Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 795–813
Vision Research
www.elsevier.com/locate/visres

Psychophysical receptive fields of edge detection mechanisms

James H. Elder *, Adam J. Sachs

Centre for Vision Research, York University, 4700 Keele Street, North York, Toronto, Ont., Canada M3J 1P3

Received 5 November 2003; received in revised form 14 November 2003

Abstract

Theories of edge detection generally assume a front-end linear stage involving some population of neural filters. Here we study these early mechanisms using psychophysical techniques, and evaluate a number of models for edge detection. We measured psychophysical efficiency for detection of noisy luminance edge stimuli over a range of stimulus sizes and shapes. The data suggest a diversity in receptive field shape and orientation bandwidth, consistent with physiological evidence, but inconsistent with standard multi-channel models of visual processing.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Edge detection; Scalespace; Psychophysical efficiency; Primary visual cortex; Simple cells; Orientation bandwidth

1. Introduction

Forty years ago, Hubel and Wiesel discovered that the receptive fields of neurons in the primary visual cortex of cat and monkey are tuned to detect oriented structure in an image (Hubel & Wiesel, 1962, 1968). They suggested that this systematic orientation tuning had evolved to support the analysis of contours.

In these early papers, Hubel and Wiesel also noted the diversity in size and shape of these receptive fields. For example, they observed that receptive fields of neurons in the area centralis of cat varied in area from 0.25–16 deg^2 (Hubel & Wiesel, 1962). Similar observations have been made in monkey cortex (DeValois, Albrecht, & Thorell, 1982; Parker & Hawken, 1988).

Neurons in early visual cortex can be classified according to their sensitivity to the position or phase of a stimulus within their receptive fields (Hubel & Wiesel, 1962, 1968; Movshon, Thompson, & Tolhurst, 1978a, 1978b). Simple cells are phase-sensitive and can be approximated as half-wave rectifying linear filters (DeValois, Albrecht et al., 1982). As an alternative to receptive field size and shape, these neurons can be characterized by their spatial frequency tuning and orientation bandwidth. Roughly speaking, larger receptive fields are tuned to lower spatial frequencies, and more elongated receptive fields have narrower orientation tuning (for fixed spatial frequency bandwidth). Broad diversity in both peak spatial frequency tuning and orientation bandwidth has been observed in both V1 (DeValois, Albrecht et al., 1982; Devalois, Yund, & Hepler, 1982; Parker & Hawken, 1988; Schiller, Finlay, & Volman, 1976) and V2 (Levitt, Kiper, & Movshon, 1994) of macaque.

At approximately the same time that Hubel and Wiesel reported their first results in monkey, human psychophysical experiments (Campbell & Robson, 1968) led to the proposal of a multi-channel model of visual processing, in which the visual image is processed simultaneously by multiple, relatively narrowband spatial filters tuned to different spatial frequencies. Subsequent psychophysical work has led to the elaboration of several multi-channel or multi-scale models of visual detection and discrimination (Marr & Hildreth, 1980; Watson, 1982, 2000; Watt & Morgan, 1984; Wilson & Bergen, 1979; Wilson & Gelb, 1984). More recently, effective multi-scale computer vision algorithms for luminance edge detection have emerged (Elder & Zucker, 1998; Lindeberg, 1998).

These theories all address the neural tuning to spatial scale or frequency, but offer no functional explanation for the observed diversity in receptive field shape. In

* Corresponding author. Tel.: +1-416-736-2100x66475; fax: +1-416-736-5857.
E-mail address: jelder@yorku.ca (J.H. Elder).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2003.11.021

fact, these theories all assume filters of a constant shape and orientation bandwidth (for fixed spatial frequency bandwidth). This divergence from existing physiological evidence is not a minor one. For example, foveal simple cells in primary visual cortex range in orientation bandwidth from less than 10 deg to more than 180 deg, and receptive field height:width ratios range from roughly 1:1 to 16:1 (DeValois, Albrecht et al., 1982; Parker & Hawken, 1988). It seems likely that such a large variation serves some functional role, and should be detectable through psychophysical experimentation. Here we propose that one functional role of this diversity is to improve the efficiency of luminance edge detection in natural images.

Fig. 1 illustrates the idea. The reliability of luminance edge detection is limited by the noise generated by the microstructure of surfaces in the scene, by quantum fluctuations in the light impinging on the eye, and by noise introduced at various stages of neural processing. Signal detection theory (Peterson, Birdsall, & Fox, 1954) informs us that, to the degree that this noise can be approximated as spatially uncorrelated, the visual system must integrate information in the neighbourhood of the edge in order to raise the signal/noise ratio to a level sufficient to ensure reliable detection.

However, a second factor limiting edge detection is the proximity of other edges in the image. If spatial integration is allowed to extend too far from the edge, this local 'clutter' may bias the detection and generate an error. As Fig. 1 illustrates, these two factors compete to determine an optimal receptive field for a given edge (Elder & Zucker, 1998). The size and shape of this receptive field will vary depending upon contrast and noise levels, and the local image context. Our hypothesis is that the visual cortex provides a diverse population of neurons from which a receptive field of appropriate size and shape (as well as orientation and spatial frequency bandwidth) can be selected in order to reliably detect a given luminance edge.

