Modeling the effect of cortical magnification on feature detection and individuation

Rachel F. Heaton (rfheaton@illinois.edu)
School of Art + Design, University of Illinois Urbana-Champaign, 408 E. Peabody Dr.
Champaign, IL 61820 USA

John E. Hummel (jehummel@illinois.edu)
Department of Psychology, University of Illinois Urbana-Champaign, 603 E. Daniel St.
Champaign, IL 61820USA

Abstract

Some researchers have argued that visual representations in the periphery differ qualitatively from those in the fovea (e.g., Balas et al., 2009; Freeman and Simoncelli, 2011; Rosenholtz et al., 2012). Consistent with this proposal, He et al. (1997) showed that crowding in the periphery disrupts the ability to individuate features but doesn't disrupt feature detection. We hypothesized that He et al.'s demonstration could be accounted for simply in terms of cortical magnification alone. We tested this hypothesis by presenting He et al.'s stimuli to a neurally-realistic model of V1 (Heaton & Hummel, 2022) that incorporates cortical magnification but posits no other differences between foveal and peripheral early visual representations. The model's performance captured He et al.'s findings, suggesting that cortical magnification alone is sufficient to account for the differences between foveal and peripheral visual perception.

Keywords: V1, striate cortex, cortical magnification, fovea, peripheral processing, summary statistics, visual representation

Introduction

Are the representations and processes in peripheral vision qualitatively different from those of central vision? Rosenholtz et al. (2012) observed that in some circumstances features in peripheral vision could be detected, but not precisely localized and individuated. For example, all the letters in a word might be detected, but their precise arrangement might be lost. Similarly, He, Intriligator, and Cavanagh (1997) demonstrated that crowding in the periphery disrupts the ability to individuate features but doesn't disrupt feature detection itself (Figure 1). They interpreted this effect in terms of a dissociation between spatial and attentional resolution. The texture tiling model uses summary statistics over pooling regions to account for peripheral crowding phenomena (Rosenholtz et al., 2012; see also Balas, 2006, and Balas et al., 2009). The TTM postulates that representations in peripheral vision are generated over large pooling regions. Like Freeman & Simoncelli's (2011) metamer model of mid-ventral processing, TTM representations over pooling regions are hypothesized to consist of encodings of large sets of summary statistics similar to the type proposed by Portilla & Simoncelli (2000).

In these texture-based models, a compressed statistical summary representation is fundamental to account for information loss, specifically localization, in peripheral vision. The idea that there may be qualitative differences in processing between the fovea and the periphery has gained traction, particularly in the visual search literature (see Hulleman & Olivers, 2017, and Wolfe, 2021). Freeman and Simoncelli's (2011) model and TTM (Rosenholtz, 2012) postulate both large pooling regions in the periphery and qualitatively different representations than those found in central vision, and they attribute the limitations of discrimination in peripheral vision to both factors. However, large receptive fields alone may be sufficient to account for these phenomena. The larger a receptive field is, the more likely it is to have many features within it, and the information about the location of those features within the receptive field is lost in the neural encoding. This information loss regardless of the type of representation that is being postulated, whether that is a summary statistical representation or whether it is a V1 receptive field. Therefore, we questioned whether the phenomena these models seek to explain require an appeal to qualitatively different kinds of representations in the periphery vs. central vision, or whether, as others have suggested, they could be accounted for more simply in terms of cortical magnification alone.

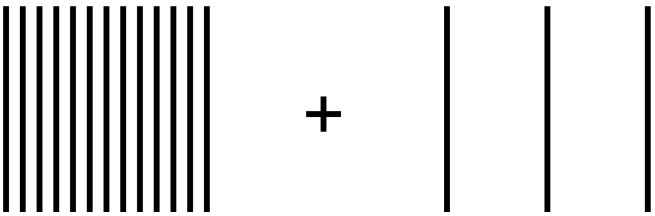

Figure 1. Recreation of the stimulus by He et al. (1997) which shows that it's possible to individuate widely spaced lines in the periphery, but difficult to individuate closely spaced lines.

