

Local Visual Energy Mechanisms Revealed by Detection of Global Patterns

Yaniv Morgenstern and James H. Elder

Centre for Vision Research, York University, Toronto, Ontario, Canada M3J 1P3

A central goal of visual neuroscience is to relate the selectivity of individual neurons to perceptual judgments, such as detection of a visual pattern at low contrast or in noise. Since neurons in early areas of visual cortex carry information only about a local patch of the image, detection of global patterns must entail spatial pooling over many such neurons. Physiological methods provide access to local detection mechanisms at the single-neuron level but do not reveal how neural responses are combined to determine the perceptual decision. Behavioral methods provide access to perceptual judgments of a global stimulus but typically do not reveal the selectivity of the individual neurons underlying detection. Here we show how the existence of a nonlinearity in spatial pooling does allow properties of these early mechanisms to be estimated from behavioral responses to global stimuli. As an example, we consider detection of large-field sinusoidal gratings in noise. Based on human behavioral data, we estimate the length and width tuning of the local detection mechanisms and show that it is roughly consistent with the tuning of individual neurons in primary visual cortex of primate. We also show that a local energy model of pooling based on these estimated receptive fields is much more predictive of human judgments than competing models, such as probability summation. In addition to revealing underlying properties of early detection and spatial integration mechanisms in human cortex, our findings open a window on new methods for relating system-level perceptual judgments to neuron-level processing.

Introduction

One of the fundamental jobs of the primate vision system is to detect objects of interest in the visible environment. The system must be able to work under a diverse range of scene conditions, with varying levels of illumination, shadows, clutter, and camouflage. These conditions can, at times, conspire to reduce the effective figure/ground contrast of the object to challenging levels. Thus, it is important for the visual system to efficiently exploit whatever information is available in the image.

Neurons in early visual cortex have highly localized receptive fields that will, in general, subtend only part of an object. Thus, efficient detection depends crucially on both the nature of these local receptive fields and on how the outputs of these neurons are combined over retinal space to yield a perceptual judgment. This process of spatial pooling is challenging to study, since it lies at the interface between local processing, involving computations at the single-neuron level, and system-level judgments, which must be studied using behavioral methods.

Our goal in this work is to develop and apply novel behavioral system identification techniques that will allow us to bridge this

gap, gaining insights into neuron-level processing from behavioral data. Our methods and findings build on previous work that relies on the injection of noise into visual stimuli to aid linear system identification (Ahumada and Lovell, 1971; Ahumada and Beard, 1999; Beard and Ahumada, 1999; Pelli and Farell, 1999; Ahumada, 2002; Solomon, 2002). These methods assume that detection is based on a linear system, where the decision variable is given by the inner product of a spatial template (the receptive or summation field) with the visual input. If the internal template is matched to the stimulus to be detected and the added visual noise is Gaussian and white, this strategy yields optimal detection performance (Green and Swets, 1966).

This linear assumption is satisfied if the neurons coding local stimulus information are themselves linear, and if their outputs are pooled additively. Note that for such a system, it is, in principle, impossible to use behavioral techniques to identify the receptive fields of the individual neurons: since addition is associative, linear summation within each receptive field cannot be distinguished from linear summation over receptive fields. For example, given inputs $x_1 \dots x_4$ and output y , one cannot distinguish the system $y = (\alpha_1 x_1 + \alpha_2 x_2) + (\alpha_3 x_3 + \alpha_4 x_4)$ from the system $y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4$.

It has, however, been demonstrated that visual detection, even for relatively simple stimuli, can be highly nonlinear (Ahumada and Beard, 1999; Solomon, 2002; Goris et al., 2008). This important observation is both a blessing and a curse for neuroscientists. It is a blessing because the existence of a nonlinearity means that, in principle, it may be possible to recover local neural parameters using behavioral data based on global perceptual judgments. It is a curse because this nonlinearity prevents the application of powerful stochastic linear system identification methods. Thus, the

Received July 29, 2011; revised Dec. 12, 2011; accepted Jan. 14, 2012.

Author contributions: Y.M. and J.H.E. designed research; Y.M. and J.H.E. performed research; Y.M. and J.H.E. analyzed data; Y.M. and J.H.E. wrote the paper.

This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Geomatics for Informed Decisions Network of Centres of Excellence, the Premier's Research Excellence Award (Ontario), and the Ontario Centres of Excellence. We thank Bosco Tjan and two anonymous reviewers for invaluable comments on this manuscript.

Correspondence should be addressed to James H. Elder, Centre for Vision Research, York University, Room 0003G, Computer Science Building, 4700 Keele Street, North York, Ontario, Canada M3J 1P3. E-mail: jelder@yorku.ca.

DOI:10.1523/JNEUROSCI.3881-11.2012

Copyright © 2012 the authors 0270-6474/12/323679-18\$15.00/0

specific issue we address here is how to adapt these stochastic techniques to a system that is highly nonlinear. In doing this, we are able to take advantage of the system nonlinearity, using behavioral judgments of global stimuli to estimate local neuron-level receptive field parameters.

Materials and Methods

The problem of nonlinear spatial pooling can be studied using a simple stimulus such as a Gaussian-windowed sinusoidal grating (Fig. 1), which has been used to study spatial perception for many decades (Campbell and Robson, 1968). It is well established that contrast detection thresholds for noiseless grating stimuli decline progressively as the stimulus width is increased, reflecting long-range spatial pooling (Howell and Hess, 1978). Whether this is also the case for noisy stimuli has been a matter of some debate (Legge and Foley, 1980; Kersten, 1984). This question is of some importance, because system identification techniques using noisy stimuli are frequently used in both psychophysical (Ahumada and Lovell, 1971; Ahumada and Beard, 1999; Beard and Ahumada, 1999; Pelli and Farell, 1999; Ahumada, 2002; Solomon, 2002) and physiological (Ringach, 2004) experiments. Thus, our first two experiments will test both for the presence of pooling nonlinearities and for changes to these pooling mechanisms induced by injection of noise into the stimuli. Our third and final experiment will then explore the use of noisy stimuli to identify underlying detection mechanisms.

Apparatus. Stimuli were displayed on a calibrated 85 Hz, 21 inch CRT using standard psychophysical control software (Brainard, 1997). The viewing distance was 57 cm. At this distance, a single pixel subtended 6 arcmin. All experiments were conducted in a dimly lit room, and stimuli were presented on a midgray (mean luminance) background.

In Experiments 1 and 3, our stimuli were 256×256 pixels in extent, subtending $25.3 \times 25.3^\circ$. In Experiment 2, we maximized the stimulus extent to 400×300 pixels (width by height), so that the stimuli subtended $38.7 \times 29.5^\circ$.

In Experiment 1, we used a video attenuator (Pelli and Zhang, 1991) to measure contrast thresholds in the absence of stimulus noise, achieving a 12-bit gamut on the green channel of the monitor (mean luminance, 20 cd/m^2).

In Experiment 3, although noise was added to all stimuli, thresholds were still too low in our low-noise condition to estimate with a standard 8-bit gamut. We therefore used a 10-bit RADEON 7000 Mac Edition graphics card. Whereas this provides only an instantaneous 8-bit gamut, it allowed us to shift the gamut between low- and high-contrast modes. (In the low-contrast mode, only the central 256 gray levels of the 1024 gray-level table were used. In the high-contrast mode, on average, every fourth gray level was used.) We used the low-contrast mode for our low-noise condition and the high-contrast mode for our intermediate and high-noise conditions.

In Experiment 2, all stimuli contained an intermediate level of noise, allowing us to use a standard 8-bit gamut. The mean luminance for both Experiments 2 and 3 was 50 cd/m^2 . The display system was linearized for each experiment.

Stimuli. All three of our experiments involved the addition of white Gaussian noise. Three levels of noise (quantified by SD over mean luminance) were used: low noise (4.3%), intermediate noise (15%), and high noise (50%).

In Experiment 1, the signal $s(x, y)$ was a vertically oriented sinusoidal grating, windowed in the horizontal (x) direction with a Gaussian windowing function:

$$s(x, y) = c \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \cos(2\pi f_s x + \phi). \quad (1)$$

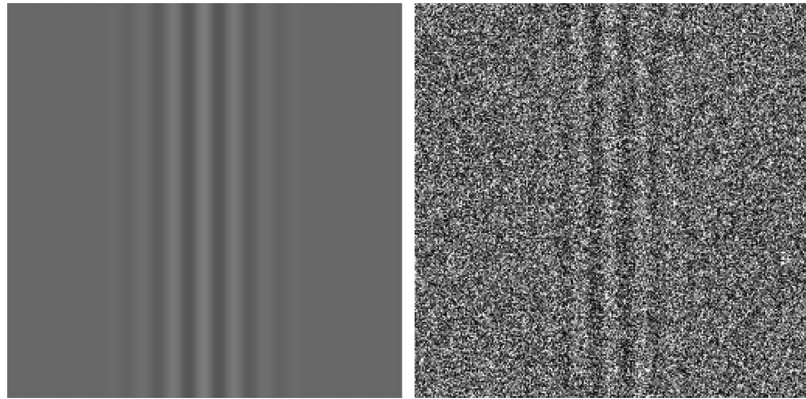


Figure 1. Experiment 1: detection of windowed grating patterns. An example stimulus is shown: $0.5 \text{ c}/^\circ$ Gaussian-windowed grating stimulus, four cycles in width, no-noise and high-noise conditions.

Two spatial frequencies were tested: $f_s = 0.5$ and $1.7 \text{ cycles } (\text{c})/^\circ$. The phase ϕ was 0 (at fixation) in the constant phase condition and uniformly distributed over $[-\pi, \pi]$ in the random condition. The Gaussian space constant σ_x was selected to subtend $n = 1, 2, 4,$ or 8 cycles of the grating between $1/e$ points: $\sigma_x = n/(2\sqrt{2}f_s)$. Contrast thresholds were measured both without noise and with a high level of noise (50%). Gray levels were clipped if they exceeded the gamut of the graphics system.

In Experiment 2, the grating signal was fixed at cosine phase and also windowed in the vertical direction:

$$s(x, y) = c \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \cos(2\pi f_s x). \quad (2)$$

The grating frequency was $f_s = 0.5 \text{ c}/^\circ$. As for Experiment 1, σ_x was selected to subtend $n = 1, 2, 4,$ or 8 cycles of the grating between $1/e$ points. The vertical space constant was fixed at $\sigma_y = 4\sqrt{2}^\circ$. An intermediate (15%) level of noise was added to the stimulus.

In Experiment 3, the stimulus was simply a noisy vertical grating of frequency $f_s = 0.5$ or $1.7 \text{ c}/^\circ$, in cosine phase:

$$s(x, y) = c \cos 2\pi f_s x. \quad (3)$$

All three levels of noise were tested.

Procedure. There were two subjects for Experiment 1 (two males), three subjects for Experiment 2 (one female and two males), and three subjects for Experiment 3 (two females and one male). One of the authors (Y.M.) was a subject in all three experiments; the others were naive to the purposes of the experiments. Other than Y.M., no subject did more than one experiment. All six subjects had normal or corrected-to-normal visual acuity.

Experiments 1 and 2 were two-interval forced-choice tasks: on each trial, a signal-absent stimulus and a signal-present stimulus were displayed in random order. Subjects were asked to press the left arrow key on a keyboard if they believed the signal to be in the first image and the right arrow key if they believed it to be in the second image. Each trial consisted of a 500 ms fixation interval, followed by a 500 ms blank screen, a 200 ms stimulus interval, a 600 ms blank screen, a second 200 ms stimulus interval, and then a blank screen again, displayed until the subject responded.

Experiment 3 was a detection (present/absent) task. On each trial, only one stimulus was displayed, randomly selected to be either signal plus noise or noise alone. Subjects were asked to press the left arrow key on a keyboard if they believed the signal to be present and the right arrow key otherwise. Each trial consisted of a 500 ms fixation interval, followed by a 500 ms blank screen, a 200 ms stimulus interval, and then a blank screen again, displayed until the subject responded.

For all experiments, an auditory tone was sounded if the response was incorrect. Subjects were asked to maintain fixation on a central cross during fixation and stimulus intervals.