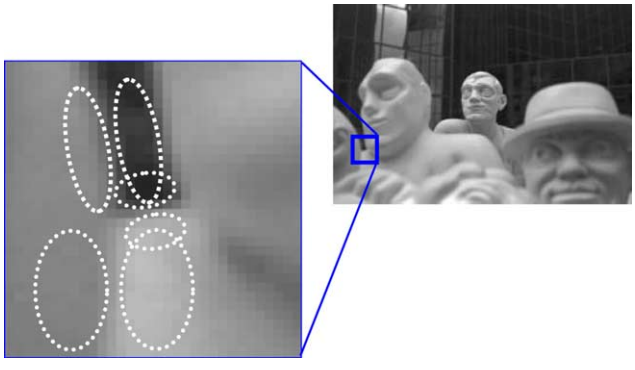

Fig. 1. The problem of luminance edge detection in natural images. The optimal receptive field size and shape for the detection of an edge depends on the local structure of the image around the edge.

While neurons in early visual cortex range widely in their spatial frequency bandwidth (DeValois, Albrecht et al., 1982; Parker & Hawken, 1988), the mechanisms most effective for edge detection are relatively broadband and odd-symmetric. Our hypothesis is that for this subset of broadband neurons, diversity in receptive field size and shape is in large part determined by the goal of reliable edge detection in the natural world. Conversely, we predict that limitations in psychophysical performance for detection of synthetic edge stimuli will be due in part to limits on this neural diversity, and thus psychophysics may allow us to estimate the parameters of the neural population.

Why limit our attention to edge stimuli? Is it plausible that this neural diversity could be optimized primarily for the detection of edges? What about other stimuli: gratings, Gabors, spots, etc? We argue that edges are different from these other psychophysical stimuli in their direct correspondence to important physical events in natural visual scenes, e.g., the boundaries of objects and shadows, changes in surface reflectance and attitude. Many psychophysical demonstrations suggest that edges play a principal role in human scene perception, and many computational vision algorithms rely upon detection of edges. Thus it is plausible that broadband mechanisms in early visual processing are determined, at least in part, by the goal of reliable edge detection, and it is of interest to understand human edge detection performance in detail.

To test our hypothesis, we conducted a series of experiments to estimate the human psychophysical efficiency of luminance edge detection. For simplicity, rather than trying to simulate the various types of clutter found in natural images, we limited the region around the edge available for integration by windowing the stimulus with a variety of elliptical apertures of various shapes and sizes. We will show that the observed data can be explained by a simple optimal filter selection model, in which filters represent the receptive fields of simple cells in early visual cortex, varying over a broad range in shape and orientation bandwidth.

These results:

1. Strengthen the original argument due to Hubel and Wiesel that the functional architecture of primary visual cortex is determined in part by the goal of contour analysis (Hubel & Wiesel, 1962, 1968).
2. Suggest a specific functional role for the diversity in receptive field shape and orientation bandwidth of neurons in early visual cortex.
3. Suggest that the efficiency of edge detection can be completely accounted for by neural mechanisms in early visual cortex (V1, and possibly V2). This implies that visual processing in other visual areas is directed toward the detection and representation of more complex stimuli.

2. Prior psychophysical studies of detection

Most prior studies of luminance contrast detection have employed sinusoidal gratings, sometimes windowed with circular or elliptical Gaussian envelopes (Campbell & Robson, 1968; Marr & Hildreth, 1980; Watson, 1982, 2000; Watt & Morgan, 1984; Wilson & Bergen, 1979; Wilson & Gelb, 1984). Were the visual system perfectly linear, response to an arbitrary stimulus could in theory be inferred from responses to sinusoidal components. However detection by definition involves some form of nonlinearity. Beyond the essential decision nonlinearity (i.e., transduction of the percept into a binary response), we must consider nonlinear selection and/or pooling of responses from multiple mechanisms. Thus edge detection performance cannot generally be inferred from grating response data.

Most prior work on detection addresses issues of spatial frequency tuning and bandwidth (Campbell & Robson, 1968; Graham, 1972; Graham & Nachmias, 1971; Graham, Robson, & Nachmias, 1978; Kersten, 1984; Legge & Foley, 1980; Stromeyer & Klein, 1975; Watson, 1982; Wilson & Bergen, 1979; Wilson & McFarlane, 1983), whereas our interest is in the two-dimensional size and shape of the mechanisms underlying edge detection, i.e., in their length and width tuning. Graham (1989) has observed that data on the length tuning of detection mechanisms is 'scanty'. The few detection studies that have systematically covaried stimulus length and width employed Gabor stimuli (Polat & Tyler, 1999; Watson, Barlow, & Robson, 1983).

Most studies of contrast detection do not employ stimulus noise or masks, hence contrast thresholds are relatively low (Campbell & Robson, 1968; Marr & Hildreth, 1980; Watson, 1982, 2000; Watt & Morgan, 1984; Wilson & Bergen, 1979; Wilson & Gelb, 1984). Under these conditions, there is substantial evidence for nonlinear pooling of responses (e.g., probability summation) from localized linear mechanisms over a number of stimulus parameters, including retinal space (Chen & Tyler, 1999; Graham, 1977, 1989; Legge & Foley, 1980; Stromeyer & Klein, 1975). This greatly complicates the task of characterizing the local mechanisms underlying detection. However, when stimulus noise or masking is employed, contrast thresholds rise, and there is evidence that pooling is effectively shut down (Kersten, 1984; Legge & Foley, 1980). Thus stimulus noise appears to be a useful tool for isolating the nature of the local detection filters from more global pooling mechanisms.