Cortical magnification is a known property of primate visual cortex, in which information in the center of the visual field is coded by more neurons in visual cortex than information in the periphery, and receptive field sizes increase with eccentricity in the visual field. If cortical magnification is sufficient to account for these phenomena, then it may not be necessary to postulate differences between attentional and spatial resolution. However, if cortical magnification cannot account for these phenomena, then that would be consistent with claims that there are qualitatively different representations in the periphery, a finding that

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

would have implications for our understanding of human vision, as well as for the development of bio-inspired computer vision models.

Methods

To investigate the effect of cortical magnification on low-level feature detection and individuation, we adapted Heaton and Hummel's (2022) neurally realistic computational model of primary visual cortex (i.e., V1) and explored the effects of crowding in the periphery vs. foveal vision in that model.

Overview of the V1 model's architecture

In the research reported here, the visual field of the V1 model fits within 1600x1600 pixels, with 1024 pixels corresponding to 30 degrees of visual angle (34.13 pixels per degree). Like the primate visual system, the V1 model consists of neurons organized into hypercolumns. Each hypercolumn processes the information in a small circular receptive field, centered at the hypercolumn's location. The hypercolumns are arranged in a hexagonal lattice such that the neurons' receptive fields overlap and collectively cover the entire visual field. The neurons within a hypercolumn filter the image for conjunctions of color, orientation, and phase information. The model processes multiple spatial scales simultaneously. There is one set of hypercolumns per spatial scale, and the locations of the hypercolumns within a spatial scale are calculated independently of other spatial scales. The number of spatial scales the model uses is configurable via a set of parameters, which in these simulations was three. The receptive field radii in the center of the visual field were 10, 14, and 18 pixels, corresponding to 0.6°, 0.85°, & 1.1° of visual angle for the small, medium, and large scales, respectively. These values were chosen based on Macaque single unit recording data from Gattass et al. (1981) and human MRI population receptive field sizes from Dumoulin & Wandell (2008).

Cortical Magnification

Cortical magnification in the V1 model is implemented as a linear increase in both receptive field radius and the spacing between adjacent hypercolumns with increasing eccentricity in the visual field (Figures 2 and 3). In the model, the cortically magnified hexagonal lattice of hypercolumn locations is calculated for each spatial scale, *s,* individually. The hypercolumn locations for a spatial scale are determined by first calculating the distances from the center of the visual field to the vertices of a series of concentric hexagons, *i*. The distance, d_i, of any vertex of hexagon *i* from the center of the visual field is determined by the equations

$$d_i = d_{i\text{-}1} + \Delta d_i \,, \tag{1}$$

$$\Delta d_i = r_0^s \,(\zeta + \mu i r_0^s) \,, \tag{2}$$

where r_0^s is the receptive field radius at scale *s* in the hypercolumn in the center of the visual field and ζ is the fraction of a hypercolumn radius that separates the centers of adjacent hypercolumns. In these simulations, ζ = 0.5. The receptive field radius, r_i^s, in hexagon *i* at scale *s* is given by

$$r_i^s = r_0^s(1+ \mu i) \,, \tag{3}$$

where μ=0.004583 is the rate of cortical magnification calculated from the empirical data from Gattass et al. (1981) and Doumalin and Wandell (2008). Once the six vertices on the *i*th concentric hexagon are calculated, *i*-1 additional points are interpolated between each pair of adjacent vertices on the hexagon.