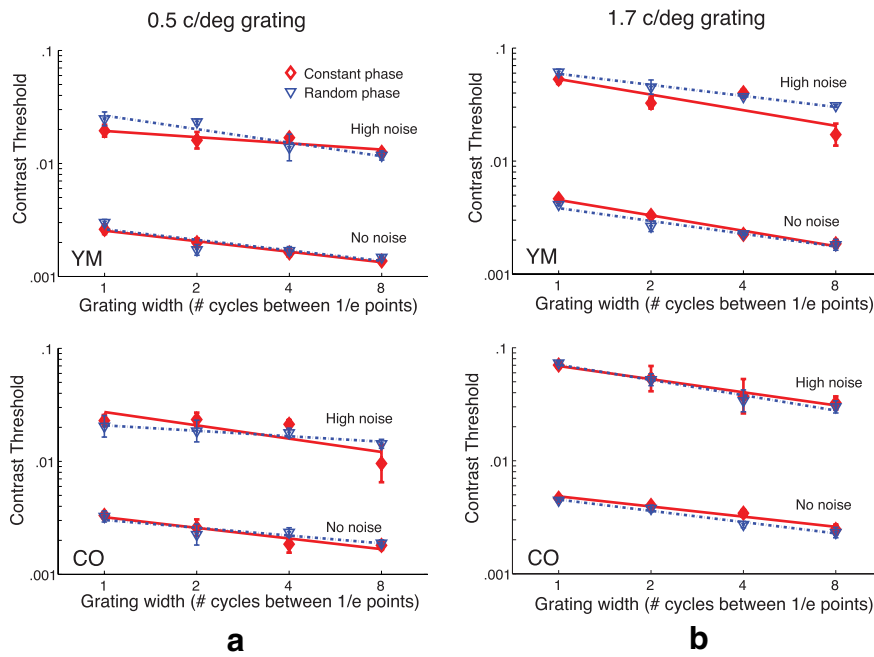


Figure 2. Experiment 1: detection of windowed grating patterns. Plots show threshold contrast required for reliable (75% correct) detection of the grating stimulus, for two subjects. Results for grating stimuli with a known, constant (cosine) phase are shown in red. Results for random phase gratings are shown in blue. Error bars indicate SEM. Lines show maximum likelihood linear fits. **a**, Results for 0.5 c/° grating stimulus. **b**, Results for 1.7 c/° grating stimulus. YM and CO are observers.

Table 1. Empirically estimated summation constants (slope of log contrast threshold vs log stimulus width)

	0.5 c/° Grating		1.7 c/° Grating	
	Without noise	With noise	Without noise	With noise
Observer C.O.				
Constant phase	-0.31	-0.39	-0.30	-0.39
Random phase	-0.23	-0.16	-0.33	-0.45
Observer Y.M.				
Constant phase	-0.31	-0.18	-0.45	-0.46
Random phase	-0.31	-0.40	-0.38	-0.32
Mean	-0.29	-0.28	-0.37	-0.41
SE	0.02	0.07	0.03	0.03

An adaptive psychometric procedure (QUEST; Watson and Pelli, 1983) was used for all experiments to estimate the threshold stimulus contrast c_t for 75% correct performance in Experiments 1 and 2 and to maintain stimulus contrast near threshold in Experiment 3.

Subjects performed 4 blocks of 40 trials for each stimulus condition in Experiment 1, 8 blocks of 40 trials for each stimulus condition in Experiment 2, and 10 blocks of 500 trials for each stimulus condition in Experiment 3. Blocks were randomly interleaved over conditions in all three experiments. In Experiment 2, at the beginning of each block of trials, an example suprathreshold stimulus was shown until subjects pressed a key to begin the block.

Results

Experiment 1: effects of noise and phase randomization on summation

In our first experiment, we measured contrast detection thresholds for Gaussian-windowed grating stimuli of various widths, both with noise (50% contrast) and without noise (Fig. 1). We tested two conditions: one in which the grating stimuli were of known and constant (cosine) phase and one in which the phase was randomized from trial to trial.

Figure 2 shows the results for two observers and two grating frequencies. Thresholds were slightly higher for the 1.7 c/° condition. This may seem surprising, since the peak of the contrast sensitivity function is generally thought to be in the 3–4 c/° range (Schade, 1956; Campbell and Robson, 1968). However, note that, consistent with Kersten (1984), our results are plotted as a function of the number of cycles visible. This means that for the plots in Figure 2, a fixed position on the abscissa represents a larger stimulus width (in degrees) for the 0.5 c/° condition than for the 1.7 c/° condition. Thus, thresholds for the 0.5 c/° condition are expected to be lower because of spatial summation. It is also the case that the exact shape of the CSF depends on many different factors (Graham, 2001).

Spatial summation

Most importantly for present purposes, thresholds were found to decline steadily as the grating window was increased in width, for all conditions. To quantify this decline, we fit both linear and quadratic models to the data. We found that the mean value of the second-order term of the quadratic term over all conditions was not significantly different from zero ($t_{(15)} = 0.05$; $p = 0.96$, two-tailed). Furthermore, a four-way ANOVA examining the effects of spatial frequency, observer, noise, and phase randomization on the quadratic term revealed no significant main effects or two-way interactions ($p > 0.1$). Thus, to a good approximation, this decline was linear in log–log space, indicating a power-law relationship.

Table 1 and Figure 3 summarize the empirical summation slopes derived from maximum likelihood linear fits to the data. A four-way ANOVA examining the effects of spatial frequency, observer, noise, and phase randomization on the summation slope revealed a significant effect of spatial frequency ($F_{(1,11)} = 5.6$; $p = 0.037$; slopes were steeper for the 1.7 c/° condition than for the 0.5 c/° condition. This difference may be attributable to uncertainty regarding the exact position and spatial extent of the windowed grating stimulus. We explore this possibility below (see The role of spatial uncertainty).

Otherwise, no significant main effects of observer ($F_{(1,11)} = 0.6$; $p = 0.5$), noise ($F_{(1,11)} = 0.2$; $p = 0.7$), or phase randomization ($F_{(1,11)} = 0.4$; $p = 0.5$) were observed. In particular, the rate of improvement with stimulus width was very similar for noiseless and noisy stimuli. We also note that previous work (Mayer and Tyler, 1986) suggests that summation slopes for grating detection are invariant to the specific threshold at which they are measured, since the slope of the psychometric function is roughly invariant to grating width. Thus, it appears that with some generality, the addition of noise to the stimulus does not change the fundamental nature of the detection and pooling mechanisms underlying the task and that system identification techniques using noisy stimuli can potentially be used to estimate these mechanisms.

Incoherent detection

For the constant-phase condition of our first experiment, the ideal linear template is a Gaussian-windowed grating function

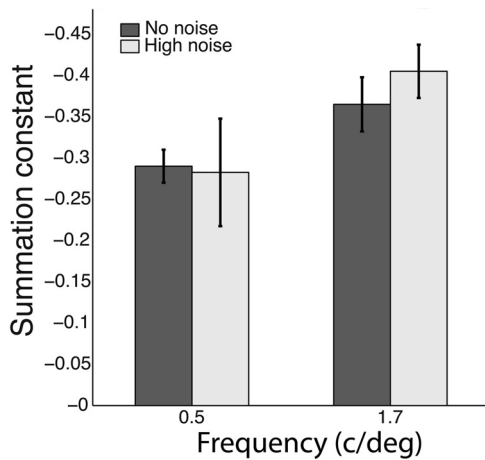


Figure 3. Experiment 1: empirical summation constants (slope of log contrast threshold vs log stimulus width), averaged over phase condition and observer. Error bars indicate SEM.

matching the frequency, orientation, position, and phase of the stimulus. Matching stimulus phase is critical here: if an incorrect phase is used, or if the system pools over multiple phases, performance declines (Green and Swets, 1966). Of course, in the random-phase condition, it is impossible for an observer to match the stimulus phase, and performance of a linear system should be lower than for the fixed phase. The fact that our human subjects perform similarly for the two conditions (Fig. 2) thus indicates that they are incoherent detectors, unable to exploit this phase information (Ahumada and Beard, 1999; Solomon, 2002), and must be using some form of nonlinear strategy.

Experiment 2: effect of noise on summation—control

The results of our first experiment suggest that the addition of stimulus noise does not qualitatively change the nature of spatial summation. However, these results are not consistent with previous work. In particular, Kersten (1984) found that adding noise to windowed grating stimuli seemed to cut off summation, leading to flat summation functions. There are specific methodological differences that we believe may account for this divergence in findings (see Discussion). However, before proceeding with classification image analysis, we want to verify that the results of our first experiment are not caused by some artifact in our stimuli.

One concern is that the sharp border between the noisy grating and the mean luminance background at the top and bottom edge of the stimulus introduces broadband components that, in principle, could be used to aid detection. Also, for the 0.5 $c/^\circ$ condition and the largest window size (eight cycles visible), the grating signal is truncated at left and right borders at 8% of peak amplitude. Thus, a small broadband artifact was introduced here as well.

There are at least two reasons why these components are unlikely to explain the divergence between our results and the results of Kersten (1984). First, stimulus truncation will only aid detection if it introduces a frequency component to which we are more sensitive. However, at the eccentricity of the stimulus boundary (12.7°), the peak of the CSF is $\sim 0.5 c/^\circ$ (Banks et al., 1991), and contrast sensitivity is generally lower than at the fovea. Second, we note that Kersten's stimuli had similar artifacts at the stimulus boundary, perhaps slightly worse since Kersten used a dark background, rather than the mean-luminance background used in our experiments, and the top and bottom boundaries were generally much closer to the fovea.

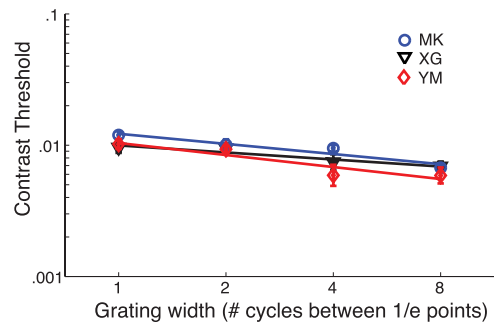


Figure 4. Experiment 2: detection of windowed 0.5 $c/^\circ$ cosine phase grating patterns. Stimuli were designed to greatly reduce stimulus artifacts introduced at the stimulus boundary and to reduce clipping of high and low intensities. Plots show threshold contrast required for reliable (75% correct) detection of grating stimulus, for three subjects. Error bars indicate SEM. Lines show maximum likelihood linear fits. MK, XG, and YM are observers.

Table 2. Experiment 2: empirically estimated summation constants (slope of log contrast threshold vs log stimulus width)

Observer	Summation constant
M.K.	-0.26
X.G.	-0.18
Y.M.	-0.30
Mean	-0.25
SE	0.04

Nevertheless, to be fully confident in our summation results, in our second experiment we again measured summation, focusing on the 0.5 $c/^\circ$, fixed-phase condition, but with a larger stimulus field ($38.7 \times 29.5^\circ$, width by height), and with a Gaussian window in both the horizontal and vertical dimensions. We fixed the vertical window space constant at $\sigma_y = 4\sqrt{2}$. These changes reduced the amplitude of the grating signal to 0.3% of peak or less at left and right boundaries and to 3% of peak at the top and bottom boundaries.

A second concern might be the clipping of the high-amplitude Gaussian noise that occurred when the intensity exceeded the gamut of the display system. For the high noise (50%) level used in Experiment 1, clipping occurred for 5% of pixels. To reduce this clipping, in Experiment 2 we use an intermediate (15%) noise level, for which clipping occurs for only 0.09% of pixels. This also allows us to assess how our results generalize over different levels of noise.

A third concern might be that the results somehow derive from subjects' uncertainty about the spatial extent of the stimuli. To address this, at the beginning of each block we displayed a suprathreshold example of the stimulus that remained on until the subject pressed a button to commence the block. By using an intermediate level of noise, we also are better able to assess whether the noise might affect the degree of spatial uncertainty.

Finally, the results of our first experiment were based on data from only two subjects. Although these two subjects were quite consistent with each other, it is possible that they are not representative of the larger population. In our second experiment, we test three subjects, including two new subjects, bringing the total number of subjects for our summation experiments to four.

The results of Experiment 2 are shown in Figure 4 and Table 2. We see that the results are very consistent with our first experiment: thresholds decline steadily with stimulus width, and the mean slope of this decline is -0.25 , not significantly different from the mean slope of -0.28 observed for the 0.5 $c/^\circ$ constant-phase, high-noise condition in our first experiment: $t_{(4)} = 0.4$;

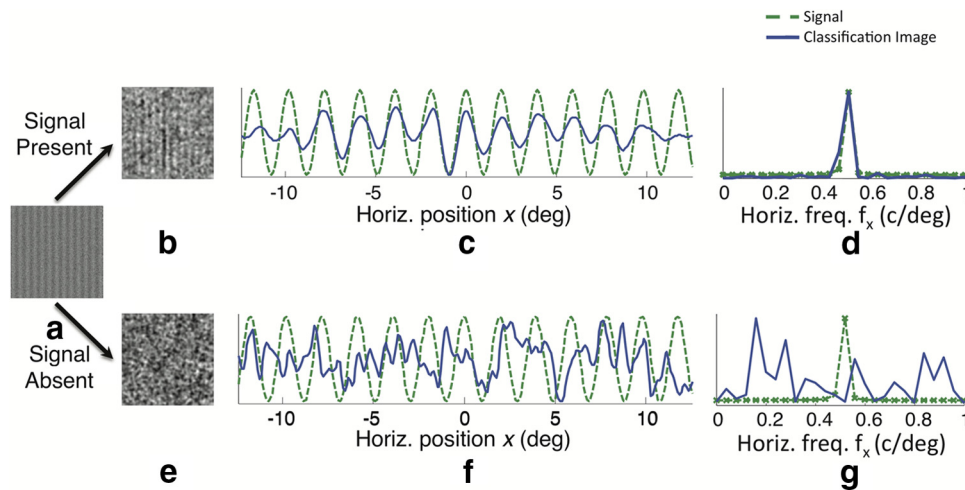


Figure 5. Experiment 3: detection of global (25.3°) grating patterns. Example results are for the 0.5 c/° grating stimulus in the intermediate (14%) noise condition, pooled over three subjects. *a*, Global 0.5 c/° grating stimulus in noise. *b*, Spatial classification image, computed from signal-present trials only. *c*, Projection of signal-present classification image onto horizontal axis. *d*, Horizontal Fourier amplitude spectrum of signal-present classification image. *e*, Spatial classification image, computed from signal-absent trials only. *f*, Projection of signal-absent classification image onto horizontal axis. *g*, Horizontal Fourier amplitude spectrum of signal-absent classification image. Results for 1.7 c/° gratings are similar. Horiz., Horizontal; freq., frequency.