Since the goal of the present study is to determine the nature of the local linear mechanisms that underlie edge detection, we measure detection thresholds under high-noise (high-contrast) conditions. The use of noise also allows us to calculate ideal observer performance as a benchmark for human performance, thus isolating the tuning of the human visual system from variations in the inherent difficulty of the task (Pelli & Farell, 1999).

The precise conditions for global pooling may depend upon task and stimulus details. Verghese and Stone (1996) found configuration-specific pooling effects for supra-threshold speed discrimination. Bonneh and Sagi (1999) found evidence for probability summation in contrast discrimination of Gabor patches on supra-threshold Gabor masks. Since we cannot be guaranteed that pooling mechanisms are not acting in our experiments, we will explicitly test whether a nonlinear pooling model can account for our data.

In summary, most prior psychophysical work on detection has focused on low-contrast sinusoidal stimuli and one-dimensional spatial frequency tuning properties of detection mechanisms. To characterize the size and shape of the local mechanisms underlying edge detection, we need data on detection performance over a range of edge stimuli varying in length and width. Prior work suggests the use of high levels of stimulus noise, in order to minimize the effects of nonlinear pooling and to permit the calculation of psychophysical efficiency. These considerations governed the design of the stimuli used in the experiments we now describe.

3. Methods

Stimuli were generated on a Macintosh G3 computer using MATLAB and the Video (Pelli, 1997) and Psychophysics (Brainard, 1997) Toolboxes. Stimuli were displayed on a 21″ Sony Trinitron® Display at a resolution of 1024×768 8-bit pixels and a refresh rate of 85 Hz. The background and mean luminance of all stimuli was 57 ± 0.5 cd/m^2.

Three adult human male subjects with corrected-to-normal vision were used, including one of the authors (AS). Head position was fixed at roughly 80 cm from the display using a chin rest and forehead strap. At this distance, a foveal pixel subtended 1.68 arcmin. Experiments were performed in a dark room. Stimuli were viewed binocularly.

Each stimulus consisted of a vertical luminance step edge windowed by a synthetic circular or elliptical aperture. The polarity of the edge (which side of the edge was lighter) was determined randomly with equal probability. White Gaussian noise with zero mean and standard deviation of 0.16 Michelson contrast units was added to each pixel of the display (Fig. 2). Prior to each stimulus presentation, subjects viewed an image consisting of a low contrast (0.08 Michelson units) fixation dot and four corners indicating the location, size and

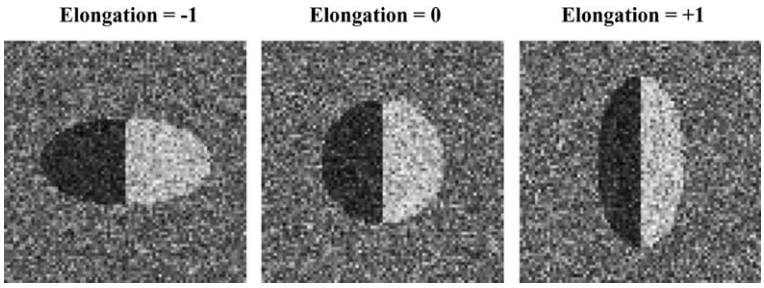

Fig. 2. Example stimuli. Stimulus contrast is well above threshold in these examples.

shape of the stimulus. The subject's task was to correctly indicate the polarity of the edge. [1]

Each experiment consisted of between 8 and 9 conditions, and each condition consisted of between 4 and 8 blocks (4 for Experiment 1, 5 for Experiment 2 and 8 for Experiment 3). In each block all stimulus parameters were fixed except for the contrast of the luminance edge, which was varied from trial to trial using an adaptive psychometric testing procedure (QUEST, Watson & Pelli, 1983) to estimate the contrast threshold for 82%-correct performance. Each block was terminated when either the estimated uncertainty in the threshold estimate fell below 0.05 log units, or 100 trials had been completed. In the latter case the block was discarded: this occurred in about 4% of the blocks. There were generally between fifty and ninety trials in each block. Subjects prepared for each block by performing a test block of at least ten trials at above-threshold contrasts.

Each trial consisted of a sequence of three images. First the fixation stimulus was presented for 500 ms. This was followed by a 153 ms stimulus interval. The stimulus was then replaced by a uniform grey screen at background luminance, until the subject responded by clicking a mouse once to indicate a light/dark edge or twice to indicate a dark/light edge. Subjects were given auditory feedback on incorrect trials. After 700 ms, the next trial began.

4. Results

We conducted three experiments to measure human psychophysical efficiency for edge detection. In the first experiment, the aperture was circular, with diameter ranging from 0.34 deg to 19.4 deg. For all three subjects, Michelson contrast threshold was found to decline with increasing aperture size (Fig. 3a).