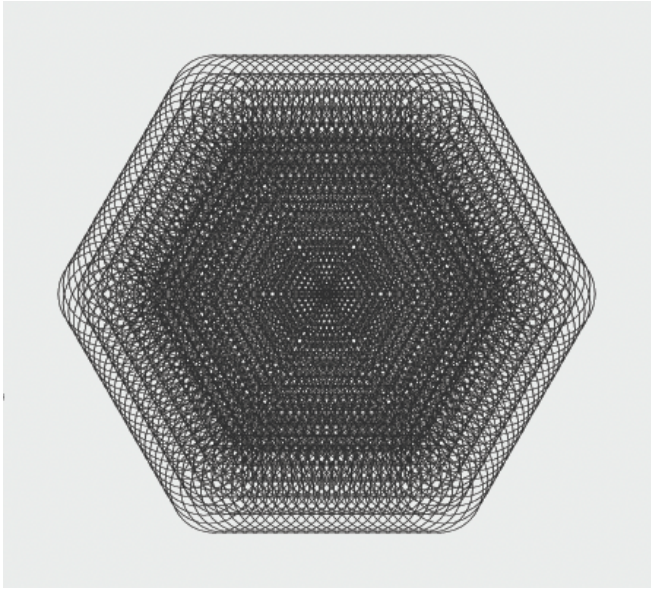

Figure 2. A visualization of the hexagonal lattice of hypercolumn locations for one spatial scale.

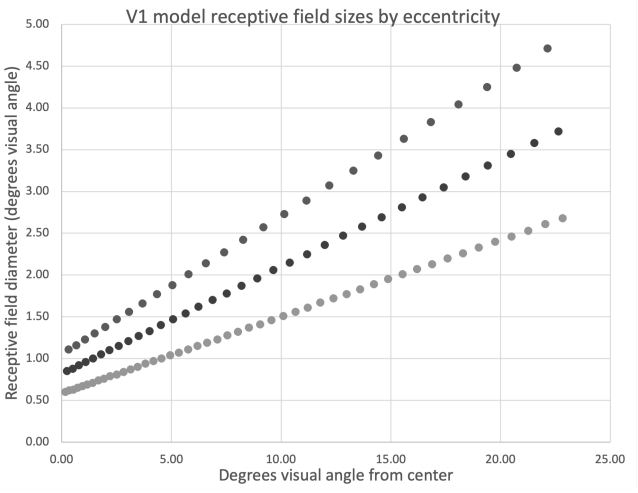

Figure 3. The V1 model's receptive field sizes for all scales plotted as a function of eccentricity.

Filtering neurons

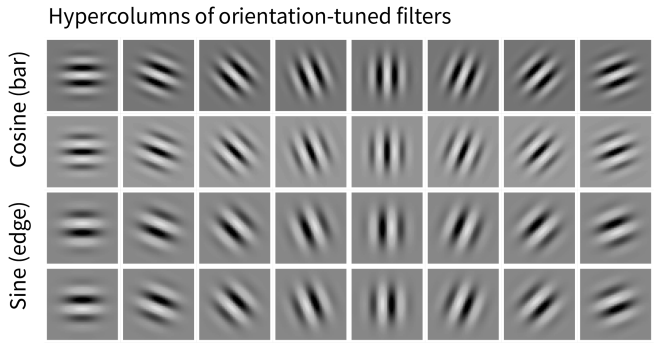

Figure 4. The set of orientation-tuned bar and edge filters for one spatial scale in the V1 model. Black is a value less than zero, white is a value more than zero, and middle gray is near zero. Each color channel has a set of filter neurons.

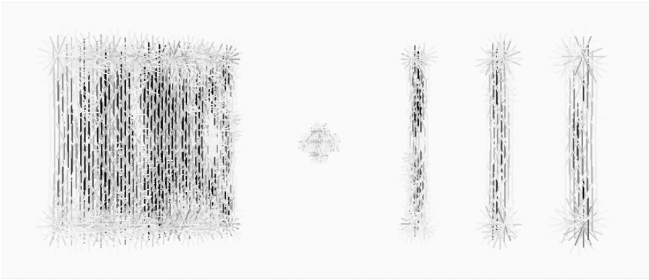

Figure 5. A visualization of the model's representations in response to the He et al. (1997) stimuli for the smallest spatial scale. Darker lines indicate more active filter neurons, and the orientation of the line corresponds to the filter's preferred orientation.