$p = 0.7$, two-tailed. Thus, we are fairly confident that the nature of spatial summation is not substantially altered by the addition of stimulus noise. We now apply the noise method to probe the neural mechanisms underlying spatial pooling and detection.

Experiment 3: classification image analysis

Our first experiment demonstrated that detection of large-field grating stimuli is highly nonlinear. Whereas it is not possible to directly identify the nature of the nonlinearity from the measurement of contrast thresholds alone, the addition of noise to the stimuli greatly increases the information available from a behavioral experiment, as one can consider the joint distribution of noise samples and subject responses over all trials. Our third experiment was designed to take advantage of this extra information. The stimuli were again vertical gratings of constant cosine phase, but this time with no window (Fig. 5*a*). We tested three subjects on two grating frequencies (0.5 and 1.7 c/°) and three levels of noise contrast (4.3, 14.8, and 50%).

Linear analysis

In this experiment, each trial has four possible signal–response outcomes: the signal is absent or present and the subject either reports seeing it or not. The standard psychophysical method for linear system identification involves computing mean noise fields for these four cases over all trials and producing a “classification image” (Ahumada, 2002) of the underlying detection template by adding the two “present” response fields and subtracting the two “absent” response fields. This classification image forms an estimate of the underlying detection template, and it has been shown that if the system is linear, this estimate is both unbiased and efficient (Murray et al., 2002).

Our first experiment, however, demonstrates that the system cannot be linear, and in such cases, it has proven useful to produce two separate classification images, one based only on signal-present trials and the other based only on signal-absent trials (Ahumada and Beard, 1999; Solomon, 2002). Specifically, the signal-present classification image is formed by subtracting the “miss” response noise field from the “hit” response noise field, and the signal-absent classification image is formed by subtracting the “correct reject” noise field from the “false alarm” noise field (Fig. 6).

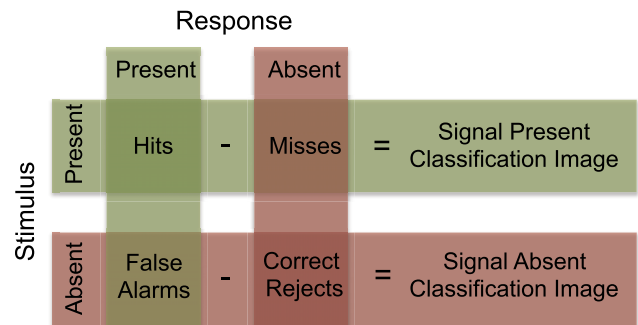


Figure 6. Calculation of signal-present and signal-absent classification images.

Figure 5, *b* and *e*, shows example signal-present and signal-absent classification images, respectively, for the 0.5 c/° stimulus in 14.8% noise, pooled over all three subjects. Whereas under the linearity assumption these two classification images are expected to look the same, they are, in fact, completely unlike. The signal-present classification image contains a structured template matched to the frequency and phase of the stimulus, attenuated a bit near the stimulus boundary but, nevertheless, sensitive to many cycles of the stimulus. The signal-absent classification image, on the other hand, contains no discernible structure. [Similar results have previously been obtained for small high-frequency or parafoveal Gabor stimuli (Ahumada and Beard, 1999; Solomon, 2002).] These differences are made clearer by projecting the classification images onto the horizontal axis (Fig. 5*c,f*). Taking the Fourier transform (Fig. 5*d,g*) makes it even more evident that the signal-present classification image is closely tuned to the spectral content of the signal, but the signal-absent classification image is not. This provides further evidence that the system is not linear over visual space, so that standard system identification techniques do not apply. How then can the underlying detection mechanisms be estimated?

Nonlinear analysis

Extending the linear methods widely used in neuroscience to nonlinear systems is a problem of great current interest. In the physiological literature, a number of techniques for identifying nonlinear neural kernels mapping stimulus sequences to spike

trains have been explored, including Wiener/Volterra expansions (Marmarelis and Marmarelis, 1978) and spike-triggered covariance (STC) analysis (de Ruyter van Steveninck and Bialek, 1988; Victor, 1992; Bialek and de Ruyter van Steveninck, 2005; Schwartz et al., 2006). More recently, related techniques have been explored in the psychophysics literature (Neri, 2004; Nandy and Tjan, 2007). Although theoretically sound, these techniques are typically limited by the curse of dimensionality: higher-order nonlinear mappings greatly increase the parameter space of the model, demanding greater quantities of data than can be realistically supplied. Whereas a neuron may fire many times per second, the quantity of data is limited by the length of time the neuron can be held. For behavioral work, where we have only a few responses per minute, even linear system identification methods can require many thousands of trials, putting unconstrained higher-order identification beyond the patience of human subjects. Thus, the key to making nonlinear system identification work for psychophysics is to apply appropriate and substantial a priori constraints that limit the dimensionality of the model.

Power spectrum analysis

Previous work can provide some insight into the constraints that might apply here. Solomon (2002) has shown that when signal-absent trials for small, higher-frequency parafoveal Gabor stimuli fail to yield structured templates in the spatial domain, they can still show tuning to the stimulus frequency in the power spectrum domain. One theory (Solomon, 2002) is that this tuning reflects a probability summation process (Green and Swets, 1966; Quick, 1974; Graham and Rogowitz, 1976; Graham, 1977; Sachs et al., 1980) in which the subject judges the stimulus to be present when any one of a number of detectors tuned to different stimulus phases responds above some threshold. However, in such a system, the behavioral decision is determined solely by the detector with maximal response, and performance is poor compared with a system that pools responses quantitatively (Green and Swets, 1966).

Local energy model

We propose an alternative theory for global detection based on a more efficient energy formulation that appears common to a diversity of neural computations in auditory (Green and Swets, 1966) and visual (Adelson and Bergen, 1985; Morrone and Burr, 1988; Landy and Bergen, 1991; Emerson et al., 1992; Heitger et al., 1992; Manahilov and Simpson, 1999; Goris et al., 2009) cortex. In this local energy model, the decision variable r is formed by summing the squared responses of an array of identical local linear filters tiling retinal space. Mathematically, this can be represented as:

$$r = \sum_{x,y} (f(x,y) * i(x,y))^2, \quad (4)$$

where $f(x, y)$ represents the receptive field of the local filter over retinal space, $i(x, y)$ represents the visual stimulus, and $*$ denotes convolution. Note that the model is stationary in space (shift invariant), which is an important and useful constraint on the nonlinear nature of the pooling network. In particular, this constraint means that the covariance matrix is determined by the autocorrelation, which is given by the inverse Fourier transform of the power spectral density. Thus, our technique can be (roughly) considered as a second-order Volterra technique under the assumption of shift invariance, which reduces the degrees of freedom from quadratic to linear in the number of image pixels. This assumption is thus crucial in limiting the degrees of freedom to allow reliable estimation.

Because of the squaring operation, the form of the local filter $f(x, y)$ cannot be directly estimated using standard classification image analysis. Remarkably, however, the model can be converted to linear form. Specifically, if $F(\omega_x, \omega_y)$ and $I(\omega_x, \omega_y)$ are the 2D Fourier transforms of the local spatial filter $f(x, y)$ and the visual stimulus $i(x, y)$, then

$$r = \sum_{\omega_x, \omega_y} |F[f(x,y) * i(x,y)]|^2 = \sum_{\omega_x, \omega_y} |F(\omega_x, \omega_y)|^2 |I(\omega_x, \omega_y)|^2, \quad (5)$$

where the first equality follows from Parseval's theorem and the second from the convolution theorem, both fundamental results in Fourier theory (Bracewell, 1978).

Equation 5 reveals that the local energy model is linear over spatial frequency in the Fourier power domain. Thus, if the model is (approximately) correct, the classification image technique can be adapted to the power spectrum, allowing direct estimation of the spectral density $|F(\omega_x, \omega_y)|$ and hence the frequency tuning and bandwidth of the early linear detection mechanisms that form the input to the nonlinear pooling network. Under the reasonable assumption that these early linear filters are localized in space, the length and width tuning can also be estimated (see below). Thus, surprisingly, it is possible, in principle, to recover the spatial properties of the local neural detection mechanisms in an energy network based only on behavioral responses to global stimuli.

Estimating the local energy model

Equation 4 implies that sensitivity is uniform across retinal space, whereas in the spatial classification images we observed in Experiment 3, sensitivity declines gradually with eccentricity (Fig. 5*b,c*). This effect can be incorporated into the local energy model by introducing a prewindowing function $g(x, y)$ that attenuates the stimulus as a function of eccentricity. Then we have:

$$r = \sum_{x,y} (f(x,y) * i'(x,y))^2 = \sum_{\omega_x, \omega_y} |F(\omega_x, \omega_y)|^2 |I'(\omega_x, \omega_y)|^2, \quad (6)$$

where

$$i'(x,y) = g(x,y)i(x,y). \quad (7)$$

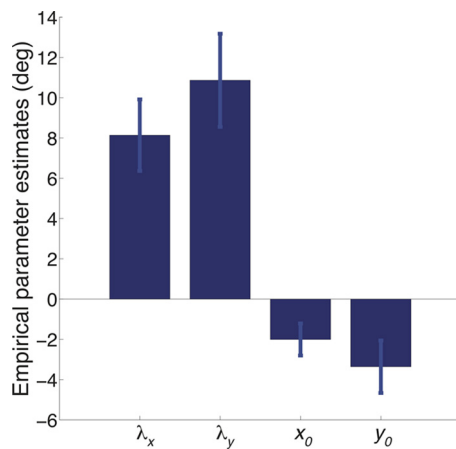
Based on previous work showing exponential falloff in sensitivity with eccentricity (Mostafavi and Sakrison, 1975), we model the spatial attenuation function $g(x, y)$ as:

$$g(x,y) = \exp\left(-\sqrt{\left(\frac{x-x_0}{\lambda_x}\right)^2 + \left(\frac{y-y_0}{\lambda_y}\right)^2}\right) \quad (8)$$

where x_0 and y_0 specify the location of peak sensitivity relative to the fovea and λ_x and λ_y are the exponential space constants in horizontal and vertical directions, respectively. To estimate these parameters, we fit the function $g(x, y) \cos 2\pi f_s x$ to the signal-present classification images computed in Experiment 3, using a standard least-squares gradient descent procedure, where f_s is the known signal frequency. The estimated parameters of $g(x, y)$ are listed in Table 3 and shown in Figure 7. We note that the summation field is, on average, displaced slightly to the lower left visual field, consistent with previous results showing higher contrast sensitivity for low to moderate spatial frequencies in the lower visual field (Rijsdijk et al., 1980; Thomas and Elias, 2011), and the general finding that tasks with a lower visual field advan-

Table 3. Empirically estimated parameters of global spatial attenuating function

Frequency (c°)	Observer	Noise contrast (%)	Laplace window parameters (degrees)			
			x_0	y_0	λ_x	λ_y
0.5	I.B.	4.3	-6.1	-0.56	6.9	7.5
		14.8	0.32	-7.4	9.4	6.2
		50	-0.92	-9.7	7.2	15
	J.R.	4.3	-9.6	0.25	12	6.2
		14.8	-5.1	4.8	7.7	18
		50	-2.5	13	5.6	47
	Y.M.	4.3	-0.82	-1.6	4.9	6.5
		14.8	-0.32	-5.1	3.2	7.5
		50	0.31	-5.9	4.1	5.8
1.7	I.B.	4.3	3.2	-1.4	13	11
		14.8	-8.3	-8.4	36	8.3
		50	1.8	-10	12	19
	J.R.	4.3	-2.8	-2.1	5.7	7.7
		14.8	-3.5	-2.7	4.6	5.4
		50	-1.8	-7.8	4.5	7.7
	Y.M.	4.3	-0.29	-4.6	4.7	5.7
		14.8	0.0	-5.1	3.8	6.1
		50	0.0	-5.9	2.0	6.5

**Figure 7.** Empirically estimated parameters of global spatial attenuating function, averaged over observers, noise contrast, and frequency. The x_0 and y_0 specify the location of peak sensitivity relative to the fovea, and λ_x and λ_y are the exponential space constants in horizontal and vertical directions, respectively. Error bars indicate SEM.

tage also tend to be biased to the left visual field (Christman and Niebauer, 1997; Thomas and Elias, 2011). We find that incorporating this global attenuation function to capture the falloff in sensitivity seen in our classification images also leads to an improvement in the trial-by-trial consistency of human and model responses (see below, Evaluation and comparison with competing models).