In the second and third experiments, the aperture was elliptical, with a fixed area and a range of elongations. We define elongation as $\log_2$ (height/width), so that positive elongation refers to vertical elongation along

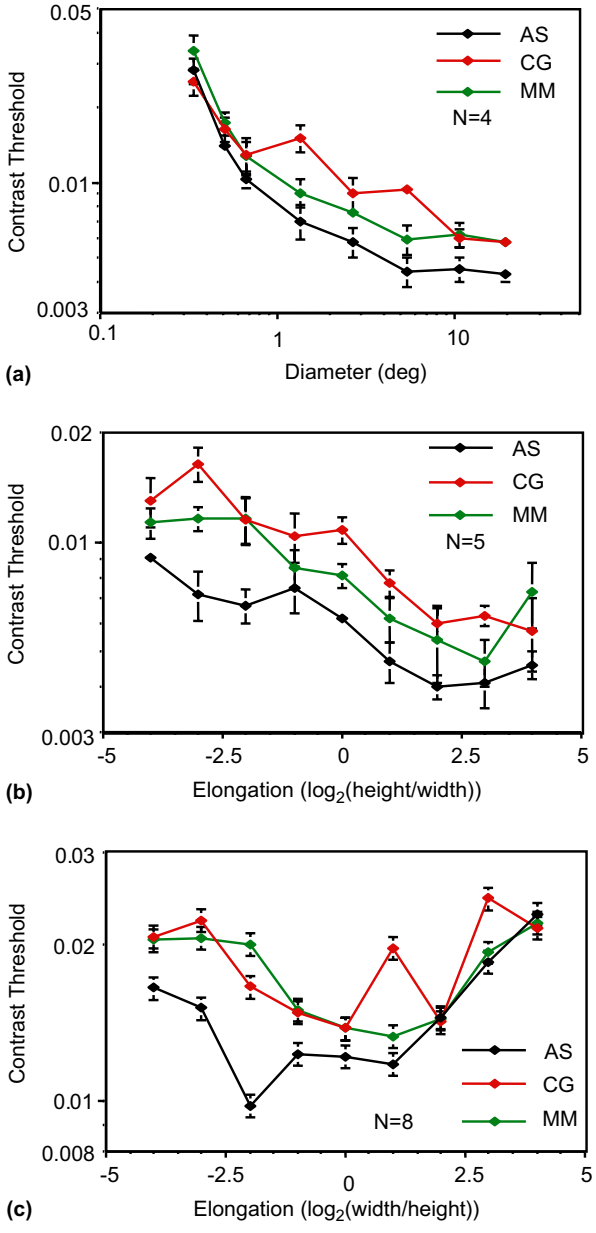

Fig. 3. Psychophysical contrast detection thresholds for (a) Experiment 1, (b) Experiment 2 and (c) Experiment 3.

[1] Control experiments indicate that performance on this discrimination task is similar to performance on a two-interval forced-choice detection task. The polarity task is simpler in that it requires only a single stimulus interval per trial.

the edge and negative elongation refers to horizontal elongation across the edge. In Experiment 2, the aperture area was fixed at 5.67 deg^2. For all three subjects, contrast threshold decreased as the aperture elongation was increased from −4 to 4, i.e., as the window was stretched along the edge (Fig. 3b). In Experiment 3, the aperture area was fixed at 0.35 deg^2. Contrast thresholds were lowest for moderate elongations (quasi-circular apertures), increasing for more extreme elongations either along the edge or across the edge (Fig. 3c).

It can be difficult to infer the nature of neural mechanisms directly from contrast thresholds, since variations in contrast threshold may be due in part to variations in the inherent difficulty of the task. In our first experiment, we found that contrast thresholds were lower for larger stimuli. Does this mean that optimal visual mechanisms have evolved for processing large stimuli? Not necessarily. Since the total contrast energy increases as a function of stimulus size, these larger stimuli are inherently easier to detect. We can eliminate this confounding factor by using ideal observer performance as a benchmark for human performance, i.e., by computing efficiency relative to the ideal observer.

The ideal observer (Barlow, 1962; Rose, 1948) discriminates the edge stimuli by selecting the alternative with the maximum a posteriori probability. For our experiments, the ideal strategy is to make a decision based on a pair of linear filters $g_1(x,y)$, $g_2(x,y)$ that are identical to the two alternative edge stimuli without noise, up to a scaling constant (i.e. the matched filters, Peterson et al., 1954). A decision is based upon which of these two filters has the greater response. This is mathematically equivalent to a decision based upon the sign of the response of a filter $g(x,y)$ formed from the difference of the two optimal filters: $g(x,y) = \frac{1}{2}(g_1(x,y) - g_2(x,y))$. Since the two alternative edge stimuli are identical in contrast but for a sign reversal, these two matched filters are also a contrast-reversed pair, hence $g(x,y) = g_1(x,y)$. For convenience, we assume that the matched filter has unit energy (L_2 norm). Since the stimulus noise is additive, white and Gaussian, the response of the filter will be normally distributed, and the contrast C_i required to