At each hypercolumn location, neurons filter the image for color, orientation, and phase information. This information is conjunctively coded by 96 neurons per hypercolumn that code for combinations of 3 color channels, 8 orientations, 2 phases (sine and cosine), and 2 multiplicative signs on the filters (e.g., a positively signed cosine filter would be white in the center and black in the surround, while the negatively signed filter would be black in the center and white in the surround). The model has red-green, blue-yellow, and white-black opponent color channels, corresponding to the opponent channels in the human visual system. Cosine filters find bars (e.g., a black line surrounded by white), while sine filters find edges (e.g., black on the left and white on the right). The number of cycles per receptive field radius for the sinusoids is defined by a parameter. In these simulations, the number of cycles per receptive field for sine filters was 4, and the number of cycles per receptive field for cosine filters was 5. Each sinusoid is damped by a gaussian envelope with σ=0.3 times the receptive field radius r_i^s. After damping, each filter is normalized so that the sum of positive and negative values in the filter sums to zero. The orientations of the filters range from 0 to $7\pi/8$ radians in increments of $\pi/8$ (see Figure 4). See Figure 5 for a visualization of the model's representations in response to the original He et al. (1997) stimuli.

Simulations

We presented the model modified versions of the peripheral spatial grating stimulus from He et al. (1997), centered either at fixation or 7.7 degrees left of fixation (Figure 6). We presented a single vertical line in the fovea of model's visual field and encoded the resulting pattern of activation as a collection of three vectors (one for each spatial scale). We repeated the same procedure in the periphery. We then showed the model widely spaced flanking vertical lines at in each location, and finally closely spaced (crowded) flanking vertical lines at each location. We again recorded the model's response to each flanking stimulus as a collection of three vectors. The measure of interest was the degree of disruption of the original vector representation of the single line by the flanking lines in each case. We hypothesized that the widely spaced flankers would have little effect on the representation of the center line in either central or peripheral vision. By contrast, due to the large receptive fields in the periphery, we hypothesized that closely spaced flankers would interfere with the representation of the center line in the periphery much more than in the fovea.

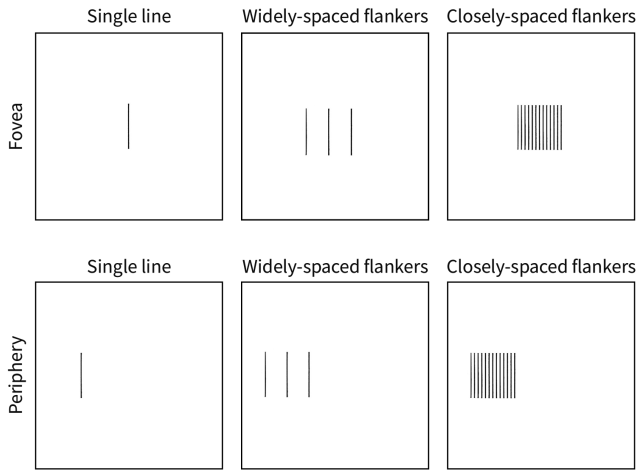

Figure 6. Single vertical lines, widely-spaced flankers, and closely-spaced flankers presented in the model's fovea and periphery.

To quantify the disruption (i.e., interference) from a given contour segment, *j*, on the representation of a nearby segment, *i*, we first encoded *i* in isolation as a *Gaussian radial basis function* (GRBF), which Poggio & Girosi (1989) demonstrated is an optimal classifier under a wide variety of circumstances. In brief, the vector, **i**, of filter outputs generated by contour segment *i* served as the mean of $G(\mathbf{i})$, the GRBF encoding segment *i*. The fit, $G(\mathbf{i},\mathbf{k})$, of any filter vector, **k**, to **i** was calculated as the height of a high

dimensional Gaussian, $G(\mathbf{i})$, with mean, $\mu_{\mathbf{i}} = \mathbf{i}$ and standard deviation, $\sigma = 1.0$:

$$G(\mathbf{i}, \mathbf{k}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-0.5\left(\frac{d(i,k)}{\sigma}\right)^2}, \quad (4)$$