We applied the estimated attenuation function to each signal-absent stimulus noise pattern before taking the Fourier transform and accumulating the noise fields in the power spectrum domain, separately for false-alarm and correct reject trials. A signal-absent power-domain classification image was then formed by taking the difference of the means of these two noise fields (Fig. 6).

Figure 8 shows the results of this analysis for both 0.5 and 1.7 c° grating stimuli and 14.8% noise contrast, using data pooled over subjects. Figure 8*a* shows the resulting power spectrum classification image, where light and dark patches indicate regions with estimated weights >3 SDs from the mean. There is a concentration of energy near the signal frequency, made more apparent by selecting a region of interest near the signal frequency (Fig. 8*b*).

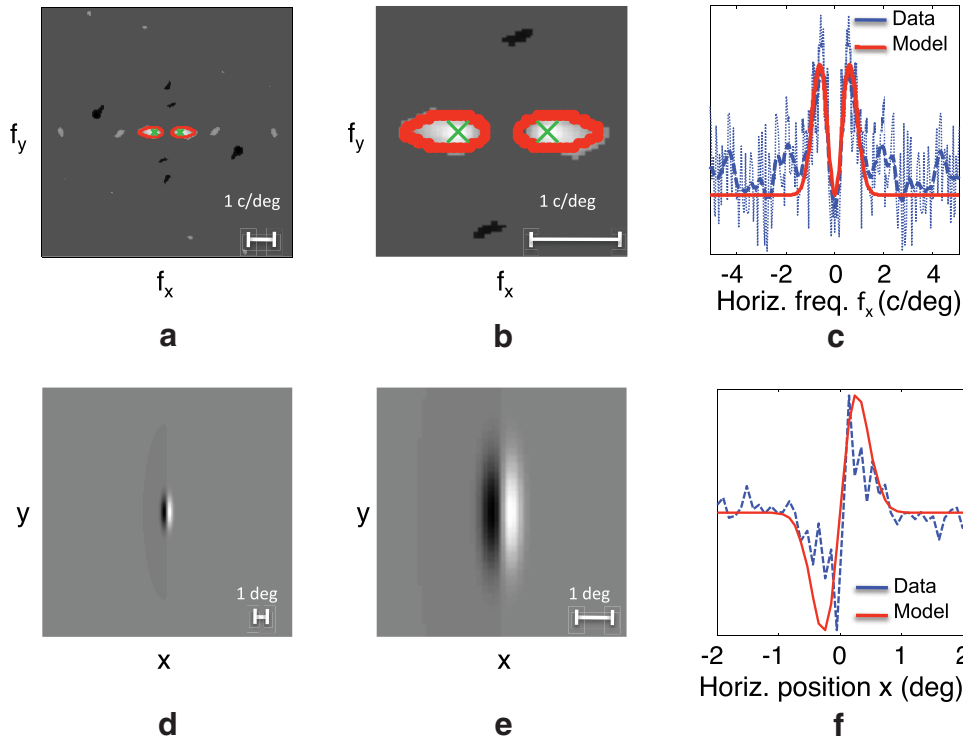
This concentration of energy near the signal frequency is roughly Gaussian in appearance, consistent with a localized Gabor filter. It is well known that maximal simultaneous localization in space and spatial frequency (in the sense of minimizing the product of the variances) is uniquely achieved by the Gabor function (Gabor, 1946), and the Gabor function is widely used as a model for early visual receptive fields (Daugman, 1980; Watson et al., 1983). Here we model the detector energy in the Fourier domain using the Fourier transform of an odd-symmetric Gabor filter:

$$f(x, y) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \sin 2\pi f_0 x, \quad (9)$$

where σ_x and σ_y are the Gabor space constants in horizontal and vertical directions, respectively, and f_0 is the carrier frequency of the Gabor. We use a standard least-squares gradient descent procedure to fit the model to the signal-absent power spectrum classification image data, determining maximum likelihood estimates for the parameters. The red contours in Figure 8, *a* and *b*, represent the bandwidth (at half-amplitude) of the maximum likelihood fit. Figure 8*c* shows the projection of both data and fit onto the horizontal spatial-frequency axis.

The horizontal and vertical spreads of the maximum likelihood Gabor fit to the power spectrum classification image completely determine the horizontal and vertical space constants of the corresponding spatial Gabor. In particular, the spread in the space domain is the reciprocal of the spread in the Fourier (amplitude) domain. Although by assumption we cannot recover the phase structure of these local filters with our method, the population of orientation-tuned simple cells in primary visual cortex of primate are known to have a large bias to odd symmetry (Ringach, 2002). Assuming odd symmetry allows us to take the inverse 2D Fourier transform of the estimated model to reveal the appearance of the filters in the spatial domain (Fig. 8*d,e*). Note that although the stimulus is $\sim 25^\circ$ in both width and height, the estimated local filters are highly localized and elongated, similar to receptive fields of neurons in early visual cortex. Figure 8*f* shows the projection of the data and model to the horizontal space axis. (This last step required half-wave rectification of the power spectrum classification data shown in Fig. 8*c*.)

0.5 c/deg grating



1.7 c/deg grating

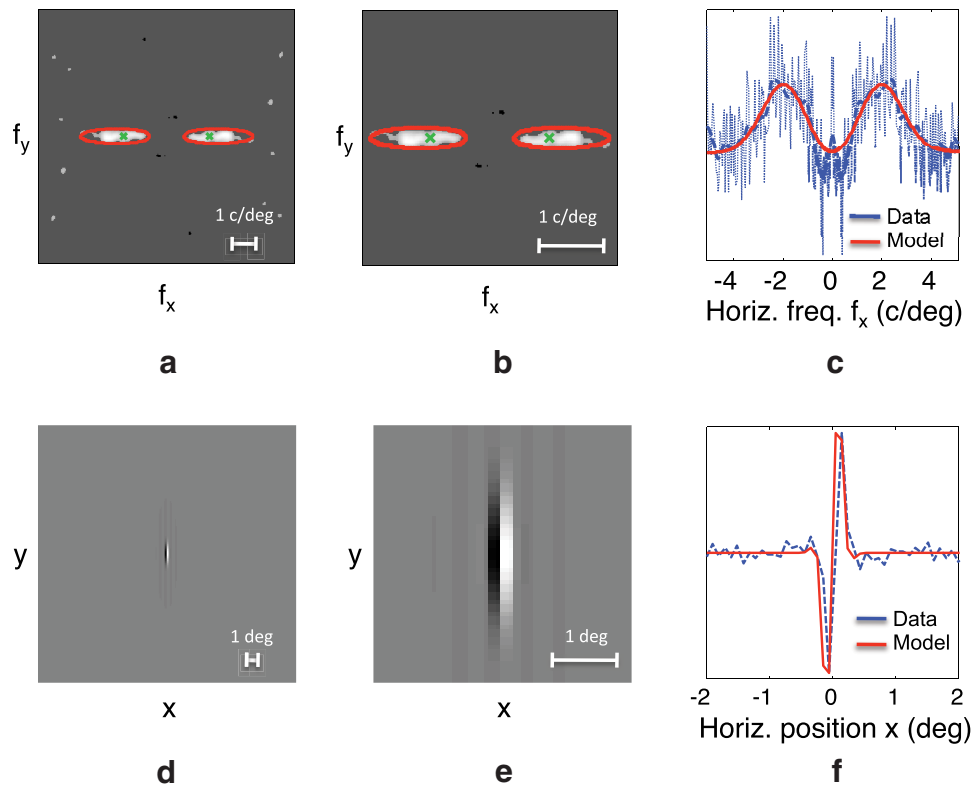
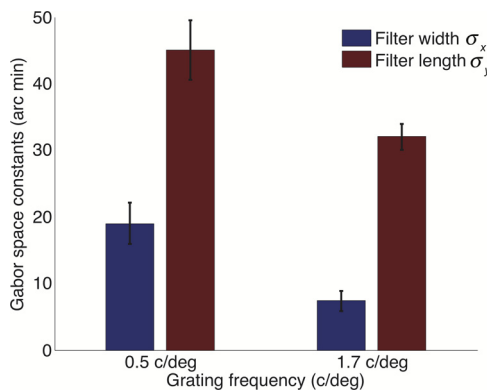


Figure 8. Recovering local mechanisms by power spectrum analysis. Example results are for the intermediate (14%) noise condition, pooled over three subjects. **a**, Power spectrum classification image. DC is at the image center. For visualization purposes only, the classification image has been smoothed using a Gaussian smoothing kernel with a space constant of 3 pixels and thresholded at ± 3 times the SD of the classification image. Green X symbols mark the signal frequency. Red contours indicate bandwidth (at half-amplitude) of a maximum likelihood sine (*Figure legend continues.*)

Table 4. Empirically-estimated parameters of local receptive field model

Frequency (c/°)	Observer	Noise contrast (%)	Octave bandwidth	Local Gabor filter parameters			
				σ_x (arcmin)	σ_y (arcmin)	f_0 (c/°)	f_p (c/°)
0.5	I.B.	4.3	2.5	13	38	0.78	0.9
		14.8	1.8	14	21	1.5	1.5
		50	2.6	11	54	0.02	0.91
	J.R.	4.3	2.6	8.8	40	0.06	1.1
		14.8	2.6	19	69	0.08	0.52
		50	2.3	26	40	0.59	0.6
	Y.M.	4.3	1.4	39	57	0.64	0.64
		14.8	2.6	15	44	0.03	0.63
		50	2.5	23	46	0.56	0.59
Mean			2.3	19	45	0.47	0.82
SE			0.14	3.1	4.5	0.16	0.11
1.7	I.B.	4.3	1	18	26	1.8	1.8
		14.8	2.4	7.4	42	1.8	1.8
		50	2.6	5.6	26	0.17	1.7
	J.R.	4.3	2.6	4.5	24	0.18	2.1
		14.8	2.4	5.3	30	2.5	2.6
		50	2.6	3.4	39	0.86	2.9
	Y.M.	4.3	2.1	9.6	34	1.8	1.8
		14.8	2.1	9	34	2	2
		50	2.6	3.7	31	0.03	2.6
Mean			2.3	7.4	32	1.2	2.1
SE			0.17	1.5	2	0.31	0.15

**Figure 9.** Empirically estimated spatial parameters of local detection filters. Error bars indicate SEM.

The maximum likelihood estimates for the spatial and spatial-frequency parameters of the local Gabor filters are listed in Table 4 and shown in Figure 9. For low-frequency odd Gabor filters, subtractive interactions between positive and negative frequencies shift the peak frequency tuning f_p to higher frequencies relative to the carrier frequency f_0 . We note a trend for the estimated local linear filters underlying detection to become slightly more broadband for higher noise levels. (Bandwidths are full amplitude at half-height.) Note that the maximum bandwidth achievable by an odd Gabor filter is 2.6 octaves, at which point the filter is indistinguishable from a Gaussian first derivative. Roughly half of the filters we estimate approach this limit.

In summary, we find that through Parseval's theorem, classification image analysis can be used to estimate the spatial and

spatial-frequency parameters of the early local linear visual mechanisms that provide the input to a nonlinear pooling mechanism. We stress that these parameters can be estimated without manipulating the spatial extent or spatial frequency of the stimulus, which is the traditional way to estimate early visual mechanisms (Watson et al., 1983; Wilson et al., 1983; Kersten, 1984). Our method does rely on the assumption that the underlying linear filters are localized in space and can be approximated as Gabors. However, this assumption is well supported by previous research (Daugman, 1980; Watson et al., 1983).

Evaluation and comparison with competing models

The ultimate test of a model is its predictive power. For this evaluation, we again make use of the noise in our stimuli, computing model responses for each noisy image and comparing these with human responses on a trial-by-trial basis. We evaluate four distinct models (Fig. 10).

1. A global coherent model in which the decision variable is determined by the inner product of the stimulus with the global spatial template $\cos(2\pi f_s x)g(x, y)$ estimated from signal-present trials. Thus

$$r = \sum_{x,y} \cos(2\pi f_s x)g(x,y)i(x,y). \quad (10)$$

This is the standard linear model.

