

A New Look at Cat Simple Cells

C. H. Anderson, Brandon Westover, and Greg DeAngelis
Dept. of Anatomy and Neurobiology
Washington University School of Medicine
St. Louis, MO 63110

Introduction:

This paper summarizes the results of applying our engineering framework for modeling neuronal systems, summarized in a recent book (Eliasmith and Anderson 2003), to the early stages of visual processing. The central assumption of this framework is that neuronal systems utilize a population of neurons to represent analog signals. In the case of simple cells, these would be the time dependent coefficients of a wavelet like representation of the image falling on the retina. While most modeling efforts accept this hypothesis, their focus is generally on either the development of a detailed low level neurobiological model to fit a particular experimental result, or the exploration of the properties of the wavelet transformation of images. In contrast, our goal is to build a model that bridges these two extremes by building a neuronal implementation of a high level signal processing model. The process of pursuing this goal has revealed a remarkable number of new properties of simple cells.

Statistical Model of Images:

The modern signal processing approach to visual processing is to define it as a problem in statistical inference. At a very abstract level this amounts to defining all the possible objects, their pose and configurations, as well as the lighting conditions and the projection of light through the optics of the lens to form the image on the retina. The problem of the system is then to estimate those objects, as well as their pose and position, that generated the observed image. At a much lower level of abstraction we can start with Olshausen and Field's (1996) statistical analysis of static images that shows to first order they can be modeled as a sparse set of coefficients A_n of an overcomplete, wavelet like representation

$$I(\mathbf{r}) = \sum_n A_n \exp(i\mathbf{k}_n \cdot \mathbf{r}) \exp(-(\mathbf{r} - \mathbf{r}_n)^2/(2\sigma_{\mathbf{r}}^2)), \quad (1)$$

where \mathbf{k}_n is what neuroscientists call the preferred spatial frequency, \mathbf{r}_n is the center of the receptive field, and $\sigma_{\mathbf{r}}^2$ the spatial variance of the receptive field. The amplitudes A_n form a set of stochastic variables that are to first order statistically independent and have a highly kurtotic probability distributions. A simple model of the early stages of the vision then begins by assuming that a particular static image $I(\mathbf{r}, \mathbf{A})$ is created by generating a set of amplitudes $\mathbf{A} = A_0, \dots, A_N$. This image is then detected, compressed by the retina, and transmitted down the optic nerve by the retinal ganglion cells. The role of the simple cells is then to generate estimates of these coefficients $\hat{\mathbf{A}}$ from the ganglion cell inputs. The activity of the population of simple cells are then presumed to contribute to the representation of these coefficients. Specifically, we define each neuronal response as a

nonlinear, rectification $\mathcal{G}_i[\cdot]$ of the output of a wavelet-like spatial filter applied to the image

$$a_i(I(\mathbf{r}, \mathbf{A})) = \mathcal{G}_i[\langle \exp(i\mathbf{k}_i \cdot \mathbf{r}) \exp(-(\mathbf{r} - \mathbf{r}_i)^2/(2\sigma_{\mathbf{r}}(i)^2)) I(\mathbf{r}, \mathbf{A}) \rangle_{\mathbf{r}}]. \quad (2)$$

We also claim one can obtain an estimate of the $\hat{\mathbf{A}}$ by using a linear population vector decoding rule

$$\hat{A}_n = \sum_i \phi_{ni}(a_i(I(\mathbf{r}, \mathbf{A})) + \eta_i), \quad (3)$$

where the ϕ_{ni} are decoding weights and the η_i represent additive noise. Defining both the encoding (2) and decoding (3) rules completes the definition of how the population of simple cells represent the amplitudes A_n .

Since we are interested in the spatio-temporal properties of the simple cells, we also need a model of how images change with time. Olshausen has been exploring this issue by extending his sparseness model into the time domain (Olshausen 2002). We utilize a simpler alternative that assumes for a short period of time, and within local spatial patches, the spatio-temporal properties can be modeled as static images rigidly translating at a constant velocity, $I(\mathbf{r}, t, \mathbf{A}) = I(\mathbf{r} - \mathbf{v}t, \mathbf{A})$, an assumption used in most video compression algorithms. Thus, we simply add the velocity \mathbf{v} as a new stochastic variable to the wavelet amplitudes \mathbf{A} . The consequences and validity of this assumption, when applied to the data, are discussed below.

