess, 1993) of local V1 interactions could explain the data. Rather, there must be multiple versions of the model in Figure 3, each constrained to pool over a narrow annulus. This advance in understanding the neural representation of RF patterns can easily be incorporated in the model by restricting the radial extent of the putative V2 receptive field to a range about equal to that of the V1 receptive fields that it pools, and incorporating multiple V4 units to sum V2 responses over different radii. See Loffler (2015) for further details. In addition to this, Kempgens et al. (2013) provided clear evidence that RF patterns of relatively large amplitude require both convex and concave curvature detectors for their representation. The model they presented to explain their data incorporated both convex and concave detectors based

on modifications of the Poirier and Wilson (2006) curvature detectors.

Discussion

The ventral pathway incorporates increasing receptive field diameter from area to area in a manner consistent with nearest neighbor pooling from the input area. This enables the system to grow from local orientation in V1 to curvature in V2 and closed curved shapes in V4. In addition, larger receptive fields permit spatial subsampling from area to area. However, the combination of larger receptive fields and subsampling alone imply that the number of neurons should be constant from area to area in the ventral pathway. This follows from the observation that subsampling by a given factor $1/N^2$ in space means that N^2 more neurons are required to encode the plethora of more complex features that are encoded via nearest neighbor pooling. The evidence that cortical neuron density decreases about six fold moving from V1 to TE (Cahalane et al., 2012) indicates that the dimensionality of shape representation must itself decrease from area to area. Recent calculations of the dimensionality of object representations in macaque TE versus the dimensionality of the input patterns themselves directly support the dimensionality reduction conclusion (Lehky et al., 2014).

Dimensionality reduction implies projection of the incoming information into a lower dimensional subspace, and we suggest two plausible ways in which this appears to be implemented in the ventral pathway. First, many areas of typical visual scenes are treated as textures, and there is psychophysical evidence that textures are represented by a small number of statistical properties, such as mean luminance, mean contrast, orientation distribution, etc. (Landy, 2014). This texture description represents an enormous dimensionality reduction compared to detailed representation of the precise location and orientation of each texture element. Indeed, the recent evidence that Glass pattern textures are encoded by mechanisms distinct from the more precise RF pattern mechanisms supports this distinction (Schmidtmann et al., 2013).

Thus far, dimensionality reduction via statistical representation of textures has been supported primarily by psychophysical evidence, so it is obviously desirable to obtain neurophysiological corroboration and elucidation of this. As a proposed step in this direction, macaques could be trained to make the same texture discriminations as humans. A study of fMRI could then be used to determine which areas are most effective for this discrimination, and this could be used to guide single unit electrophysiological recordings in

alert macaques. The prediction is that there would be a small number of cell types extracting texture properties. This approach to texture neurophysiology is analogous to the study of monkey face patches by Tsao et al. (2006).

Second, biologically important shapes are not random but rather show significant correlations among their parts. Principal component analysis (PCA) is one plausible way in which the correlation structure of objects can be used for dimensionality reduction, and there is now evidence that at least a few principal components are automatically learned when studying new faces (Gao & Wilson, 2014). Independent component analysis (ICA) can provide similar dimensionality reduction benefits (Bartlett, Movellan, & Sejnowski, 2002; Draper, Back, Bartlett, & Beveridge, 2003). It remains for future research to determine whether dimensionality reduction via statistical representation of textures and PCA or ICA representation of object correlations by themselves suffice to generate the dimensionality reduction that is implied by neuron density reductions. However, biologically based models of visual pattern recognition have shown that subsampling and dimensionality reduction of this sort can be quite powerful (Osadchy, LeCun, & Miller, 2007).

It was mentioned at the beginning that V1 contains an overcomplete representation relative to the retinal input. Thus, it is natural to question whether this representation might aid the dimensionality reduction process in higher areas. It is known that different subsets of V1 neurons process motion direction, contour orientation, and disparity, and each of these projects in parallel to different higher level areas (e.g., MT for motion, V4 for orientation and color, etc.). Thus, we conjecture that dimensionality reduction can be most effectively accomplished in different ways for different visual attributes. Obviously, further theoretical and experimental work is required.

One major element missing from this ventral pathway scenario is any role for the ubiquitous feedback connections among areas (VanEssen et al., 1992). One possibility is that low spatial frequency information moves rapidly up the ventral pathway to generate a neural “hypothesis” about the object category: face, quadruped, house, etc. This information would then be fed back to lower areas to enhance more detailed processing (Bar, 2007; Bar et al., 2006). The same circuitry could also be used in top-down selective attention.

A second missing ingredient is a role for the “skipping connections” that bypass an area to connect with the next higher one (Nakamura et al., 1993). This would speed processing while leading to diminished precision. One untested conjecture is that it might be low frequency information that is rapidly and crudely conveyed by skipping connections, which is consistent

with ideas of Bar and colleagues above. Indeed, based on the Nyquist theorem, vastly fewer spatial samples are required to represent very low spatial frequency information, so skipping connections may result in part from fewer processing demands at low spatial frequencies.

