

# Visual neuroscience: Resonating to natural images

Allison B. Sekuler and Patrick J. Bennett

**Visual neurons may be optimized to produce sparse, distributed responses to natural scenes. This proposal, along with recent results from monkey fMRI and electrophysiology, may force us to re-interpret many neuroimaging results.**

Address: Department of Psychology, McMaster University, 1280 Main Street W., Hamilton, Ontario, L8S4K1, Canada.  
E-mail: sekuler@mcmaster.ca

*Current Biology* 2001, 11:R733–R736

0960-9822/01/\$ – see front matter  
© 2001 Elsevier Science Ltd. All rights reserved.

*A wing would be a most mysterious structure if one did not know that birds flew.* Barlow (1961)

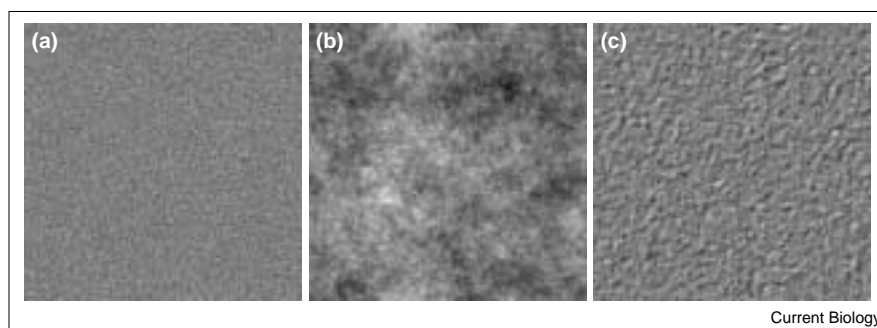
Neuroscientists have amassed an impressive array of data describing the anatomical and physiological characteristics of neurons at every level in the visual pathway. But what do visual neurons do? Why, for example, do the receptive fields of retinal ganglion cells have a center-surround structure? Why do simple cells in visual cortex exhibit spatial frequency and orientation tuning? Answers to these types of question cannot be answered by examining the response properties of visual neurons in ever greater detail. Instead, Barlow [1] suggested that it would be useful to consider the kinds of computations performed by visual neurons — to interpret the physiology within the context of a computational theory of sensory processing [2]. Barlow [1] and others (for example [3]) argued that a primary function of visual neurons is to re-code the input in a way that reduces redundancy and maximizes the information transmitted in the output.

At the time he made his suggestion, Barlow pointed out that there was little evidence to support or refute it: the response properties of visual neurons were only starting to be discovered, and even less was known about the statistical structure of the visual environment. All that has changed, of course, and in recent years tremendous progress has been made in characterizing the regularities in natural scenes, developing optimal coding strategies that take advantage of this structure, and relating these coding strategies to the physiological properties of visual neurons in a variety of species [4–7]. Even if Barlow's original redundancy-reduction hypothesis needs to be revised [8], there is little doubt that understanding the spatial and temporal structure in natural stimuli has led to a deeper understanding of the response properties of sensory neurons. A new study by Rainer *et al.* [9], published recently in *Current Biology*, suggests that understanding the structure of natural scenes, and the cortical mechanisms that represents that structure, also may be important for interpreting more global measures of neural activity, such as those obtained by functional magnetic resonance imaging (fMRI).

Perhaps the most obvious spatial structure in natural images is the similarity in intensities and wavelengths of light reflected from adjacent regions, and the decline with distance of the similarity or correlation between regions. This pattern of pairwise correlations is the reason why the amplitudes of Fourier components of natural scenes are roughly inversely proportional to frequency (see Figure 1). Returning to our initial question of why retinal ganglion cells have center-surround receptive fields, Attick [10] showed that such organization is nearly ideal for removing

## Figure 1

Three different kinds of noise. (a) White noise, in which each pixel's contrast was selected randomly from a zero-mean Gaussian distribution. On average, the amplitude spectrum of white noise is flat, yet the perceived structure is dominated by high spatial frequencies. (b) What the noise looks like after the amplitude spectrum is filtered so that it exhibits the  $1/f$  characteristic of natural images. Like natural images, the  $1/f$  noise has structure at multiple scales, but it lacks higher-order statistical structure that corresponds to sparse distributions of local features. (c) A wavelet texture constructed using the algorithm described by Field [13], which has higher-order structure that more closely resembles the kind found in natural scenes.