2. A global incoherent model, in which the decision variable is formed by the sum of squared responses of quadrature-pair global filters:

$$r = \left(\sum_{x,y} \cos(2\pi f_s x)g(x,y)i(x,y) \right)^2 + \left(\sum_{x,y} \sin(2\pi f_s x)g(x,y)i(x,y) \right)^2. \quad (11)$$

Sometimes called the narrowband incoherent detector, this is the maximum likelihood detector when the phase is completely uncertain (Kay, 1998) and has been proposed (Murray, 2011) as a possible explanation for the incoherent detection of small high-

(Figure legend continued.) **Gabor model fit to the classification image in the power spectrum domain. b.** Magnification of relevant region of power spectrum. **c.** Projection of power spectrum classification image onto horizontal axis. **d.** Estimated local linear spatial filter, shown in sine phase. **e.** Magnification of local linear spatial filter. **f.** Projection of estimated local linear spatial filter onto horizontal axis. Horiz., Horizontal; freq., frequency.

frequency or parafoveal Gabor stimuli (Ahumada and Beard, 1999; Solomon, 2002).

3. Our local energy model, in which the decision variable is determined by the sum of squared responses over a bank of local linear detection filters $f(x, y)$ within the global spatial window $g(x, y)$:

$$r = \sum_{x,y} (f(x,y) * (g(x,y)i(x,y)))^2. \tag{12}$$

4. A local probability summation model, in which the decision variable is determined by the maximum response over a pool of local linear detection filters within the global spatial window $g(x, y)$:

$$r = \max_{x,y} (f(x,y) * (g(x,y)i(x,y))). \tag{13}$$

Probability summation is a standard model for spatial pooling (Quick, 1974; Graham and Rogowitz, 1976; Sachs et al., 1980).

For all models, we use the global exponential spatial attenuation function $g(x, y)$ of Equation 8. For the last two models, we use the local Gabor filters $f(x, y)$ of Equation 9. We repeat the equations below for convenience:

$$g(x,y) = \exp\left(-\sqrt{\left(\frac{x-x_0}{\lambda_x}\right)^2 + \left(\frac{y-y_0}{\lambda_y}\right)^2}\right),$$

$$f(x,y) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \sin 2\pi f_0 x, \tag{14}$$

We used the empirically estimated parameters for $g(x,y)$ and $f(x,y)$ recorded in Tables 3 and 4, respectively.

To compare the real-valued decision variables produced by these four models with the binary decisions made by our human subjects, we first partitioned trials into signal-present and signal-absent subsets. Then, for each of these subsets, we computed the t score for the difference in the mean model response when observers responded present versus when they responded absent. To be consistent with human judgments, the model should generate high values when the user responds present and low values when the user responds absent, thus producing a large positive t score.

The results of this analysis are shown in Figure 11. Note that three of the four

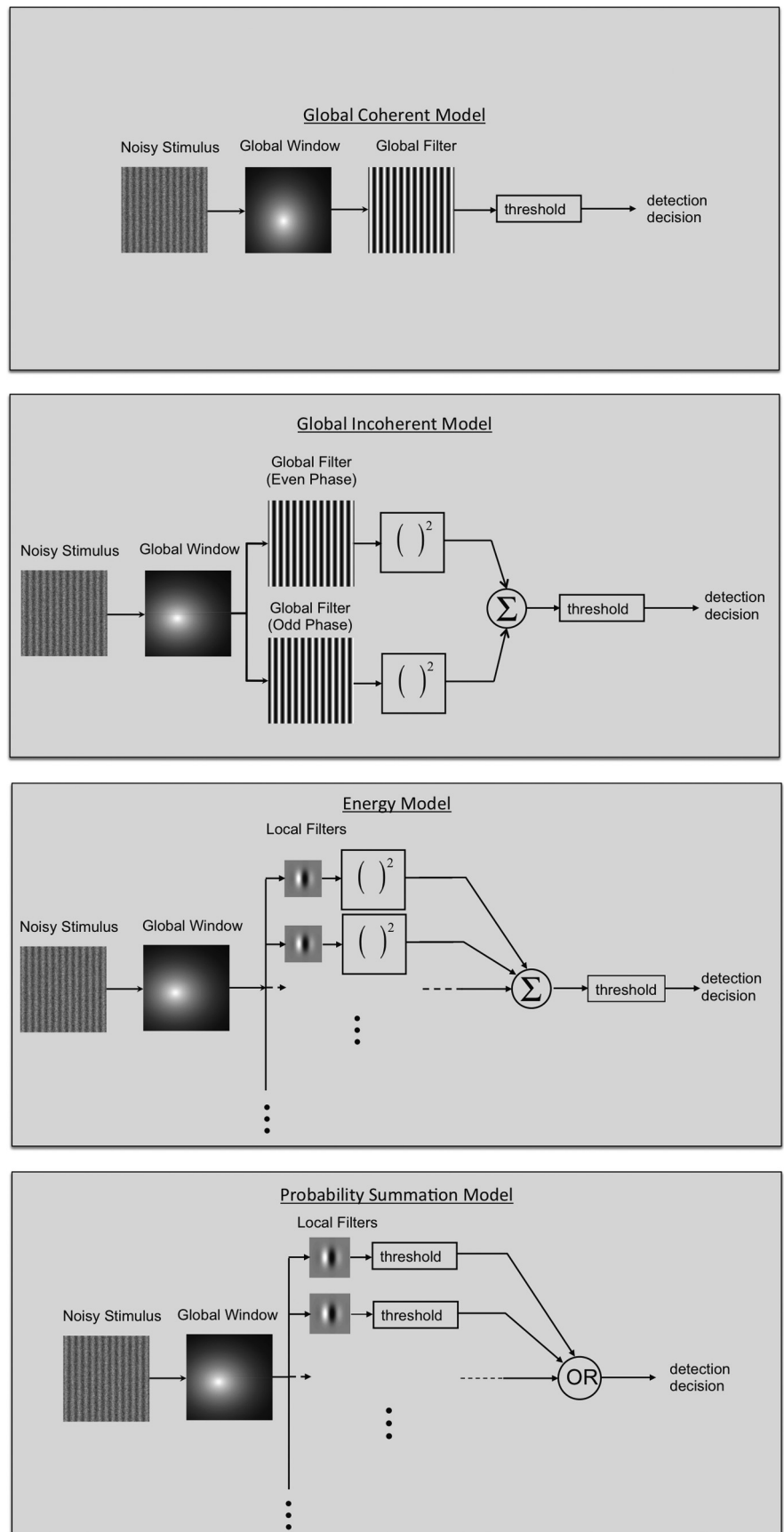


Figure 10. The four detection models evaluated.

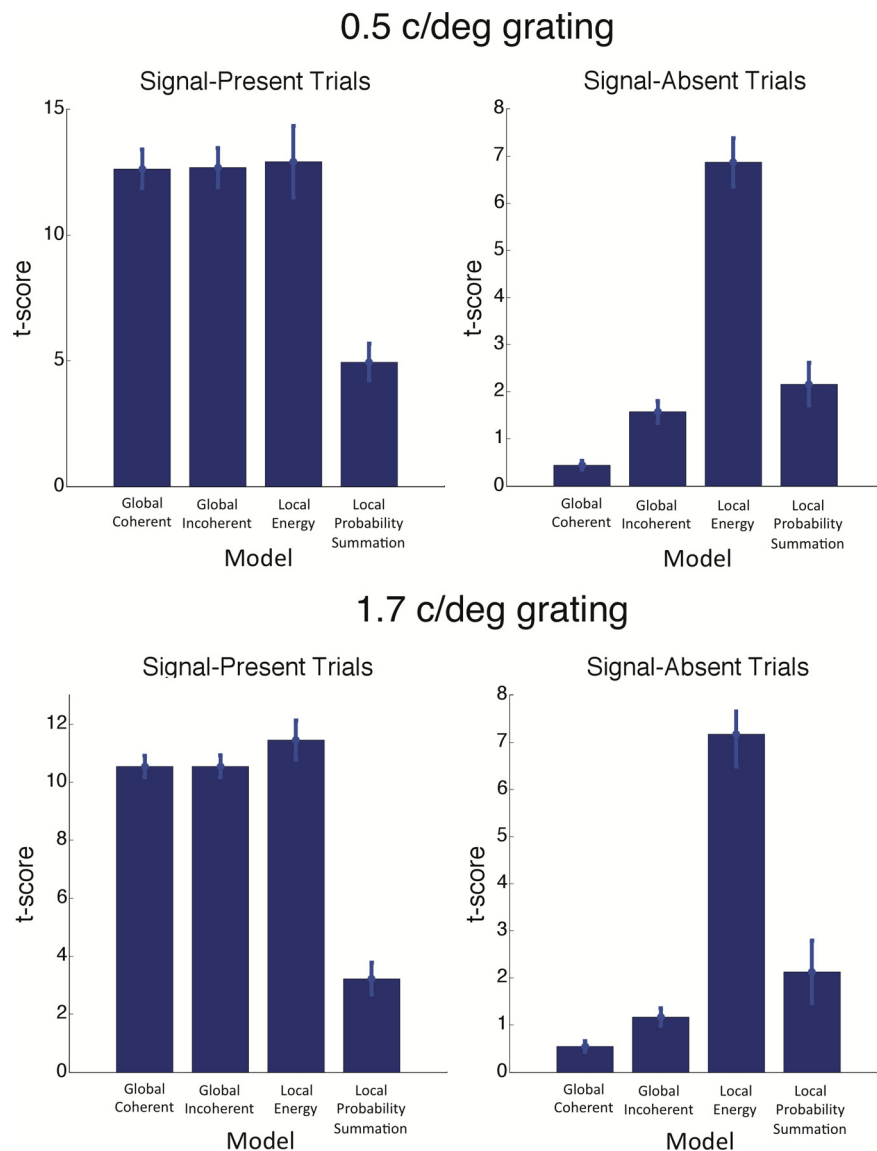


Figure 11. Evaluation of spatial pooling models based on trial-by-trial consistency with human judgments, for Experiment 3. The t scores are based on the difference between the mean model response for trials on which the human subject reports the signal present and the mean model response for trials on which the human subject reports the signal absent. Higher t scores indicate greater consistency between model and human data. Error bars indicate SEM. Planned pairwise t tests reveal that the local energy model is significantly more consistent with human judgments than the other three models for signal-absent trials, for both 0.5 c° grating stimuli ($p < 4e - 6$, Bonferroni adjustment $\alpha = 0.05/3 = 0.017$) and 1.7 c° grating stimuli ($p < 8e - 5$, Bonferroni adjustment $\alpha = 0.05/3 = 0.017$).

models (global coherent, global incoherent, and local energy) show similar t scores for the signal-present case. This is not surprising, since all three models and the human responses correlate positively with signal contrast, which varied moderately from trial to trial. (Since the local probability summation model depends only on the maximum local response, it correlates more weakly with signal contrast.) Thus, the crucial test is the signal-absent case. Here we see that the local energy model predicts the human data much better than the alternatives, including probability summation.

To determine whether the failure of probability summation to match human behavior might be attributable to the local filter parameters used, we also evaluated the local probability summation model using parameters extracted from Kersten (1984). Specifically, we assumed a Gabor filter that was matched to the

frequency of the signal, with a 1.7 octave bandwidth, as estimated by Kersten, and equal filter length and width. We found that the t score for the local probability summation model in the signal-absent condition was roughly the same with this filter as with the filters we estimated from the power spectrum. We also checked for dependence on the length of the underlying filter, varying the aspect ratio (length to width) from 1 to 8. In the signal-absent condition, the consistency of the local probability summation model with the human data varied only modestly, achieving a peak t score between 2 and 3 at an intermediate aspect ratio, well below the t score of 7 achieved by the local energy model. In summary, it seems that the performance of the probability summation model is not a strong function of the precise filter parameters. Rather, the failure to match the human data seems to derive from the maximum-response rule on which the model is based.

Whereas our t score analysis shows that the local energy model is better than competing models, this is a relative judgment that does not directly indicate the goodness of fit of the local energy model to the human data. To assess this, we fit a two-parameter cumulative normal psychometric function to the human response data for Experiment 3, as a function of the response of the local energy model (Fig. 12).

For plotting purposes only, we have separated the data based on whether the signal was absent or present in the stimulus. We estimated the probability of a “signal present” human response using a variant of kernel density estimation. In particular, each trial in which the observer’s response was negative contributed a Gaussian estimate of the local density of negative responses, with mean equal to the model response for that trial. The SD was selected such that the model responses for 10 other trials were within 1

SD on either side. The density of positive responses was estimated in similar fashion, and the probability of a positive response was computed as the ratio of the positive density to the sum of the positive and negative densities.