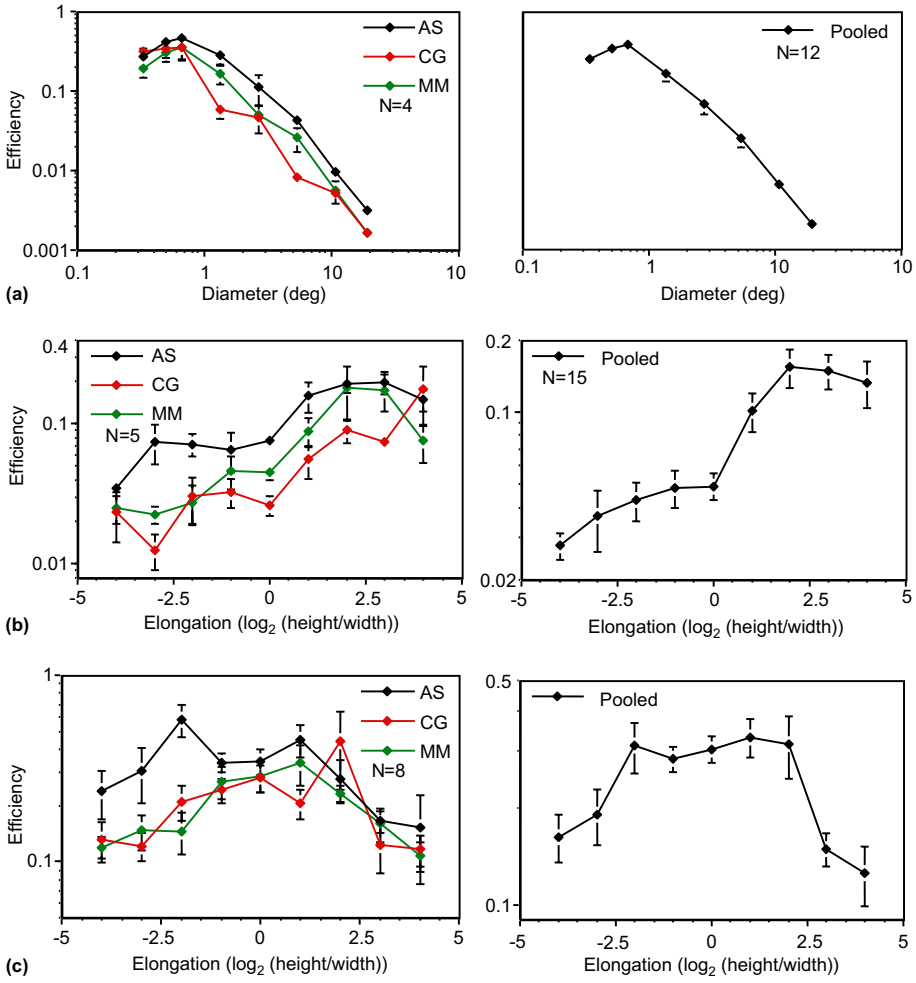

Fig. 4. Psychophysical detection efficiencies for (a) Experiment 1, (b) Experiment 2 and (c) Experiment 3. Plots on the left show results of individual subjects, plots on the right show results pooled over all three subjects.

achieve 82%-correct performance can be determined using a standard normal table. Specifically,

$$z_{0.82} = \frac{C_i r_i}{\sigma_n} \Rightarrow C_i = \frac{z_{0.82}\sigma_n}{r_i} = \frac{0.9154\sigma_n}{r_i}$$

where $z_{0.82}$ is the signal-to-noise ratio required for 82% correct performance; r_i is the (0-noise) response of the matched filter $g_1(x,y)$ to its associated unit-contrast stimulus; σ_n is the standard deviation of the stimulus noise. Human efficiency η_h is then defined as the squared ratio of ideal to human contrast threshold (Pelli, 1990): $\eta_h = (C_i/C_h)^2$.

Fig. 4a shows psychophysical efficiency for circular apertures (Experiment 1). Pooling the data for the three subjects, efficiency was found to peak for viewing apertures roughly 0.7 deg in diameter, declining for both smaller apertures (e.g., 0.34 deg, $p < 0.05$) and larger apertures (e.g., 1.34 deg, $p < 0.05$). Fig. 4b and c show efficiency for elliptical apertures (Experiments 2 and 3). For larger stimuli (Experiment 2), efficiency was found to be higher for positive elongations along the edge than for negative elongations across the edge ($p < 0.05$). For smaller stimuli (Experiment 3), efficiency was found to be higher for moderate elongations than for extreme elongations (elongations of −2 to 2 vs. elongations of ±3 to 4, $p < 0.05$).

Note that by considering the results in terms of efficiency, we have a more direct indication of the relationship between the stimulus and the underlying mechanisms. For example, in Fig. 4a, we find that efficiency peaks for edge stimuli with a diameter of 0.67 deg. If detection filters existed over a wide range of scales, as suggested by many theories, we would expect there to be a broad plateau in efficiency. The existence of this well-defined peak suggests that these detection filters span only a relatively small range of scales.

5. Modelling

Can these results be understood in terms of neural information processing in early visual cortex? To answer this question, we require a theory for how the brain maps responses from a population of visual neurons to a unitary decision ('The edge is light/dark.' or 'The edge is dark/light.'). Here we propose a simple theory of optimal selection. We assume that the visual system bases its decision on the response of a population of visual neurons whose receptive fields roughly match the stimulus in preferred orientation and spatial frequency bandwidth. These early neuronal receptive fields may be modeled as half-wave rectified linear spatial filters. [2] Each filter $f_i(x,y)$ may correspond to a single neuron, or to a small assembly of neurons with effectively identical receptive field properties. The output of each filter is normalized by its energy (L_2 norm), so that each has equivalent noise-sensitivity. For a given stimulus, the discrimination is based on the pair of filters $(f_1(x,y), f_2(x,y))$ most closely-matched (in the least-squares sense) to the two alternative edge stimuli without noise. A decision is based upon which of these two filters has the greater response. As for the matched filter case, this is mathematically equivalent to a decision based upon the sign of the response of a filter $f(x,y)$ formed from the normalized difference $\frac{1}{2}(f_1(x,y) - f_2(x,y))$ of the two optimal filters. Since the two alternative edge stimuli are identical in contrast but for a sign reversal, these two optimal filters are also a contrast-reversed pair, hence $f(x,y) = f_1(x,y)$. As for the matched filter, the response of this filter will be normally distributed, and the contrast C_f required to achieve 82% performance can be determined using a standard normal table. Specifically,

$$z_{0.82} = \frac{C_f r_f}{\sigma_n} \Rightarrow C_f = \frac{z_{0.82}\sigma_n}{r_f} = \frac{0.9154\sigma_n}{r_f}$$

where $z_{0.82}$ is the signal-to-noise ratio required for 82% correct performance; r_f is the response of the filter $f_1(x,y)$ to its associated unit-contrast stimulus; σ_n is the standard deviation of the stimulus noise.