Current Biology

the spatial correlations, or redundancy, corresponding to  $1/f$  amplitude spectra.

But the similarity among nearby regions represents only a part of the statistical structure in natural images. After all, random textures with  $1/f$  amplitude spectra are never confused with natural images [11]. Unlike random  $1/f$  textures, natural images contain conspicuous local features such as lines and edges (see Figure 1). This structure is not captured by the pattern of pairwise correlations that produces  $1/f$  amplitude spectra, and instead corresponds to higher-order correlations that introduce structure in the phase spectra of natural images [12–16].

Field [17] and Daugman [18] have suggested that the spatial frequency and orientation tuning of simple cells in primary visual cortex are, in some sense, matched to this higher-order structure. They noted that the distribution of responses of bandpass linear filters to natural images was not Gaussian: most of the filters had a zero or very small response, and a small subset had very robust responses. This high kurtosis in the response distribution — which Field referred to as sparseness — occurs for natural images but not for random  $1/f$  textures, and is greatest when the spatial frequency and orientation bandwidths of the filters correspond to the average bandwidths found in simple cells. These findings led to the proposal that the tuning properties of simple cells have been optimized to produce sparse distributed responses, with most of the information about a natural scene being carried by a relatively small set of cells, and with different scenes being represented by different sets of cells [12,13,19,20]. This proposal has received indirect support from recent single unit recordings showing that simple cells in fact are rarely active when stimulated with natural scenes [21], and that a great deal of stimulus information is carried by the few cells that do respond vigorously [22].

To what extent are measures of global neural activation, such as fMRI, affected by the way in which the visual system seems to have evolved to encode information about natural scenes? fMRI provides a measure of the blood-oxygenation-level-dependent (BOLD) signal, which depends on activity within a given region of the brain. A number of recent articles have noted a link between the BOLD signal and action potentials — in some cases, higher BOLD signals in humans occur under conditions that lead to higher levels of spiking in neurons in non-human primates [23,24]. Thus, it is often implicitly assumed that a greater BOLD signal in a particular area is a sign that neurons in that area respond preferentially to the specific stimulus that led to the increased BOLD signal. If Field and Daugman are correct, however, this might be a false assumption — a proper interpretation of the BOLD signal would depend not only on neuronal

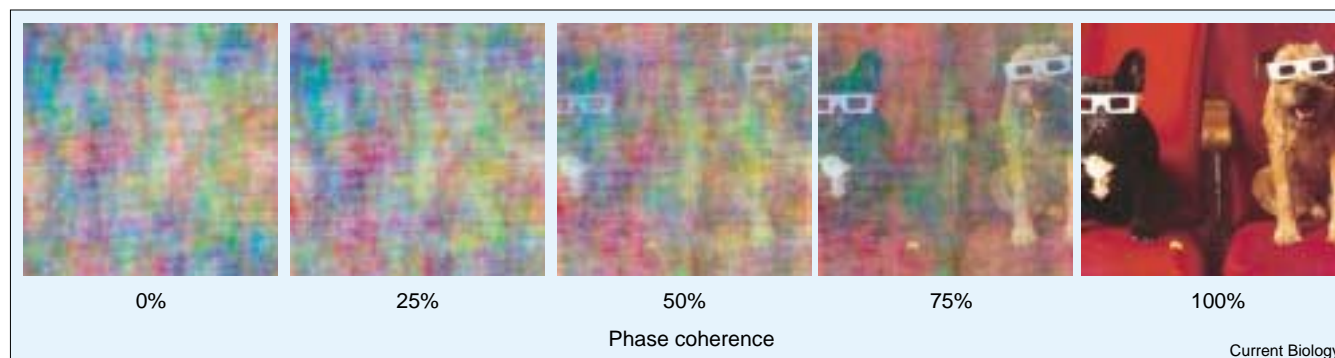
responses, but also on the sparseness of those responses. For example,  $1/f$  random textures and natural images could evoke similar average responses but differ substantially in sparseness, and fMRI might be blind to this difference.