Finally, it should be emphasized that most of the data discussed here relate primarily to foveal vision. As roughly the central 5.0° of the visual field occupies about 50% of striate cortex, a focus on central vision certainly seems appropriate. However, one can ask whether the same principles operate to reduce dimensionality in the periphery. Although much further work is needed, it is reasonable to conjecture that the cortical representation of peripheral vision is even more heavily focused on representation of regions by texture statistics. Furthermore, perhaps projection onto very few principal components, a number just sufficient for basic categorization, may operate in the peripheral representation in cortex. These possibilities await elucidation by future research.

Here we have primarily dealt with the curved structure inherent in radial frequency patterns, which is clearly related to V4 physiology (Pasupathy & Connor, 2001, 2002). Quite obviously, many visual shapes include angular structure, such as most buildings, jagged rocks, most chairs and tables, etc. A beginning has been made with the recent introduction of angular frequency patterns, which are the angular analog of RF patterns (Wilson & Propp, in press). Much additional research is clearly required on such angular patterns and on patterns combining angles with curves. A promising fMRI approach to this has shown that responses in intermediate and higher ventral pathway areas are highly correlated with the angled and curved structure of the objects themselves (see Andrews, Watson, Rice, & Hartley, 2015). We believe that the observations and models for curved shapes presented here will inform and illuminate these future studies.

Keywords: ventral pathway, V4, form perception

Acknowledgments

This work was supported in part by CIHR grant #172103 to the authors, NSERC grants OGP0007551 to FEW, OPG227224 to HRW, and a CIFAR grant to HRW.

Commercial relationships: none.

Corresponding author: Hugh R. Wilson.

Email: hrwilson@yorku.ca.

Address: Centre for Vision Research, York University, Toronto, Ontario, Canada.

References

- Achtman, R. L., Hess, R. F., & Wang, A. (2003). Sensitivity for global shape detection. *Journal of Vision*, 3(10):4, 616–624, <http://www.journalofvision.org/content/3/10/4>, doi:10.1167/3.10.4. [PubMed] [Article]
- Anderson, N. D., Habak, C., Wilkinson, F., & Wilson, H. R. (2007). Evaluating shape aftereffects with radial frequency patterns. *Vision Research*, 47, 298–308.
- Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision*, 15(7):3, 1–12, <http://www.journalofvision.org/content/15/7/3>, doi:10.1167/15.7.3. [Article]
- Anzai, A., Peng, X., & VanEssen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, 10, 1313–1321.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11, 280–289.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, USA*, 103, 449–452.
- Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Trans. Neural Networks*, 13, 1450–1464.
- Bell, J., & Badcock, D. R. (2009). Narrow-band radial frequency shape channels revealed by sub-threshold summation. *Vision Research*, 49(8), 843–850.
- Bell, J., Wilkinson, F., Wilson, H. R., Loffler, G., & Badcock, D. R. (2009). Radial frequency adaptation reveals interacting contour shape channels. *Vision Research*, 49, 2306–2317.
- Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *Journal of Comparative Neurology*, 306, 554–575.
- Cahalane, D. J., Charvet, C. J., & Finlay, B. L. (2012). Systematic, balancing gradients in neuron density and number across the primate isocortex. *Frontiers in Neuroanatomy*, 6, 28, doi:10.3389/fnana.2012.00028.
- Chubb, C., Landy, M. S., & Econopouly, J. (2004). A visual mechanism tuned to black. *Vision Research*, 44, 3223–3232.
- Diamantaras, K. I., & Kung, S. Y. (1996). *Principal component neural networks*. Toronto, Ontario: John Wiley.
- Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329, 438–441.
- Draper, B. A., Back, K., Bartlett, M. S., & Beveridge, J. R. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, 91, 115–137.
- Elston, G. N. (2002). Cortical heterogeneity: Implications for visual processing and polysensory integration. *Journal of Neurocytology*, 31, 317–335.
- Elston, G. N., & Rosa, M. G. P. (1998). Morphological variation of layer III pyramidal neurones in the occipitotemporal pathway of the macaque monkey visual cortex. *Cerebral Cortex*, 8, 278–294.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local “association field.” *Vision Research*, 33, 173–193.
- Freiwald, W. A., Tsao, D. T., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, 12, 1187–1196.
- Gallant, J. L., Braun, J., & VanEssen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, 259, 100–103.
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76, 2718–2739.
- Gallant, J. L., Shoup, R. E., & Mazer, J. A. (2000). A human extrastriate cortical area that is functionally homologous to macaque area V4. *Neuron*, 27, 227–235.
- Gao, X., & Wilson, H. R. (2014). Implicit learning of geometric eigenfaces. *Vision Research*, 99, 12–18.
- Glass, L. (1969). Moiré effect from random dots. *Nature*, 223, 578–580.
- Habak, C., Wilkinson, F., Zakher, B., & Wilson, H. R. (2004). Curvature population coding for complex shapes in human vision. *Vision Research*, 44, 2815–2823.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A. S., J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- Hegd e, J., & Van Essen, D. C. (2003). Strategies of