Using these thoughts as background, we examined receptive fields of 292 cat simple cells that were recorded during previous studies (DeAngelis et al. 1993, 1999). Our analyses were conducted using the parameters from fits to a model described by DeAngelis et al. (1999), rather than the raw data itself.

Spatial Properties of Simple Cell Responses

The spatial properties of simple cells eq. (2) that we are interested in are the preferred spatial frequency k_c , the spatial frequency standard deviation (SD) $\sigma_{\mathbf{k}} = 1/\sigma_{\mathbf{r}}$ orthogonal to the preferred orientation, and the spatial phase of the sinewave relative to the center of the Gaussian envelope ϕ_r . The spatial phase was shown by DeAngelis et al. (1993, 1999) to be uniformly distributed within the population, which our analysis confirms. A new observation we made was that the spatial frequency SD of the cells $\sigma_{\mathbf{k}}$ is linearly correlated to the spatial frequency of the cells

$$\sigma_{\mathbf{k}}(k_c) = a(k_c + k_0), \quad (4)$$

where for the cat $a = 0.39 \pm 0.03$ and $k_0 = 0.20 \pm 0.03$ cycles/deg. Previous studies in the macaque have shown that the number of side lobes in simple cells decreases with decreasing spatial frequency (De Valois and De Valois 1988, and Ringach 2000), as is also implied by this result. However, they did not analyze their data in this manner and missed this simple, but important result. In particular, replotting the DeValois graph that shows

the spatial frequency bandwidth, as measured in octaves, as a function of spatial frequency, clearly displays a linear relationship where $a = 0.39$ and $k_0 = 0.6$ cycles/deg.

The interpretation of this result begins with noting that in a pure scale invariant representation $\sigma_{\mathbf{k}}$ is strictly proportional to the magnitude of \mathbf{k}_c , which holds in this case when $|\mathbf{k}_c| > k_0$. The fact that the scale factor a is the same for both species provides support for the hypothesis that the shape of the spatial filters is determined by the statistics of natural images. However, all implementations of scale invariant representations must deal with the problem of what to do as the spatial frequency goes to zero, because the corresponding spatial extent of the filter goes to infinity. Multiscale pyramids and most orthonormal wavelets handle this by stopping at a particular scale and representing the remaining data as a lowpass sampled image. The result given by equation (4) shows that neurobiology found it beneficial to make this transition in a more smooth fashion. The different values of k_0 for the cat and monkey reflects a slightly different tradeoff that evolution settled as to where these two systems make the transition to a scale invariant representation.

Temporal Properties of Simple Cells:

The temporal properties of simple cells were also modeled by DeAngelis et al (1999) as Gaussian modulated sinewaves. So again, the primary parameters of interest are the temporal carrier frequency ω_c , the temporal frequency SD σ_ω , and the temporal phase ϕ_t . In contrast to the spatial filters, the temporal phase is highly peaked around $\pi/4$, and there is no significant correlation between σ_ω and ω_c .

Combined Spatio-Temporal Properties:

Combining the spatial and temporal filters introduces a new parameter that determines the degree to which cells preferentially respond to coherent motion. The original analysis (DeAngelis et al. 1999) introduced the parameter α to combine two quadrature pair filters

$$f(\mathbf{r}, t) = G_r(\mathbf{r})G_t(t) \cdot (\cos(\mathbf{k}_c \cdot \mathbf{r} + \phi_r) \cos(\omega_c t + \phi_t) + \alpha \sin(\mathbf{k} \cdot \mathbf{r} + \phi_r) \sin(\omega_c t + \phi_t)), \quad (5)$$

where $f(\mathbf{r}, t)$ is the spatio-temporal filter. The Gaussian envelopes $G_r(\mathbf{r})$ and $G_t(t)$ are not directly relevant to this discussion so are ignored. The modulation terms of the filter, i.e. the $\sin()$ and $\cos()$ parts, can be rewritten as a weighted average of two counter propagating traveling waves,

$$\cos(\epsilon) \cos(\mathbf{k}_c \cdot \mathbf{r} + \omega_c t + \phi^+) + \sin(\epsilon) \cos(\mathbf{k}_c \cdot \mathbf{r} - \omega_c t + \phi^-), \quad (6)$$

where $\phi^+ = \phi_r + \phi_t$, $\phi^- = \phi_r - \phi_t$ and

$$\cos(\epsilon) = \frac{1 + \alpha}{\sqrt{2(1 + \alpha^2)}}; \quad \sin(\epsilon) = \frac{1 - \alpha}{\sqrt{2(1 + \alpha^2)}}. \quad (7)$$

The angle ϵ is limited to the range $0 \leq \epsilon \leq \pi/4$, corresponding to $1 \geq \alpha \geq 0$, where the data is reordered such that the preferred direction is always in the direction of decreasing r , i.e.