Rainer *et al.*'s [9] recent study addresses precisely this issue. They used fMRI to assess the neural activity in anesthetized monkeys when viewing natural images and degraded versions of those images. The question was: how does neural activity vary as a function of the coherent structure of an image? By interpolating between a natural image and  $1/f$  noise, Rainer *et al.* [9] constructed a series of images with identical amplitude spectra but different amounts of phase coherence (Figure 2). In previous behavioural and physiological work with similar images, Rainer and Miller [25] found that monkeys' discrimination thresholds decreased and the response of prefrontal neurons increased with increasing phase coherence. Consequently, one might expect that the BOLD signal should increase with increasing phase coherence (that is, as the image assumes the appearance of a natural scene).

Rainer *et al.* [9] did find consistent BOLD signals in several parts of the visual system: the primary visual cortex (V1), extrastriate cortex (V2, V3), and the superior temporal sulcus (STS). Furthermore, as one might expect based on the behavioral results for these sorts of stimuli, in each brain region the BOLD signals were greater with 100% phase coherence than with 0% phase coherence. When the complete parametric design was included, however, it became clear that this relationship was neither linear nor monotonic in primary visual cortex or extrastriate cortex. In both cases, the BOLD signals initially *declined* with increasing coherence — despite the fact that stimuli are perceptually more similar to natural scenes as coherence increases (see Figure 2), and despite the fact that discrimination improves as coherence increases [25].

Although this non-monotonic result is surprising at first glance, it may be related to the ideas of sparse coding in the visual cortex. Recall that the distribution of responses of bandpass filters to  $1/f$  noise is approximately Gaussian, but that the response distribution for natural scenes is more sparse (has higher kurtosis). Rainer *et al.* [9] suggest that their results are directly related to this response distribution shift. They assume that relatively large BOLD signals are found for 0% and 100% coherence stimuli for different reasons. In the case of 0% phase coherence, Rainer *et al.* [9] suggest that most neurons fire at relatively low levels, but that there are enough of these neurons doing something that the total response of all neurons included in the region of interest is relatively high. In the case of 100% phase coherence, there are conspicuous, locally oriented, high-contrast regions, which produce large responses only in neurons with appropriate receptive

Figure 2



An example of varying phase coherence, as used in the work of Rainer *et al.* [9]. Here, the amplitude spectra are identical in each panel, but only the image on the right (100% phase coherence) contains the original phase spectrum, so that it best represents this 'natural' scene.

The phase information is completely randomized in the image on the left (0% phase coherence), and intermediate levels of phase coherence are formed by weighted linear combination of the pixels in the 0 and 100% stimuli.

fields. Thus, although most neurons will produce only very small responses, some will fire quite vigorously — and the total neuronal response in the region may be higher than the total response for the 0% coherence stimulus.

For intermediate levels of phase coherence, Rainer *et al.* [9] suggest that there is a sort of trade-off, such that fewer neurons will respond than for 0% coherence, but they do not respond at high enough levels to overcome the reduction in the number of neurons responding. Consequently, the BOLD signal initially decreases with increasing structure, before rising again as structure increases further. Only at the level of the STS, where the response properties of neurons are considerably more complicated, do we see anything resembling a monotonic, linear relationship between BOLD signals and phase coherence.