From Figure 12, we see that, qualitatively at least, the cumulative normal psychometric function accurately captures the relationship between the model response and the human response in almost all cases. We assessed the fit of the model quantitatively for each observer and condition by Monte Carlo sampling 2000 datasets of the same size as the empirical dataset from the estimated psychometric function and by computing the resulting squared error (Wichmann and Hill, 2001). (Since the empirical model responses are not uniformly distributed, we sample with replacement from the empirical model response values and then sample from the cumulative normal model at these values of the model response.) We

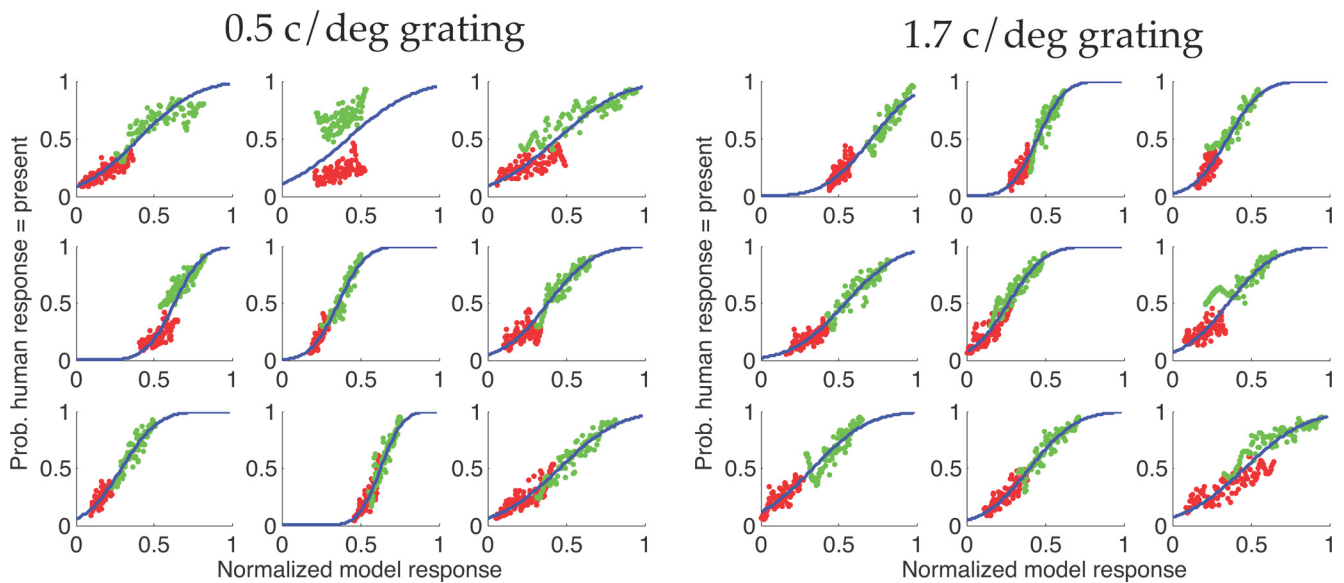


Figure 12. Psychometric functions relating response of local energy model to human responses for Experiment 3. Top to bottom, Observers I.B., J.R., and Y.M. Left to right, Noise contrast equals 4.3, 14.8, and 50%. Data from trials in which the signal was absent are plotted in red, and signal-present data are in green. The least-squares fit of the cumulative normal psychometric function is shown in blue.

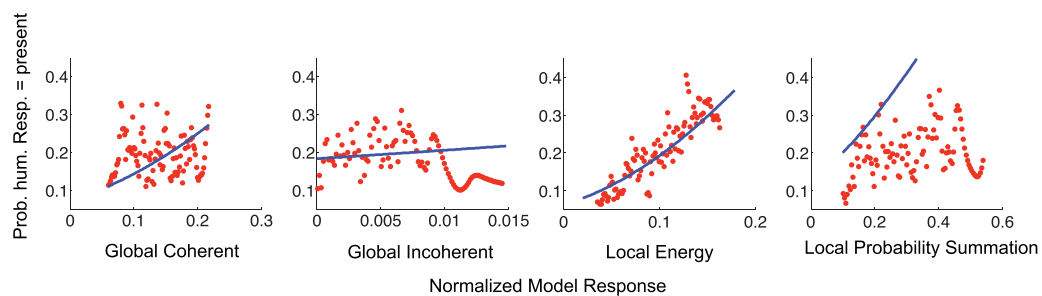


Figure 13. Psychometric functions relating model responses to human responses for an example condition of Experiment 3 (0.5 c/deg , observer J.R., 14.8% noise contrast). Only data for signal-absent trials are shown.

find that the proportion of our sampled datasets for which the squared error exceeds our empirical error ranges from 0.15 to 0.59 across our 18 conditions, indicating a reasonable fit to the combined (signal-present plus signal-absent) data.

However, when we separate out the signal-present and signal-absent stimulus conditions (Fig. 12), we do observe one of the 18 conditions in which the model fails (0.5 c/deg , observer I.B., 14.8% noise). From Table 4, we see that this is the one condition for which our estimate of the local Gabor filter parameters seems to be in error: the peak frequency is three times higher than the signal frequency, and the estimated length of the filter is roughly half that estimated for the other conditions involving the 0.5 c/deg grating. Thus, we suspect that the poor correspondence between model and human data for this one condition stems from these errors in model parameter estimates. Increasing the number of trials and/or regularizing the estimation would likely correct this problem.

We emphasize that the crucial support for the model derives from the positive correlation between model and human response for the signal-absent condition (Fig. 12, shown in red), which is clear in every one of our 18 conditions, and in the fact that generally the signal-absent and signal-present data together form a single smooth, continuous distribution that can be modeled by a common psychometric function. As our *t* score analysis reveals, this strong correlation between

model and human response for the signal-absent condition is not found for the other models. Figure 13 illustrates this for a typical example condition, where we have zoomed in on the signal-absent data alone. We see that a strong correlation between model and human response is evident only for the local energy model. For the local probability summation model, we also see a displacement of the psychometric function from the signal-absent data, indicating that the signal-absent and signal-present data cannot be accounted for by a single psychometric function, a strong sign that probability summation is not the basis for human judgments in this task.

Neural basis

The energy-based classification image method allows properties of local detection mechanisms to be estimated solely from human judgments of global visual stimuli. These local mechanisms may be identified with single neurons in early visual cortex. To match the properties of the local filters we have estimated behaviorally, these neurons must integrate visual information in a substantially linear fashion within their receptive fields, and these receptive fields must be orientation tuned. These requirements are met by the population of simple cells in primary visual cortex (Hubel and Wiesel, 1968; Movshon et al., 1978).

To assess quantitative correspondence, we rely on data for receptive fields of simple cells in macaque primary visual cortex

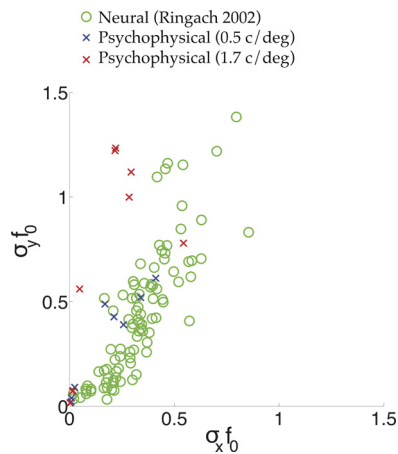


Figure 14. Comparison of our psychophysical estimates of spatial and spatial-frequency parameters of local mechanisms with neurons in macaque primary visual cortex (Ringach, 2002). Both are based on Gabor models, where f_0 is the carrier frequency and σ_x and σ_y are the Gaussian space constants determining the width and length of the mechanisms, respectively.

(Ringach, 2002). Here the spatial selectivity of each neuron is characterized by the product of the carrier frequency f_0 and the width and length space constants σ_x and σ_y of a Gabor receptive field model. Figure 14 compares the distribution of 93 neurons with the 18 psychophysical estimates made from our third experiment (3 subjects \times 2 spatial frequencies \times 3 noise contrast levels).

We find that the physiological and psychophysical estimates are generally compatible, although some of the psychophysical estimates for the 1.7 c° condition are slightly more elongated than physiological measurements. This could be attributable to a number of factors: interspecies differences, sampling bias toward elongated receptive fields caused by our grating stimuli, and/or to differences in visual eccentricity of the human neurons underlying detection in our experiments and the macaque neurons studied physiologically, since spatial receptive field properties are known to vary with visual eccentricity (Smith et al., 2001).

Spatial summation

The results of Experiments 1 and 2 reveal summation slopes (log contrast threshold vs log stimulus width) for grating detection in the range of -0.27 to -0.28 for the 0.5 c° grating stimuli, and -0.37 to -0.41 for the 1.7 c° stimuli. What summation slope would a local energy model predict?

To address this question, we analyzed how the sensitivity d' of a grating detector varies as a function of the contrast c of the grating and the number n of independent local filters. Since the experiments were blocked by the width of the windowed grating, we make the assumption that subjects can approximately limit pooling to the signal region. (Below, we consider what happens when this assumption does not hold.) Then n is proportional to the width of the stimulus.

Without loss of generality, we assume that the detectors are normalized to have unit expected response to a grating of unit contrast and let σ_n represent the SD of each (normal, independent, and identically distributed) detector response to the stimulus noise.

Coherent linear summation

It is useful to first consider the simpler case of an optimal linear, coherent pooling mechanism over identical, spatially distributed

mechanisms. This is an easier case because when the detector is linear and the noise is normally distributed, the decision variable is also normally distributed.

Letting r_i represent the response of local detection filter i , the system response (decision variable) r will be:

$$r = \sum_{i=1}^n r_i, \quad (15)$$

In the signal-absent case, the decision variable has expected value $\bar{r}_a = 0$ and variance $\sigma_a^2 = n\sigma_n^2$. In the signal-present case, for a signal of contrast c , the decision variable has expected value $\bar{r}_p = cn$ and variance $\sigma_p^2 = n\sigma_n^2$.

The signal-to-noise ratio can be defined as:

$$d' = \sqrt{2SNR} = \sqrt{2} \frac{\bar{r}_p - \bar{r}_a}{\sqrt{\sigma_p^2 + \sigma_a^2}} = \frac{cn}{\sqrt{2n\sigma_n^2}} = \frac{c\sqrt{n}}{\sqrt{2}\sigma_n} \propto c\sqrt{n}. \quad (16)$$

Assuming constant observer bias, a criterion level of performance (e.g., 75% correct) is achieved at a fixed value of d' . Thus, as stimulus width (the number n of stimulated detectors) increases, contrast threshold c_t decreases as $1/\sqrt{n}$. On a log–log plot, c_t falls with a slope of -0.5 , steeper than the falloff we observe.

Incoherent energy summation

For an incoherent local energy detector, the relationship between d' , n , and c turns out to be more complicated. In particular, d' is given by:

$$d' = \frac{c^2\sqrt{n}}{\sigma_n\sqrt{2(2c^2 + 7\sigma_n^2)}}. \quad (17)$$

(See mathematical proof below.)

Note that although d' is still proportional to \sqrt{n} , it is no longer proportional to the contrast c . As a consequence, although d' will rise as \sqrt{n} when c is fixed, the threshold c_t will no longer fall as $1/\sqrt{n}$ when d' is fixed (i.e., at criterion performance). Specifically, solving Equation 17 for c at threshold, we obtain:

$$c_t^2 = \frac{d'\sigma_n^2}{n} \left(2d' + \sqrt{4d'^2 + 14n} \right). \quad (18)$$

Thus, for large n ,

$$c_t^2 \approx \sqrt{14d'\sigma_n^2/\sqrt{n}} \rightarrow c_t \propto 1/n^{1/4}. \quad (19)$$

On a log–log plot, c_t falls with a slope of -0.25 .

For smaller n , the relationship between contrast threshold c_t and n is not exactly a power law, and so the summation function $c_t(n)$ will not be exactly linear in log–log coordinates. However, in the neighborhood of a particular value of n , the summation function will be approximately linear, with a slope that can be calculated analytically.

In particular, taking the log of both sides of Equation 18, we obtain:

$$\log c_t = \frac{1}{2} \left(\log d' + 2\log \sigma_n - \log n + \log \left(2d' + \sqrt{4d'^2 + 14n} \right) \right). \quad (20)$$

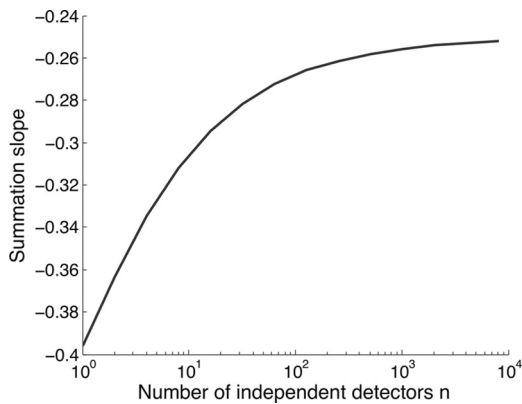


Figure 15. Slope of summation function $c_t(n)$ for an incoherent energy pooling mechanism as a function of the number n of independent detectors.