$\cos(\epsilon) \geq \sin(\epsilon)$. One can then define a traveling wave index (TWI) as the relative power of the response to motion in the preferred direction minus the power in the reverse direction.

$$\text{TWI} = \cos(\epsilon)^2 - \sin(\epsilon)^2 = \frac{2\alpha}{1 + \alpha^2} \quad (8)$$

Our stochastic model of the spatio-temporal properties of images as static patches moving at a constant velocity suggests the cells would be tuned to a particular velocity, $v_c = \omega_c/k_c$, and a particular direction, i.e. the TWI should be peaked near 1. Instead the histogram of this index is found to be broadly distributed over the entire range. There is a small peak at $\text{TWI}=0$, $\epsilon = \pi/4$, corresponding to a pure counter flicker response and no directional selectivity; and a second, broader peak, centered at $\text{TWI}=0.5$, $\epsilon = \pi/8$, which lies midway between counter flicker $\epsilon = \pi/4$ and pure traveling wave $\epsilon = 0$. Our present conjecture is that our model leaves out saccades and other processes that produce sudden transients, which can be modeled as counter propagating traveling waves of equal intensity. While our prediction did not hold, asking it led to new insights into the spatio-temporal properties of simple cells.

However, our model led to another prediction about the spatio-temporal properties that was found to be true. The temporal bandwidth of the output of the wavelet spatial filters to an image moving at a constant velocity v is equal to $\sigma_\omega = \sigma_{\mathbf{k}} \cdot v = \sigma_{\mathbf{k}} \omega_c/k_c$. If we assume the temporal filters are tuned to pass all the information at this velocity, then we would predict the measured temporal bandwidth should obey the constraint

$$\sigma_\omega \geq \sigma_{\mathbf{k}} \omega_c/k_c. \quad (9)$$

This relationship was found to hold for 248 cells out of the 292 cells in the data set. While this constraint was derived for purely traveling waves, it can be generalized to the case of transients for the reason mentioned in the last paragraph.

We also found a weak, but significant correlation between the temporal phase ϕ_t and the TWI

$$\phi_t = b \cdot \text{TWI} + \phi_0, \quad (10)$$

where $b = 0.04 \pm 0.02$ and $\phi_0 = 33.0 \pm 3.0$ degrees. For the pure counter flicker case, $\text{TWI}=0$, the temporal phase $\phi_t = 33$ degrees, which produces a nearly monophasic temporal filter. At the other extreme of a pure traveling wave, $\text{TWI}=1$, the temporal phase $\phi_t = 70$ degrees, making the filter more biphasic. This makes sense because a monophasic temporal response matches transient changes while a biphasic response is essential for matching coherent motion.

Population Codes and the Issue of Signal to Noise:

Finally, we address an issue that is central to our framework (Eliasmith and Anderson 2003) and generally ignored by the visual neuroscience community. Namely, the SNR of

the wavelet coefficients A_n is larger at low spatial frequencies than it is at the highest spatial frequencies, and this excess is supported by increasing the population of simple cells at the lower spatial frequencies. It is not widely appreciated that most visual tasks utilize the high SNR information available in the low spatial frequency channels, and relatively few tasks utilize the low signal to noise high spatial frequency channels.

The increase in the SNR at low spatial frequencies arises from the pooling of large numbers of retinal ganglion cell inputs. However, the SNR of individual simple cells are presumably the same, thus the number of simple cells must be increased to handle this increase in the dynamic range. Quantitatively, the number of degrees of freedom, or the density of the coefficients covering a fixed spatial extent of an image, in a multiscale representation is proportional to the square of the spatial frequency k^2 . Thus, as k drops by a factor of 2, the density of the coefficients drops by a factor of 4. However, the convergence of the retinal inputs goes up by a corresponding factor of 4, which means, under the assumption of additive noise, that the SNR of the coefficients goes up by a factor of $\sqrt{4} = 2$. To encode this increase of 2 in the SNR with noisy neurons (modeled as additive noise of zero mean and constant variance eq. (3)) requires an increase in their number by a factor of 4. Thus, our first order estimate is that the density of simple cells at a particular location in primary visual cortex should be independent of spatial frequency. Furthermore, the SNR of the wavelet coefficients encoded by the population of simple cells scales as $1/k$.