- shape representation in macaque visual area V2. *Visual Neuroscience*, *20*, 313–328.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face recognition. *The Journal of Neuroscience*, *17*, 4302–4311.
- Kelly, D. M., Bischof, W. F., Wong-Wylie, D. R., & Spetch, M. L. (2001). Detection of Glass patterns by pigeons and humans: Implications for differences in higher-level processing. *Psychological Science*, *12*, 338–342.
- Kempgens, C., Loffler, G., & Orbach, H. S. (2013). Set-size effects for sampled shapes: Experiments and model. *Frontiers in Computational Neuroscience*, *7*, 67.
- Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–867.
- Kurki, I., & Saarinen, J. (2004). Shape perception in human vision: Specialized detectors for concentric spatial structures? *Neuroscience Letters*, *360*, 100–102.
- Landy, M. S. (2014). Texture analysis and perception. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 639–652). Cambridge, MA: MIT Press.
- Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, C. (2014). Dimensionality of object representations in monkey inferotemporal cortex. *Neural Computation*, *26*, 2135–2162.
- Lestou, V., Lam, J. M. L., Humphreys, K., Kourtzi, Z., & Humphreys, G. W. (2014). A dorsal visual route necessary for global form perception: Evidence from neuropsychological fMRI. *Journal of Cognitive Neuroscience*, *26*, 621–634.
- Leuba, G., & Kraftsik, R. (1994). Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. *Journal of Anatomy and Embryology*, *190*, 351–366.
- Loffler, G. (2008). Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research*, *48*, 2106–2127.
- Loffler, G. (2015). Probing intermediate stages of shape processing. *Journal of Vision*, *15*(7):1, 1–19, <http://www.journalofvision.org/content/15/7/1>, doi:10.1167/15.7.1. [Article]
- Loffler, G., Wilson, H. R., & Wilkinson, F. (2003). Local and global contributions to shape discrimination. *Vision Research*, *43*, 519–530.
- Miller, D. J., Balaram, P., Young, N. A., & Kaas, J. H. (2014). Three counting methods agree on cell and neuron number in chimpanzee primary visual cortex. *Frontiers in Neuroanatomy*, *8*(36), 1–11.
- Nakamura, H., Gattass, R., Desimone, R., & Ungerleider, L. G. (1993). The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *The Journal of Neuroscience*, *13*, 3681–3691.
- Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology*, *1*(28), 1–13.
- O’Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*, 580–590.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
- Op de Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, *426*, 505–518.
- Osadchy, M., LeCun, Y., & Miller, M. L. (2007). Synergistic face detection and pose estimation with energy-based models. *Journal of Machine Learning Research*, *8*, 1197–1215.
- Ostwald, D., Lam, J. M., Li, S., & Kourtzi, Z. (2008). Neural coding of global form in human visual cortex. *Journal of Neurophysiology*, *99*, 2456–2469.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82*, 2490–2502.
- Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, *86*, 2505–2519.
- Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*, 1332–1338.
- Pei, F., Pettet, M. W., Vildaviski, V. Y., & Norcia, A. M. (2005). Event related potentials show configural specificity of global form processing. *NeuroReport*, *16*, 1427–1430.
- Poirier, F. J., & Wilson, H. R. (2006). A Biologically Plausible Model of Human Radial Frequency Perception. *Vision Research*, *46*, 2443–2455.
- Rodieck, R. W. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer Associates.
- Schmidtman, G., Gordon, G. E., Bennett, D. M., & Loffler, G. (2013). Detecting shapes in noise: Tuning characteristics of global shape mechanisms.

- Frontiers in Computational Neuroscience*, 7(37), 1–14.
- Seu, L., & Ferrera, V. P. (2001). Detection thresholds for spiral glass patterns. *Vision Research*, 41, 3785–3790.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311, 670–674.
- VanEssen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, 255, 419–423.
- Wang, L., Mruczek, R. E., Acaro, M. J., & Kastner, S. (2014). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, 1–21.
- Wilkinson, F., James, T. W., Wilson, H. R., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). Radial and concentric gratings selectively activate human extrastriate form areas: An fMRI study. *Current Biology*, 10, 1455–1458.
- Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, 38, 3555–3568.
- Wilson, H. R. (1999). Non-Fourier cortical processes in texture, form, and motion perception. In P. S. Ulinski & E. G. Jones (Eds.), *Cerebral cortex, 13: Models of cortical circuitry* (pp. 445–477). New York: Plenum.
- Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23, 873–882.
- Wilson, H. R., & Propp, R. (in press). Detection and recognition of angular frequency patterns. *Vision Research*.
- Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. *Vision Research*, 38, 2933–2947.
- Wilson, H. R., & Wilkinson, F. (2014). Configural pooling in the ventral pathway. In J. S. Werner & L. Chalupa (Eds.), *The new visual neurosciences* (pp. 617–626). Cambridge, MA: MIT Press.
- Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, 37, 2325–2330.