Another recent paper from the Logothetis laboratory [26] complicates the interpretation of BOLD signals further. This paper reports the first study to directly compare BOLD signals with a variety of simultaneously obtained electrophysiological measures. They found that BOLD signals were most related to local field potentials, rather than to action potentials (even when action potentials were combined globally within a region). Because local field potentials are related to membrane changes — which are dominated by dendrites — Logothetis *et al.* [26] suggest that BOLD signals are more likely related to the input to, and intracortical connections within, a region than to the output (action potentials) of that region. This conclusion differs substantially from the popular view that the BOLD signal reflects action potentials [23,24], but it is consistent with conclusions from optical imaging (a technique that also depends on blood oxygenation levels for the inferences it draws [27]).

If Logothetis *et al.* [26] are correct, then the V-shaped function Rainer *et al.* [9] observed in V1 may be more related to the inputs to V1 — both feedforward and feedback — than to the outputs from that area. Similarly, the BOLD signal in extrastriate cortex may in fact be dominated by the output from V1. In fact, V-shaped functions are observed both at the level of V1 and extrastriate cortex, so the sparseness-related explanation put forward by Rainer *et al.* [9] would still apply. But the simple neural model Rainer *et al.* [9] describe would fit the data less well, as it would inaccurately predict the location of the V's minimum. It remains to be seen whether a more realistic model would provide an accurate quantitative fit — and whether that fit would be more consistent with the BOLD signal in extrastriate cortex or V1 (perhaps providing support or lack of support, respectively, for the Logothetis hypothesis).

Regardless of whether the BOLD response reflects action potentials or local field potentials, the results of Rainer *et al.*'s [9] study have dramatic implications for the study of the neural processing of natural images, and for the interpretation of fMRI results more generally. First, the results imply that different underlying patterns of local neuronal activity can lead to similar global activation patterns. In other words, an X% increase in BOLD signal may be signaling very different things under different circumstances. Here, if one were to consider only the results from the 0 and 100% conditions, one might erroneously conclude that individual neurons respond similarly to complete phase coherence and complete phase incoherence, with only a slight increase in firing for coherent stimuli. Only with some understanding of what is happening at the level of individual neurons does one come to the realization that the result 'means' different things in the two cases. Obviously, this sort of analysis can only be done when one has a

relatively good understanding of the response properties of individual cells in a given brain region — an understanding that is currently more complete in lower levels of visual processing than higher levels, such as the STS.

The results are also important because they clearly highlight the necessity of running parametric studies, rather than simply comparing two conditions as is often done. If Rainer *et al.* [9] were to have included only the 0 and 100% conditions, they would have concluded that there is a weak linear relationship between BOLD signal and phase coherence. If only 25 and 100% had been included, the linear relationship would have seemed even stronger. In either case, the conclusions would have led to inaccurate interpretations of the results. Only by including a full range of intermediate values can one reveal the non-monotonicity of the results — leading to a very different conclusion about the brain and the visual system than one might have otherwise made.

Finally, the results highlight the fact that different parts of the brain encode information about the same stimuli in different ways. For example, the pattern of BOLD signals changes systematically as one moves from V1, through extra-striate visual cortex, to higher level visual processing (area STS). Presumably, the increased linearity reflects changes in the response properties of individual neurons at each level, and the varied inputs to each of those levels. A complete understanding of the neural system underlying behavior on a particular task requires an understanding of how information is encoded in the full range of contributing regions of the brain. This systems-approach is particularly relevant in the context of Logothetis *et al.*'s [26] results. For example, if BOLD signals are telling us primarily about the inputs to a particular brain region (both feedforward and feedback), then it is critical to understand the functional connections across the brain, and how those connections vary across observers and conditions [28–30].