The first modification to this simple model is that there is a point where the SNR reaches a sufficiently high level that the system does not allocate more resources as the spatial frequency decreases below a critical value k_{cutoff} , leading to a constant SNR for $k \leq k_{cutoff}$. This makes the density of neurons follow the expected k^2 behavior of the density of the wavelet coefficients for small k . Indeed, the histograms of the preferred spatial frequency of both the cat (DeAngelis et al 1999) and monkey simple cells (De Valois' 1988) show this behavior for spatial frequencies in the neighborhood of $k = 0$. The second modification is that the SNR of the wavelet coefficients has to go to zero as the spatial frequency approaches the Nyquist limit of the system, as set for example at the center of the fovea in primates by the optics of the eye. Thus, the density of simple cells will fall to zero as $k \rightarrow k_{max}$, where k_{max} is the maximum spatial frequency conveyed by the high resolution X cells in cat and P cells in monkey at a given angle of eccentricity. This roll off at the high spatial frequencies is observed in both the cat and monkey data.

It is tempting to connect these observations to the measured psychophysical modulation transfer function (MTF) of both the cat and monkey. However, there are many complicating factors that mitigate against doing this. For example, the MTF measurements depend on how the stimuli are presented, there is a lack of information about the shape of the spatial frequency prefilters that are applied before the signals are encoded into the retinal ganglion cells, and the cortical magnification factor has to be taken into account. In spite of these uncertainties, it seems clear that the SNR in primary visual cortex will rapidly decrease, i.e. at least as fast as $1/k$, as the spatial frequency increases above the point where the quadratic increase in the density of cells stops, k_{cutoff} . This occurs in the

cat data at around 0.2 cycles/deg, and a plot in the De Valois book shows the cat MTF peaking at around 0.4-0.5 cycles/deg. The De Valois's also show a histogram of the monkey simple cells spatial frequency responses where the quadratic behavior stops around 2-4 cycles/deg. in the fovea and 1-2 cycles/deg in the parafovea, while the MTF peaks around 4-5 cycles/deg. Given the uncertainties we feel there is reasonable agreement here.

Finally, we note that in CNS02 we provided an analysis that shows the low spatial frequency information as encoded by the ganglion cells in monkeys is highly distorted and must be modified before it arrives at the level of simple cells (Westover 2002). We suggested this could be accomplished by the introduction of a high degree of diversity in the nonlinear response properties of the large number of cells in layer 4 of V1.

In summary, our effort to connect the properties of individual simple cells to system level performance has revealed a number of new and important properties: (1) a linear relationship between $\sigma_{\mathbf{k}}$ and $|\mathbf{k}_c|$ in both cat and monkey, (2) a new TWI for simple cells and its highly informative histogram, (3) a correlation between this TWI and the temporal phase ϕ_t that makes signal processing sense, (4) a previously unnoticed and important relationship between σ_{ω} and $(\sigma_{\mathbf{k}}, \omega_c, \mathbf{k}_c)$, and (5) a quantitative understanding of why there is an excess of cells encoding the low spatial frequencies. We feel this is a remarkable number given that this is one of the oldest and most studied topics in visual neuroscience.

We thank Jonathan Victor for his insightful suggestions for this abstract.

References:

- DeAngelis, G.C., Ghose, G.M., Ohzawa, I., and Freeman, R.D. Functional micro-organization of primary visual cortex: Receptive-field analysis of nearby neurons. *Journal of Neuroscience*, 19: 4046-4064, 1999.
- DeAngelis, G.C., Ohzawa, I., and Freeman, R.D. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69: 1091-1117, 1993a.
- De Valois, R. L., and De Valois, K. K., *Spatial Vision*, Oxford Science Publications, 1998.
- Eliasmith, C., and Anderson, C. H., *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*, MIT Press, 2003.
- Olshausen B.A., Field, D.J., "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images" *Nature*, 381: 607-609, 1996.
- Olshausen, B. A., "Sparse Codes and Spikes", In: *Probabilistic Models of the Brain: Perception and Neural Function*. R.P.N. Rao, B. A. Olshausen, and M.S. Lewicki, Eds. MIT Press, pp. 257-272, 2002.
- Westover, M. B., and Anderson, C. H., "Layer 4C in Monkey V1 May Improve the Information Content of Simple Cell Responses", *CNS* 2002.