Approximating n as a continuous variable, we have:

$$\frac{d \log c_t}{d \log n} = -\frac{1}{2} + \frac{7n}{2\sqrt{4d'^2 + 14n(2d' + \sqrt{4d'^2 + 14n})}} \quad (21)$$

Thus, the theoretical summation slope is -0.5 for $n = 0$, declining monotonically to -0.25 for large n (Fig. 15). Of course, if $n = 0$, nothing will be detected. A value of $n = 1$ produces a more realistic upper bound on the slope: for $d' = 1.349$, corresponding to an unbiased observer at 75% correct, the slope at $n = 1$ is -0.39 .

For the $0.5 c^\circ$ grating stimuli, we observed summation slopes of -0.28 to -0.29 (Table 1), squarely in the predicted range. For the $1.7 c^\circ$ gratings, we observed somewhat steeper slopes, between -0.37 and -0.41 . This may be caused by observer uncertainty in the horizontal position of the narrow windowed stimuli; in other words, observers are not able to precisely limit pooling to the signal region. A fixed uncertainty will reduce performance more for narrower stimuli, resulting in a steeper summation curve. This will be less an issue for the $0.5 c^\circ$ gratings, which are more than three times wider than the $1.7 c^\circ$ gratings. We explore this hypothesis quantitatively below (see The role of spatial uncertainty).

Mathematical proof of local energy summation law

Here we prove Equation 17 relating d' , n , and c for the incoherent local energy detector.

The local energy detector forms the decision variable as a sum of squared detector responses r_i :

$$r = \sum_{i=1}^n r_i^2 \quad (22)$$

In the signal-absent condition, since the individual filter responses are independent, identically distributed, zero-mean normal random variables, the decision variable follows a X_n^2 distribution, with mean and variance given by:

$$\bar{r}_a = n\sigma_n^2, \quad \sigma_a^2 = 2n\sigma_n^4 \quad (23)$$

In the signal-present condition, the response r_i of each filter is still normally distributed, but the sum of squared responses is not exactly X_n^2 because the expected response of each filter is different, depending on the phase of the filter relative to the signal. To see this, consider a one-dimensional signal $s(x) = c \sin(2\pi f_s x)$ and a

local Gabor filter $f_i(x)$ matched in frequency but with relative phase ϕ_i :

$$f_i(x) = a \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \sin(2\pi f_s x + \phi), \quad (24)$$

where a is an arbitrary gain parameter. The expected response of the filter to the noisy stimulus is equal to the response to the signal alone:

$$\mathbb{E}[r_i] = \int f_i(x)s(x)dx \quad (25)$$

$$= ac \int \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \sin(2\pi f_s x + \phi_i) \sin(2\pi f_s x) dx \quad (26)$$

$$= acc \cos \phi_i \int \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \sin^2(2\pi f_s x) dx \quad (27)$$

$$= acr_0 \cos \phi_i, \quad (28)$$

where r_0 is a constant equal to the response of a unit-amplitude phase-locked filter to a signal of unit contrast. Without loss of generality, we set $a = 1/r_0$, so that

$$\mathbb{E}[r_i] = c \cos \phi_i. \quad (29)$$

The second moment of the local filter response also depends on the phase:

$$\mathbb{E}[r_i^2] = \mathbb{E}[r_i]^2 + \sigma_n^2 = c^2 \cos^2 \phi_i + \sigma_n^2. \quad (30)$$

Although in the signal-present condition the decision variable r is not exactly X_n^2 , we can use the first and second moments of the local filter responses to derive its mean and variance. First, from Equation 30, we observe that the mean \bar{r}_p of the decision variable in the signal-present condition can be written as:

$$\bar{r}_p = \sum_{i=1}^n \mathbb{E}[r_i^2] = c^2 \sum_{i=1}^n \cos^2 \phi_i + n\sigma_n^2. \quad (31)$$

Assuming that the phase of the local linear filters is uniformly distributed, we can write:

$$\sum_{i=1}^n \cos^2 \phi_i \approx \frac{n}{2\pi} \int_{-\pi}^{\pi} \cos^2 \phi_i d\phi_i = n/2. \quad (32)$$

Substituting Equation 32 into Equation 31, we obtain:

$$\bar{r}_p = n(c^2/2 + \sigma_n^2). \quad (33)$$

To derive the variance σ_p^2 of the decision variable in the signal-present condition, we first note that since the filters are independent, it can be written as the sum of the variances of the squared filter responses:

$$\sigma_p^2 = \sum_{i=1}^n \mathbb{E}[(r_i^2 - \mathbb{E}[r_i^2])^2] = \sum_{i=1}^n \left(\mathbb{E}[r_i^4] - \mathbb{E}[r_i^2]^2 \right). \quad (34)$$

The fourth moment of the normally distributed local filter response is given by:

$$\mathbb{E}[r_i^4] = \mathbb{E}[r_i]^4 + 6\mathbb{E}[r_i]^2\sigma_n^2 + 2\sigma_n^4. \quad (35)$$

Substituting Equations 30 and 35 into Equation 34, we obtain:

$$\sigma_p^2 = \sum_{i=1}^n \left(\mathbb{E}[r_i]^4 + 6\mathbb{E}[r_i]^2\sigma_n^2 + 2\sigma_n^4 - \left(\mathbb{E}[r_i]^2 + \sigma_n^2 \right)^2 \right) \quad (36)$$

$$= \sum_{i=1}^n \left(4\mathbb{E}[r_i]^2\sigma_n^2 + \sigma_n^4 \right) \quad (37)$$

$$= 4\bar{r}_p\sigma_n^2 + n\sigma_n^4 \quad (38)$$

$$2nc^2\sigma_n^2 + 5n\sigma_n^4. \quad (39)$$

Combining Equations 23, 31, and 39 yields Equation 17 relating d' to n and c for the incoherent local energy detector:

$$d' = \sqrt{2} \frac{\bar{r}_p - \bar{r}_a}{\sqrt{\sigma_p^2 + \sigma_a^2}} \quad (40)$$

$$= \frac{c^2 n}{\sqrt{2(\sigma_p^2 + \sigma_a^2)}} \quad (41)$$

$$= \frac{c^2 \sqrt{n}}{\sigma_n \sqrt{2(2c^2 + 7\sigma_n^2)}}. \quad (42)$$

Role of spatial uncertainty

The incoherent local energy model incorporates uncertainty about the phase of the stimulus, but our analysis assumes that the observer does know the exact extent of the horizontal grating window used in Experiments 1 and 2 and is able to restrict summation exactly to this region. This approximation may be reasonable for larger stimulus windows, but for smaller stimulus windows it is likely that spatial uncertainty about the exact position and extent of the window will play a larger role in determining performance.

To gain quantitative insight into this effect, suppose that the uncertainty in the horizontal position of the stimulus can be described by a zero-mean Gaussian random variable with variance $\sigma_{\Delta x}^2$. The effect can be modeled by shifting the observer's coordinate frame left or right on each trial by this random amount. The expected signal received in the observer's coordinate frame is now a wider Gaussian-windowed grating with space constant $\sigma_x = \sqrt{\sigma_x^2 + \sigma_{\Delta x}^2}$, where σ_x is the horizontal space constant of the Gabor stimulus. Were the deviations of the stimulus from this expected signal Gaussian and white, the optimal summation window would match this larger window. Although, in fact, the deviations are more complex, this is still a good first-order strategy for dealing with spatial uncertainty.

To see the effect on summation, we let n represent the number of independent local filters being pooled and n_s the number of these falling within the stimulus window. With this notation, the mean and variance of the decision variable remain unchanged in the signal-absent condition (Equation 23), but in the signal-present condition, they become:

$$\bar{r}_p = n_s c^2 / 2 + n \sigma_n^2, \quad \sigma_p^2 = 2n_s c^2 \sigma_n^2 + 5n \sigma_n^4. \quad (43)$$

Substituting into Equation 40, we obtain:

$$d' = \frac{c_t^2 n_s}{\sigma_n \sqrt{2(2n_s c_t^2 + 7n \sigma_n^2)}}. \quad (44)$$

Solving for c_t as before (Equations 40–42), we obtain:

$$c_t^2 = \frac{d' \sigma_n^2}{n_s} \left(2d' + \sqrt{4d'^2 + 14n} \right) \quad (45)$$

$$\rightarrow \log c_t = \frac{1}{2} \left(\log d' + 2 \log \sigma_n - \log n_s + \log \left(2d' + \sqrt{4d'^2 + 14n} \right) \right). \quad (46)$$

To relate this to spatial uncertainty, we note that the number n_s of local filters falling within the stimulus window will be proportional to the width σ_x of the stimulus window and, similarly, the number of local filters n being pooled will be proportional to the width $\sigma_{\bar{x}} = \sqrt{\sigma_x^2 + \sigma_{\Delta x}^2}$ of the perceptual window. Thus, Equation 46 can be written as:

$$\log c_t = \alpha - \frac{1}{2} \log \sigma_x + \frac{1}{2} \log \left(2d' + \sqrt{4d'^2 + \beta \sqrt{\sigma_x^2 + \sigma_{\Delta x}^2}} \right). \quad (47)$$

where constant terms have been collected into the unknown α .

Using $d' = 1.349$ corresponding to unbiased 75% performance, we fit Equation 47 to the 1.7 c° summation data from Experiment 1, computing maximum likelihood estimates of the free parameters α , β , and $\sigma_{\Delta x}$, obtaining mean values for the spatial uncertainty $\sigma_{\Delta x}$ and proportional constant β of $\sigma_{\Delta x} = 1.2^\circ$ and $\beta = 65$.

Given these parameter estimates, we can address the hypothetical question: What would the summation slope for the 1.7 c° condition be if there were no spatial uncertainty? Setting $\sigma_{\Delta x}$ to zero and taking the derivative of Equation 47 with respect to σ_x , we obtain:

$$\frac{d \log c_t}{d \log n} = -\frac{1}{2} + \frac{\beta \sigma_x}{4 \sqrt{4d'^2 + \beta \sigma_x} (2d' + \sqrt{4d'^2 + \beta \sigma_x})}. \quad (48)$$

Using $d' = 1.349$ and our maximum likelihood estimate of $\beta = 65$, we find that over the range of stimulus widths we tested (one to eight cycles between $1/e$ points), the mean predicted summation slope for the 1.7 c° condition in the absence of spatial uncertainty is -0.31 , very similar to the mean empirical slope measured for the 0.5 c° condition (-0.29). This supports the hypothesis that spatial uncertainty is primarily responsible for the steeper empirical slopes observed for the 1.7 c° condition.

Discussion

Previous results on summation in noise

Our first two experiments suggest that stimulus noise raises thresholds for grating detection but does not alter the nature of summation: the decline in threshold as stimulus width increases follows a power law with an exponent that is roughly the same with or without stimulus noise.

These results are quite different from those of Kersten (1984), who found that for noisy gratings, contrast thresholds did not improve beyond about one cycle in width. For noiseless gratings,

on the other hand, Kersten found results very similar to ours: thresholds declined as a power law with an exponent on the order of -0.3 . The consistency in our results for the no-noise condition suggests that our divergent results have something to do with differences in the nature of the stimulus noise used.

We used 2D white noise that did not vary in time. Kersten (1984), on the other hand, used 1D white noise (varying randomly horizontally, but constant vertically) that also varied randomly over time. In other words, the noise appeared as random, time-varying vertical bars.

We see two ways in which these differences could contribute to the divergent results.

1. 2D versus 1D spatial noise. It is reasonable to assume that internal noise is 2D in space, i.e., varies randomly in both horizontal and vertical dimensions. In the low-noise condition, the observer will therefore benefit from spatial summation in both of these dimensions.

In the high-noise condition, where stimulus noise dominates, the benefit of summation may depend on the nature of the noise. For Kersten's stimuli, the observer only derives benefit from summation horizontally. In our experiments, observers benefit from summation in both dimensions. This does not predict a difference in performance for the ideal observer, since in both experiments the observer sees the full vertical extent of the grating. However, it is likely that for human observers, sensitivity to the grating stimuli is highest near the fovea, so that summation over a stimulus region of a fixed area is most effective if the region is roughly circular and centered at the fovea. Furthermore, the brain may be wired to more efficiently summate over roughly circular regions. Thus, even though the stimuli are windowed only in the horizontal dimension, the observer may effectively apply an internal vertical window roughly proportional to the horizontal stimulus window.

If this is the case, the extent of vertical summation will also increase as the horizontal window is enlarged, which benefits the observer when the noise is 2D as in our experiment, but not when the noise is 1D, as in Kersten's experiment. This predicts that summation will be more effective in the high-noise condition for our stimuli than for Kersten's.

2. Static versus dynamic noise. It is known that temporal sensitivity increases as a function of visual eccentricity (Allen and Hess, 1992). Therefore, it is possible that the temporally varying noise used by Kersten was more effective at masking the grating signal away from the fovea. This would predict a reduction in the benefit of summation in Kersten's high-noise condition.

Future experimental work could help to determine more precisely how these two factors determine the effects of stimulus noise on summation.