The efficiency η_f of this filter mechanism for detecting the stimulus is thus

$$\eta_f = \left(\frac{C_i}{C_f}\right)^2 = \left(\frac{r_f}{r_i}\right)^2$$

Within this simple framework, we can test whether neural processing in early visual cortex can account for human edge detection efficiency. We will restrict our attention to the population of simple cells found in early visual cortex (Baizer, Robinson, & Dow, 1977; Hubel & Wiesel, 1968; Levitt et al., 1994), and ask whether edge detection can be accounted for at this relatively primitive stage of cortical processing.

While the receptive field tuning of simple cells in primate visual areas V1 and V2 ranges over all orientations and across a broad range of spatial-frequency phases and bandwidths, our stimuli are at fixed vertical orientation, odd phase and relatively broad horizontal spatial frequency bandwidth. The filters with strongest response will be as closely matched to the stimulus as possible, in other words, vertically-oriented, odd-phase, and broadband. We can model this subpopulation as oriented first-derivative of Gaussian filters (Koenderink, 1984; Young, 1991) (Fig. 5):

$$f(x,y) = -\frac{2x}{\pi\sigma_x^3\sigma_y} e^{-[(x/\sigma_x)^2 + (y/\sigma_y)^2]} \quad (1)$$

where σ_x and σ_y are the Gaussian scale constants across and along the neuron's preferred orientation.

[2] Half-wave rectification reflects the fact that visual cortical neurons cannot directly signal negative responses.

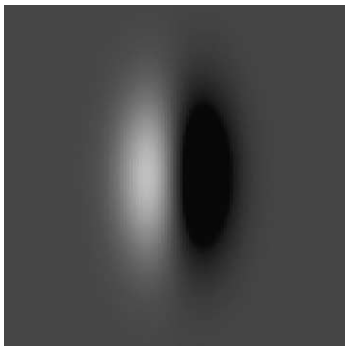

Fig. 5. Example receptive field using oriented first derivative of Gaussian model.

Based upon this family of linear mechanisms, we consider five possible models for edge detection, in increasing order of complexity.

5.1. Model 1. Single channel model

In our most primitive model, we assume that the observer uses the same filter pair, centred on the stimulus, to discriminate each of the edge stimuli in all conditions. To evaluate the model, we treat the scale constants σ_x and σ_y as free parameters, optimized to fit the data. To model additional stimulus-independent sources of inefficiency, we employ a single factor γ, which represents the intrinsic efficiency of the filters:

$$\eta_f = \gamma\tilde{\eta}_f$$

where $\eta_f =$ actual efficiency of the neuron, and $\tilde{\eta}_f =$ ideal efficiency of the neuron.

The intrinsic filter efficiency has a global effect over all stimulus conditions: efficiency of the model rises as γ is increased, and falls as γ is decreased. Thus the single channel model has three free parameters to fit 26 data points from our three experiments. We employed a standard Nelder–Mead simplex optimization algorithm (MATLAB function FMINSEARCH) to compute maximum likelihood estimates (MLEs) for model parameters that minimize the squared deviation of the model from the human data. The optimization was repeated hundreds of times from random initial conditions to confirm results. [3]

Table 1 shows the maximum likelihood estimate (MLE) values for these parameters. Fig. 6 (top row) shows the fit of the model to the data. While the model accounts for the results of Experiment 1 reasonably well, it fails to capture the increase in efficiency for positive elongations observed for our human observers in Experiment 2. The model also fails for Experiment 3, predicting a far steeper falloff in efficiency with elongation than is seen in the human data. Clearly an optimal filter selection model of edge detection based upon a single linear channel is inadequate.

5.2. Model 2. Multi-scale model

The standard multi-channel model of visual processing (Marr & Hildreth, 1980; Watson, 1982, 2000; Watt & Morgan, 1984; Wilson & Bergen, 1979; Wilson & Gelb, 1984) assumes a population of neural filters of fixed shape and orientation tuning, but varying in scale over a range of 6:1 (Wilson & Bergen, 1979), 9:1 (Watt & Morgan, 1984), 20:1 (Wilson & Gelb, 1984) or 32:1 (Watson, 2000).

Could such a population of filters, of diverse scales but identical shape and orientation tuning, account for human edge detection efficiency? To answer this question, we assumed a filter selection model, based on a family of linear filters of fixed shape but varying over some range in scale. This requires three parameters in addition to the intrinsic filter efficiency γ:

$$\alpha_x < \sigma_x < \beta_x$$

$$\sigma_y/\sigma_x = \varepsilon$$

where ε represents the shape (elongation) of the filters. For each condition, detection was assumed to be determined by the filter most closely matching the stimulus (optimal filter selection). Again, we used Nelder–Mead simplex optimization to compute the MLE values for these parameters, given the human data (Table 1). Fig. 7 shows the range of filters in the optimal model, on a two-dimensional scalespace plot. The red line indicates the range of filters available in the population, and the 'x' point markers indicate the actual scales selected for the stimuli we used. The optimal elongation parameter, represented by the slope of this line, was found to be $\varepsilon = 1.8$. The fit of this model to the data is shown in Fig. 6 (second row). Surprisingly, the multi-scale model fits the data only marginally better than the single channel model.