where $d(\mathbf{i}, \mathbf{k})$ is the Euclidean distance between **i** and **k**. Let vector **k** be the sum of vectors **i** and **j**, representing segments *i* and *j* respectively. To the extent that segment *j* activates any of the same neurons as segment *i*, then vector elements in **i** will change in the presence of **j**, relative to their values when **i** is presented in isolation. As a result, $G(\mathbf{i},\mathbf{k})$ will be less than $G(\mathbf{i},\mathbf{i})$ because the fit to the GRBF representing segment *i* will be reduced in the presence of *j*, relative to when *i* is presented alone. By contrast, if *j* does not activate any of the same filters activated by *i*, then $G(\mathbf{i},\mathbf{k})$ will equal $G(\mathbf{i},\mathbf{i})$: The vector representing segment *i* will be unaffected by the presence of *j*, so **k** (the sum of **i** and **j**) will, from the perspective of $G(\mathbf{i})$, be equivalent to **i**, and $G(\mathbf{i},\mathbf{k})$ will be equal to $G(\mathbf{i},\mathbf{i})$. In the simulation results that follow, all GRBF values have been normalized to the range 0…1 by dividing by all $G(\mathbf{i},\mathbf{k})$ by $G(\mathbf{i},\mathbf{i})$, the fit of vector **i** to the GRBF encoding it, which is the height of the Gaussian at its mean, **i**.

We computed the disruption (i.e., the difference between $G(\mathbf{i},\mathbf{k})$ and $G(\mathbf{i},\mathbf{i})$) at each spatial scale and also across all spatial scales combined.

The model code, stimuli, and GRBF script are available at: https://github.com/rachelfheaton/V1_crowding_GRBF

Results

The addition of widely spaced flankers had no effect on the representation of a single vertical line anywhere in the visual field, which is consistent with the He et al. (1997) demonstration (see Table 1).

The effect of closely spaced flankers was greater in the periphery than in the fovea at every spatial scale, as well as across all spatial scales (Table 2). When all the spatial scales were combined, the disruption from closely spaced flankers was more than twice as large in the periphery as in the fovea. In the fovea, the representation in the finest spatial scale was almost entirely unperturbed by the addition of the closely spaced flankers. Given that people can selectively attend to specific spatial frequencies (Baas et al., 2002), an observer who attended primarily to the finest spatial scale would have no difficulty differentiating the close flankers in foveal vision. Even in foveal vision, however, the middle spatial scale was somewhat affected by proximity and the largest spatial scale was impacted almost as it was in the periphery. The difference between the disruption in the periphery and that in the fovea was smaller with increasing spatial scale. At the smallest spatial scale, the GRBF disruption in the periphery was about 42 times the disruption in the fovea, but at the largest spatial scale the disruption in the periphery was only about 1.22 times larger than in the fovea.

Table 1: GRBF disruption, $G(\mathbf{i}, \mathbf{i}) - G(\mathbf{i},\mathbf{k})$, with widely spaced flankers.

Spatial scale	Fovea	Periphery
Small	0.0000	0.0000
Medium	0.0000	0.0000
Large	0.0000	0.0000
All scales	0.0000	0.0000

Table 2: GRBF disruption, $G(\mathbf{i},\mathbf{i}) - G(\mathbf{i},\mathbf{k})$, with closely spaced flankers.

Spatial scale	Fovea	Periphery
Small	0.0031	0.1290
Medium	0.0583	0.1225
Large	0.1418	0.1732
All scales	0.0606	0.1354

Discussion

We used a neurally realistic model of V1 to investigate the effects of crowding on foveal and peripheral visual representations and found that the phenomenon pointed out by He et al. (1997), as well as Balas (2006), Balas et al. (2009), and Rosenholtz et al. (2012), which is that it is difficult to individuate some features in the periphery but that it is not difficult to know what they are, arises as a natural consequence of cortical magnification. As a result, it is not necessary to appeal to either different spatial and attentional resolutions as postulated by He et al., or to lossy compressed peripheral representations like those postulated by Balas (2006), Balas et al. (2009), Freeman and Simoncelli (2011) and Rosenholtz et al. (2012).

Our simulation results suggest that is possible to account for effects like those reported by Balas, He, Rosenholtz, and others, in terms of the concept of *entanglement*, which is well understood in the neural modeling community as a failure of independence in representation.