## References

- Barlow HB: Possible principles underlying the transformation of sensory messages. In *Sensory Communication*. Edited by Rosenblith W. Cambridge, MA: MIT Press; 1961:217-234.
- Marr D: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman. 1982:xvii, 397.
- Attneave F: Informational aspects of visual perception. *Psychol Rev* 1954, **61**:183-193.
- Pelach A: Special issue: the vision of natural and complex images (preface). *Vis Res* 1997, **37**:3201-3202.
- Moorhead I, Troscianko T: Special Issue: natural images (Guest Editorial). *Perception* 2000, **29**:1013-1015.
- Reinagel P, Laughlin L: Special issue: natural stimulus statistics (editorial). *Network: Computation in Neural Systems* 2001, **12**:237-240.
- Simoncelli EP, Olshausen BA: Natural image statistics and neural representation. *Annu Rev Neurosci* 2001, **24**:1193-1216.
- Barlow HB: Redundancy reduction revisited. *Network: Computation in Neural Systems* 2001, **12**:241-253.
- Rainer G, Augath M, Trinath T, Logothetis NK: Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr Biol* 2001, **11**:846-854.
- Attick JJ: Could information theory provide an ecological theory of sensory processing? *Network* 1992, **3**:213-251.
- Thomson MG, Foster DH, Summers RJ: Human sensitivity to phase perturbations in natural images: a statistical framework. *Perception* 2000, **29**:1057-1069.
- Field DJ: Scale invariance and self-similar 'Wavelet' transforms: an analysis of natural scenes and mammalian visual systems. In *Wavelets, Fractals and Fourier Transforms*. Edited by Farge M, Hunt JCR, Vassilicos JC. Oxford: Clarendon Press, 1993:151-193.
- Field DJ: What is the goal of sensory coding? *Neural Comput* 1994, **6**:559-601.
- Thomson MG: Visual coding and the phase structure of natural scenes. *Network* 1999, **10**:123-132.
- Thomson MGA: Higher-order structure in natural scenes. *J Opt Soc Am* 1999, **16**:1549-1553.
- Thomson MGA: Beats, kurtosis and visual coding. *Network: Computation in Neural Systems* 2001, **12**:271-287.
- Field DJ: Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am* 1987, **4**:2379-2394.
- Daugman JG: Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Trans Biomed Eng* 1989, **36**:107-114.
- Olshausen BA, Field DJ: Natural image statistics and efficient coding. *Network: Computation in Neural Systems* 1996, **7**:333-339.
- Olshausen BA, Field DJ: Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res* 1997, **37**:3311-3325.
- Vinje WE, Gallant JL: Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 2000, **287**:1273-1276.
- Geisler WS, Albrecht DG: Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Vis Neurosci* 1997, **14**:897-919.
- Heeger DJ, Huk AC, Geisler WS, Albrecht DG: Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nat Neurosci* 2000, **3**:631-633.
- Rees G, Friston K, Koch C: A direct quantitative relationship between the functional properties of human and macaque V5. *Nat Neurosci* 2000, **3**:716-723.
- Rainer G, Miller EK: Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron* 2000, **27**:179-189.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A: Neurophysiological investigation of the basis of the fMRI signal. *Nature* 2001, **412**:150-157.
- Shoham D, Glaser DE, Arieli A, Kenet T, Wijnbergen C, Toledo Y, Hildesheim R, Grinvald A: Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes. *Neuron* 1999, **24**:791-802.
- Bennett PJ, Sekuler AB, McIntosh AR, Della-Maggiore V: The effects of aging on visual memory: evidence for functional reorganization of cortical networks. *Acta Psychol* 2001, **107**:249-273.
- Della-Maggiore V, Sekuler AB, Grady CL, Bennett PJ, Sekuler R, McIntosh AR: Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *J Neurosci* 2000, **20**:8410-8416.
- McIntosh AR, Sekuler AB, Penpeci C, Rajah MN, Grady CL, Sekuler R, Bennett PJ: Recruitment of unique neural systems to support visual memory in normal aging. *Curr Biol* 1999, **9**:1275-1278.