Although a truly scale-invariant model would predict identical performance for Experiments 2 and 3, the limits on the range of scales available in the optimized multi-scale model result in different predictions for the two experiments. To illustrate this, Fig. 8 shows the scales selected for each condition. Since the stimuli in Experiment 2 are relatively large, the maximum filter scale within the population (β_x) is selected for processing all of these stimuli. Performance falls off for extreme elongations, when the filter exceeds either the width or height of the stimulus.

Since the stimuli in Experiment 3 are closer in scale to the filter population available, model performance is qualitatively different. Strongest filter responses are obtained when the filter roughly matches the size of the

[3] We employ the same method to compute MLE parameters for Models 2–4.

Table 1
MLE model parameters

Single-channel		Multi-scale		2D scalespace		Hyperbolic scalespace		Nonlinear pooling	
Param.	Value	Param.	Value	Param.	Value	Param.	Value	Param.	Value
γ	0.69	γ	0.57	γ	0.37	γ	0.37	γ	0.62
σ_x	9.9	α_x	4.5	α_x	8.0	α_x	7.4	σ_x	7.7
σ_y	24.4	β_x	12.6	β_x	11.1	β_x	16.9	σ_y	7.0
		ε	1.8	α_y	–	a_1	−0.034		
				β_y	59.8	a_2	−0.200		
						a_3	0.029		

Scale constants measured in arcmin.

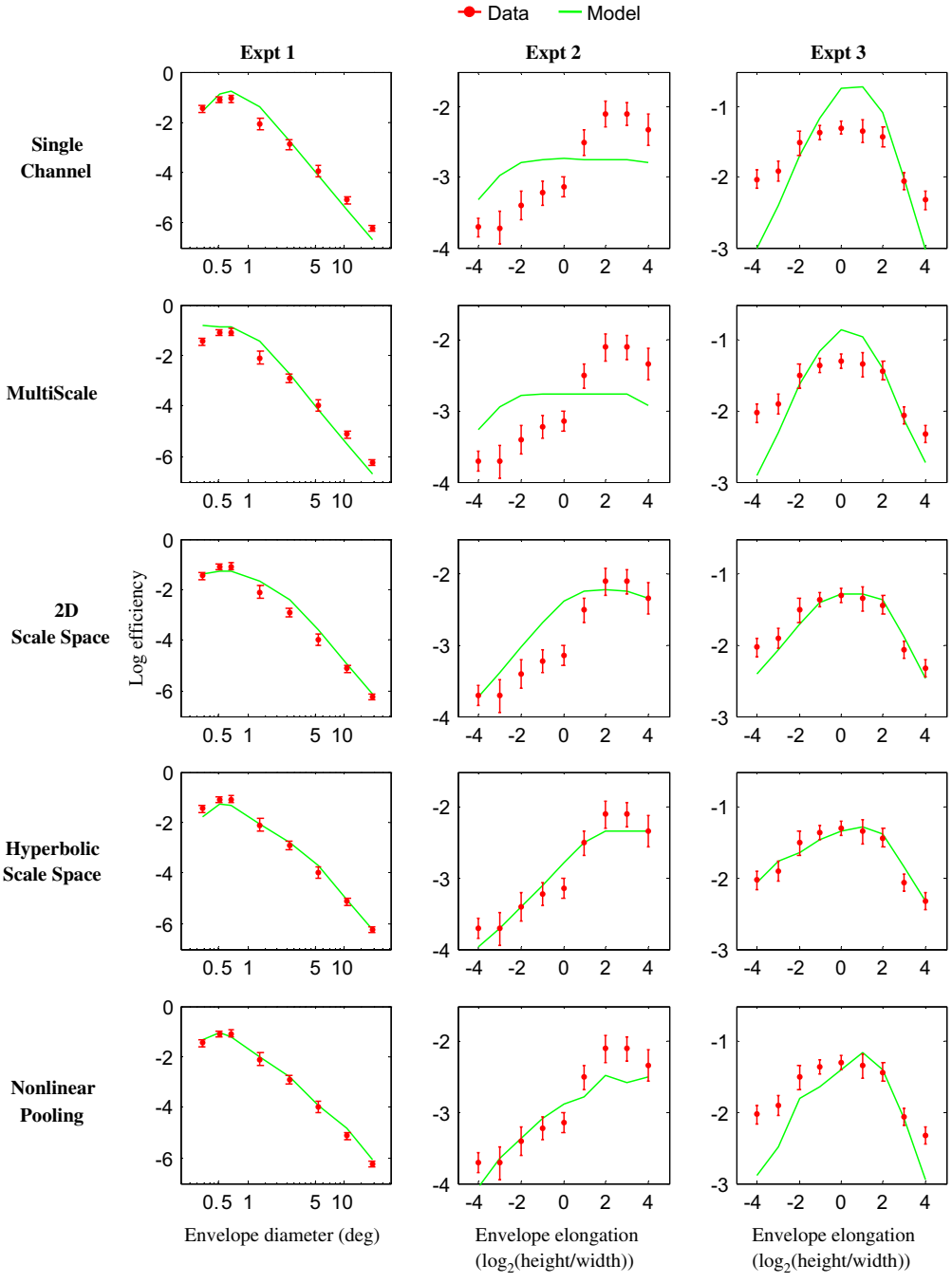

Fig. 6. Model fits to human edge detection efficiency data.