Previous methods for nonlinear systems identification

In this study, we have developed and applied a method for identifying the nature of local linear filters underlying detection of global stimuli, in the context of a nonlinear pooling network. In the neurosciences, techniques for nonlinear systems identification have perhaps been most widely explored for the problem of inferring nonlinear kernels mapping an input stimulus to the sequence of recorded spikes generated by a single neuron.

One approach is to use a second-order Wiener/Volterra expansion of the mapping (Marmarelis and Marmarelis, 1978; Victor, 1992), modeling the nonlinearity as quadratic. Neri (2004) explored an extension of this approach to behavioral data and demonstrated its application to a low-dimensional visual stimulus. Whereas this approach is general, it increases the degrees of

freedom of the model quadratically, making it impractical for larger visual stimuli. Neri (2004) relied on 12,500 trials for his low-dimensional example ($n = 13$). Scaling this up to the $n = 256 \times 256$ pixel images used here would require roughly 300 billion trials.

An alternative approach is spike-triggered covariance analysis (de Ruyter van Steveninck and Bialek, 1988; Victor, 1992; Bialek and de Ruyter van Steveninck, 2005; Schwartz et al., 2006), which uses the covariance of the spike-triggered stimulus ensemble to identify a low-dimensional linear subspace containing the nonlinear neural kernel. This approach reduces the number of free parameters from $O(n^2)$ to $O(kn)$, where k is the (hopefully small) dimensionality of the subspace, plus the parameters required to model the low-dimensional nonlinear kernel.

There remain issues in the practical application of the method to the psychophysical detection of global stimuli. First, although the method ultimately uses only $O(kn)$ parameters, it does require explicit computation of the $O(n^2)$ coefficients of the covariance matrix, which can be problematic for realistic stimulus images.

Second, STC analysis relies heavily on the assumption that the subspace containing the nonlinear mapping is of very low dimensionality. Is this the case for the problem we consider here? Our results suggest that the mapping can be modeled as a nonlinear pooling of identical but shifted local linear filters. This implies that the system is shift invariant, meaning that the eigenvectors of the covariance matrix must be sinusoids. It appears from our results that the underlying linear filters are fairly broadband (like V1 neurons), and so the dimensionality of the subspace is actually fairly high. Thus, the assumptions of the STC method are not satisfied.

Defeating the curse of dimensionality that plagues general higher-order methods for nonlinear systems analysis requires application of a priori constraints to reduce the effective dimensionality of the parameter space. Our power spectrum method achieves this. The key assumption is that the system is stationary (shift invariant) over retinal space. This means that the covariance matrix is determined by the autocorrelation, which has only n degrees of freedom, and is given by the inverse Fourier transform of the power spectral density. The shift invariance assumption is also important in that it allows the estimated kernels to be easily visualized in the Fourier domain and also (with an additional assumption about the spatial localization of the kernel) in the space domain.

Most closely related to our approach are methods developed in a recent study on crowding. Nandy and Tjan (2007) examined two-way identification of small, eccentrically presented letter stimuli with and without flankers, assuming an underlying model in which targets are identified by cross-correlation with a localized internal template. Because of spatial uncertainty, however, the observer monitors multiple spatially displaced copies of this internal template. The result is that first-order classification image analysis of error trials reveals a localized template for the missed letter but not for the false-alarmed letter.

Nandy and Tjan (2007) projected out the first-order missed component from the noise fields and then computed the autocorrelation over a limited rectangular region of interest (ROI) of the image to estimate the spatial features of the template for the false-alarmed letter. The optimal ROI is estimated using a maximum likelihood method.

As noted by Nandy and Tjan, (2007) accumulation of the autocorrelation of noise fields is equivalent to accumulation of the power spectral density, which is the method we use. Thus,

although applied in different ways to different problems, the two methods are closely related. In some ways, our problem is easier, as 50% of our trials contain only white noise. Restricting our analysis to signal-absent trials allows us to avoid the problem of projecting out the first-order classification image. Using a large-field stimulus also allows us to directly determine the ROI monitored by the observer, without using maximum likelihood techniques.

Whereas Nandy and Tjan (2007) do not explicitly model how multiple channels are combined, we focus specifically on the fact that assuming an energy model of pooling makes the system linear in the power spectrum domain, thus allowing for unbiased and efficient estimation of the classification image. Explicit testing of the energy model versus probability summation and other alternative models against the human data supports this assumption. Furthermore, our mathematical modeling of the summation slope predicted by an energy model, including the effects of spatial uncertainty, shows that the rate of decline of contrast thresholds as a function of stimulus width (slopes on the order of -0.3) is what one would expect from such an energy model.

Efficiency

The primate visual system has evolved to allow reliable visual detection even when object contrast is poor because of low-light conditions or camouflage. When the exact stimulus parameters (location, size, shape, etc.) are known and limiting noise can be approximated as additive, Gaussian and white, the most efficient detection strategy is matched filtering, which involves cross-correlation with a template matched to the known stimulus (Green and Swets, 1966). Under conditions of stimulus uncertainty, this simple strategy may no longer work. Probability summation (Green and Swets, 1966; Quick, 1974; Graham and Rogowitz, 1976; Graham, 1977; Sachs et al., 1980; Solomon, 2002) has long been a popular model for information pooling in the brain under such conditions. However, performance is poor compared with a system that pools responses quantitatively (Green and Swets, 1966).

The narrowband incoherent detector, based on a quadrature pair of global detectors, is the maximum likelihood grating detector when the phase is completely uncertain (Kay, 1998) and is a much more efficient alternative. However, this global detector still requires the brain to maintain large global templates with precise internal phase coherence. In contrast, existing physiological evidence suggests a flexible, compositional system based on nonlinear combinations of local detectors, yielding global detectors with a balance of selectivity and invariance (Cadiou et al., 2007; Connor et al., 2007). The local energy model is of this class, yielding superior detection ability compared with probability summation, without the global coherence requirements of the narrowband detector. Qualitatively similar local energy mechanisms seem to be common to a diversity of neural computations in auditory (Green and Swets, 1966) and visual (Adelson and Bergen, 1985; Morrone and Burr, 1988; Landy and Bergen, 1991; Emerson et al., 1992; Heitger et al., 1992; Manahilov and Simpson, 1999; Goris et al., 2009) cortex. Thus, the brain seems to have evolved a relatively efficient and general-purpose strategy for the neural pooling of visual information, even in the face of uncertainty.

References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299.
- Ahumada AJ Jr (2002) Classification image weights and internal noise level estimation. *J Vis* 2:121–131.
- Ahumada AJ, Beard B (1999) Classification images for detection. *Invest Ophthalmol Vis Sci* 40:S572.
- Ahumada AJ, Lovell J (1971) Stimulus features in signal detection. *J Acoust Soc Am* 49:1751–1756.
- Allen D, Hess RF (1992) Is the visual field temporally homogeneous. *Vision Res* 32:1075–1084.
- Banks MS, Sekuler AB, Anderson SJ (1991) Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling. *J Opt Soc Am A* 8:1775–1787.
- Beard B, Ahumada AJ (1999) Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *J Opt Soc Am A* 16:755–763.
- Bialek W, de Ruyter van Steveninck R (2005) Features and dimensions: motion estimation in fly vision. <http://arxiv.org/abs/q-bio/0505003>.
- Bracewell R (1978) *The Fourier transform and its applications*. New York: McGraw-Hill.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Cadiou C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. *J Neurophysiol* 98:1733–1750.
- Campbell FW, Robson JG (1968) Applications of Fourier analysis to the visibility of gratings. *J Physiol* 197:551–566.
- Christman S, Niebauer C (1997) The relation between left-right and upper-lower visual field asymmetries (or what goes up goes right, while what's left lays low). In: *Cerebral asymmetries in sensory and perceptual processing* (Christman S, ed), pp 263–296. Amsterdam: Elsevier Science.
- Connor CE, Brincat S, Pasupathy A (2007) Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* 17:140–147.
- Daugman J (1980) Two-dimensional analysis of cortical receptive field properties. *Vision Res* 20:846–856.
- de Ruyter van Steveninck R, Bialek W (1988) Coding and information transfer in short spike sequences. *Proc R Soc Lond B Biol Sci* 234:379–414.
- Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Res* 32:203–218.
- Goris R, Zaenen P, Wagemans J (2008) Some observations on contrast detection in noise. *J Vis* 8:1–15.
- Goris RLT, Wichmann FA, Henning GB (2009) A neurophysiologically plausible population code model for human contrast discrimination. *J Vis* 9:1–15.
- Graham N (1977) Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Res* 17:637–652.
- Graham N (2001) *Visual pattern analyzers*. Oxford: Oxford UP.
- Graham N, Rogowitz BE (1976) Spatial pooling properties deduced from the detectability of FM and quasi-AM gratings: a reanalysis. *Vision Res* 16:1021–1026.
- Green D, Swets S (1966) *Signal detection theory and psychophysics*. New York: Wiley.
- Heitger F, Rosenthaler L, von der Heydt R, Peterhans E, Külber O (1992) Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res* 32:963–981.
- Howell ER, Hess RF (1978) The functional area for summation to threshold for sinusoidal gratings. *Vision Res* 18:369–374.
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–243.
- Kay S (1998) *Fundamentals of statistical signal processing: detection theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Kersten D (1984) Spatial summation in visual noise. *Vision Res* 24:1977–1990.
- Landy MS, Bergen JR (1991) Texture segregation and orientation gradient. *Vision Res* 31:679–691.
- Legge GE, Foley JM (1980) Contrast masking in human vision. *J Opt Soc Am* 70:1458–1471.
- Manahilov V, Simpson W (1999) Energy model for contrast detection: spatiotemporal characteristics of threshold vision. *Biol Cybern* 81:61–71.
- Marmarelis P, Marmarelis V (1978) *Analysis of physiological systems: the white noise approach*. New York: Plenum.
- Mayer MJ, Tyler CW (1986) Invariance of the slope of the psychometric function with spatial summation. *J Opt Soc Am A* 3:1166–1172.
- Morrone MC, Burr DC (1988) Feature detection in human vision: a phase-dependent energy model. *Proc R Soc Lond B Biol Sci* 235:221–245.

- Mostafavi H, Sakrison DJ (1975) Structure and properties of a single channel in the human visual system. *Vision Res* 16:957–968.
- Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol* 283:53–77.
- Murray RF (2011) Classification images: a review. *J Vis* 11:1–25.
- Murray RF, Bennett PJ, Sekuler AB (2002) Optimal methods for calculating classification images: weighted sums. *J Vis* 2:79–104.
- Nandy A, Tjan B (2007) The nature of letter crowding as revealed by first- and second-order classification images. *J Vis* 7:1–26.
- Neri P (2004) Estimation of nonlinear psychophysical kernels. *J Vis* 4:82–91.
- Pelli DG, Farell B (1999) Why use noise? *J Opt Soc Am A* 16:647–653.
- Pelli DG, Zhang L (1991) Accurate control of contrast on microcomputer displays. *Vision Res* 31:1337–1350.
- Quick RF Jr (1974) A vector-magnitude model of contrast detection. *Kybernetik* 16:65–67.
- Rijsdijk JP, Kroon JN, van der Wildt GJ (1980) Contrast sensitivity as a function of position on the retina. *Vision Res* 20:235–241.
- Ringach DL (2002) Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol* 88:455–463.
- Ringach DL (2004) Mapping receptive fields in primary visual cortex. *J Physiol* 558:717–728.
- Sachs MB, Nachmias J, Robson JG (1980) Spatial-frequency channels in human vision. *J Opt Soc Am* 61:1176–1186.
- Schade OH Sr (1956) Optical and photoelectric analog of the eye. *J Opt Soc Am* 46:721–739.
- Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. *J Vis* 6:484–507.
- Smith AT, Singh KD, Williams AL, Greenlee MW (2001) Estimating receptive field size from fmri data in human striate and extrastriate visual cortex. *Cereb Cortex* 11:1182–1190.
- Solomon JA (2002) Noise reveals visual mechanisms of detection and discrimination. *J Vis* 2:105–120.
- Thomas NA, Elias LJ (2011) Upper and lower visual field differences in perceptual asymmetries. *Brain Res* 1387:108–115.
- Victor J (1992) Nonlinear systems analysis in vision: overview of kernel methods. In: *Nonlinear vision: determination of neural receptive fields, function and networks* (Pinter RB, Nabet B, eds), pp 1–37. New York: CRC.
- Watson AB, Pelli DG (1983) Quest: a Bayesian adaptive psychometric method. *Percept Psychophys* 33:113–120.
- Watson AB, Barlow HB, Robson JG (1983) What does the eye see best? *Nature* 302:419–422.
- Wichmann FA, Hill NJ (2001) The psychometric function: I. fitting, sampling, and goodness of fit. *Percept Psychophys* 63:1293–1313.
- Wilson HR, McFarlane DK, Phillips GC (1983) Spatial frequency tuning by orientation selective units estimated by oblique masking. *Vision Res* 23:873–882.