Entanglement is typically discussed in terms of how neurons respond to different properties of the same stimulus: To the extent that the neural representation of one stimulus property is affected by other properties of the stimulus, those properties are entangled. For example, if a neuron responds to a conjunction of phase and orientation (like the neurons in V1), then phase and orientation are entangled in that representation because the neuron's response depends on both. Conversely, to the extent that the representation of one stimulus property is unaffected by another property, the two are disentangled.

The problem of entanglement, which in the example above concerns the entanglement of the phase and orientation of the same stimulus, also impacts the neural representation of different stimuli. To the extent that a neural architecture cannot discriminate the properties of one stimulus (e.g., one vertical line) from the properties of another stimulus (e.g.,

another vertical line), the two representations become an entangled, undifferentiable "whole".

We argue that this is precisely what is happening in peripheral vision with He et al.'s (1997) closely spaced lines. Peripheral vision is not recoding the lines into a reduced statistical representation a la Rosenholtz et al., (2012) and Freeman and Simoncelli (2011). Rather, the neural representation of the lines in the periphery is simply too entangled to differentiate one line from another. The neural representation of one line overlaps so much with the neural representation of another that it is impossible to tell where one line stops and the next starts (see Figure 5).

At the same time, those same neural representations are sufficient to tell that all the lines are roughly vertical. What tells you whether the line you are looking at is vertical or horizontal (i.e., what features are present) is the orientation selectivity of the neurons responding to the stimulus. As is evident in Figure 5, the orientation selectivity of the responses to all the vertical lines in the stimulus is similar in the closely spaced (crowded) case to the orientation selectivity in the widely spaced case. In other words, the model is equally able to detect that the stimuli are vertical in both cases, just as people are. By contrast, what enables you to individuate one object from another is perceptual grouping, which is known to be affected by proximity.

In the fovea, there is not as much overlap in the receptive fields responding to the vertical lines, and as a result the representation of a crowded stimulus in the fovea is similar to the representation of a widely spaced stimulus in the periphery. In both cases, there is enough separation between the neurons responding to the individual lines to individuate (i.e., perceptually segment) them.

Our simulation results are not definitive about whether visual representations in the periphery really are qualitatively different from those in the fovea, but they demonstrate that one does not have to postulate such qualitative differences to account for findings like those reported by He et al. (1997) and Rosenholtz et al (2012).

Acknowledgments

This research was supported by AFOSR Grant AF-FA9550-12-1-003. This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) the State of Illinois, and as of December, 2019, the National Geospatial-Intelligence Agency. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. This work was partly supported by the National Science Foundation under Grant No BCS1921735. The authors thank Alejandro Lleras and Simona Buetti for mentioning the theories of Rosenholtz et al. (2012) and He et al. (1997).

References

Baas, J. M. P., Kenemans, J. L., & Mangun, G. R., Selective attention to spatial frequency: an ERP and source localization analysis. Clinical Neuropsychology, 113 (11), 1840-1854.

Balas, B. J. (2006). Texture synthesis and perception: Using computational models to study texture representations in the human visual system. Vision research, 46(3), 299-309.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. Journal of vision, 9(12), 13-13.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. Neuroimage, 39(2), 647-660.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. Nature neuroscience, 14(9), 1195-1201.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. Journal of Comparative Neurology, 201(4), 519-539.

He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. Trends in cognitive sciences, 1(3), 115-121.

Heaton, R., & Hummel, J. (2022). A computational model of binding by temporal synchrony in visual area V1. Journal of Vision, 22(14), 3586-3586.

Hulleman, J., & Olivers, C. N. (2017). The impending demise of the item in visual search. Behavioral and Brain Sciences, 40, e132.

Poggio, T., & Girosi, F. (1990). Networks for approximation and learning. Proceedings of the IEEE, 78(9), 1481-1497.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. International journal of computer vision, 40, 49-70.

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012a). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. Frontiers in psychology, 3, 13.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. Psychonomic bulletin & review, 28(4), 1060-1092.