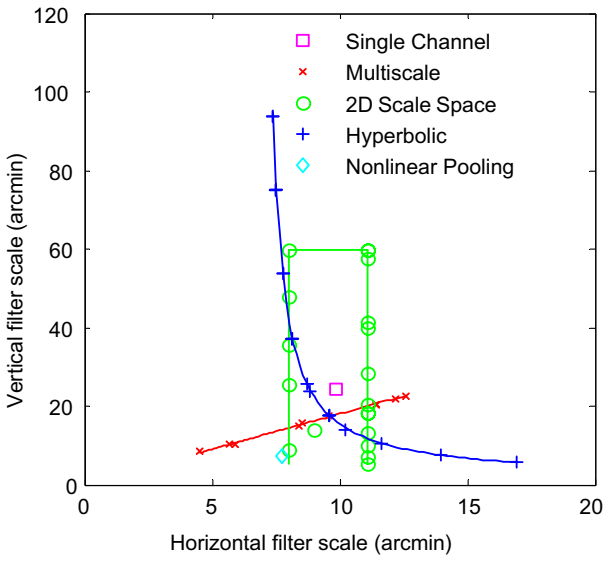

Fig. 7. Filter scales employed by optimized edge detection models. Curves indicate range of scales spanned by filter population, point markers indicate scales automatically selected for stimuli tested.

stimulus in the horizontal direction (across the edge). For negative elongations, the stimulus is wider than the available filters, and filter scale again reaches its maximum value (β_x). As elongation is increased toward the positive, the width of the stimulus comes into the range of the filters available, and filter scale tracks this width down to near its minimum.

In summary, a filter selection theory based on filters ranging in scale but fixed in shape is inconsistent with the human data for edge detection.

5.3. Model 3. Two-dimensional scalespace model

The standard multi-channel model provides for diversity in filter scale, but not filter shape, in contrast with the physiological record (DeValois, Albrecht et al., 1982; Parker & Hawken, 1988). In order to reflect a broader neural diversity, we modified the multi-scale filter selection model to permit independent variation in both horizontal and vertical scale parameters, within a two-dimensional region of scalespace (Fig. 7):

$$\alpha_x < \sigma_x < \beta_x$$

$$\alpha_y < \sigma_y < \beta_y$$

The model has 5 free parameters: these four scalespace bounds and the intrinsic filter efficiency γ. The MLE parameters are shown in Table 1, and the region of scalespace occupied by the population of filters is shown in Fig. 7. The lower bound α_y on the length scale constant was not exploited by the model, presumably because the stimulus set did not include stimuli short enough to demand filters below the minimum filter length available in the neural population. Thus only 4 free parameters were actually used to fit the data. Observe that while selected filters vary over a broad range in length tuning along the preferred orientation of the filter (5.1–59.8 arcmin), there is far less variation in width tuning orthogonal to the preferred orientation (8.0–11.1 arcmin).

Fig. 6 (middle row) shows that this 2D scalespace model accounts much better for the data than either single channel or multi-scale models. The model captures the decline in efficiency for both small and large stimuli (Experiment 1), as well as the higher efficiencies observed for positive elongations relative to negative elongations, seen in Experiment 2. Finally, it more accurately captures the moderate rolloff in efficiency with elongation observed in Experiment 3.

Why is this particular region of scalespace found to be optimal in the context of this model? We conducted a perturbation analysis to gain insight into the effects of the parameters on model performance. The results correspond well with intuition:

- Decreasing α_x leads to an increase in efficiency for narrow stimuli (Experiment 1, Condition 1; Experiment 3, Conditions 6–9). Conversely, increasing α_x leads to a decrease in efficiency for these stimuli.
- Decreasing β_x leads to a decrease in efficiency for wide stimuli (Experiment 1, Conditions 3–8; Experiment 2, all conditions; Experiment 3, Conditions 1–5). Conversely, increasing β_x leads to an increase in efficiency for these stimuli.
- Decreasing α_y has no effect on the model. Increasing α_y beyond 5.1 arcmin leads to a decrease in efficiency for the shortest stimulus (Experiment 3, Condition 1).
- Decreasing β_y leads to a decrease in efficiency for tall stimuli (Experiment 1, Conditions 5–8; Experiment 2, Conditions 5–9; Experiment 3, Condition 9). Conversely, increasing β_y leads to an increase in efficiency for these stimuli.

While the 2D scalespace model is much more consistent with the human data than single channel or multi-scale models, it overestimates efficiency for intermediate-sized stimuli in Experiment 1, and for a range of elongations (−3 to +1) in Experiment 2. For these conditions, the model selected the 5 scales indicated by circles along the upper portion of the right boundary of the optimized scalespace region in Fig. 7 (i.e., $\sigma_x = 11$ arcmin and $\sigma_y = 28$, 40, 41, 58 and 60 arcmin). The fact that the model outperforms the human data for these conditions suggests that these filters that are relatively large in both width and length are not in the population of neural filters available to the human visual system. This suggests some joint constraint on the two filters scales, and leads us to consider our fourth model.

5.4. Model 4. Hyperbolic scalespace model

Our analysis of the 2D scalespace model suggests a limit on the joint length and width tuning of the detection

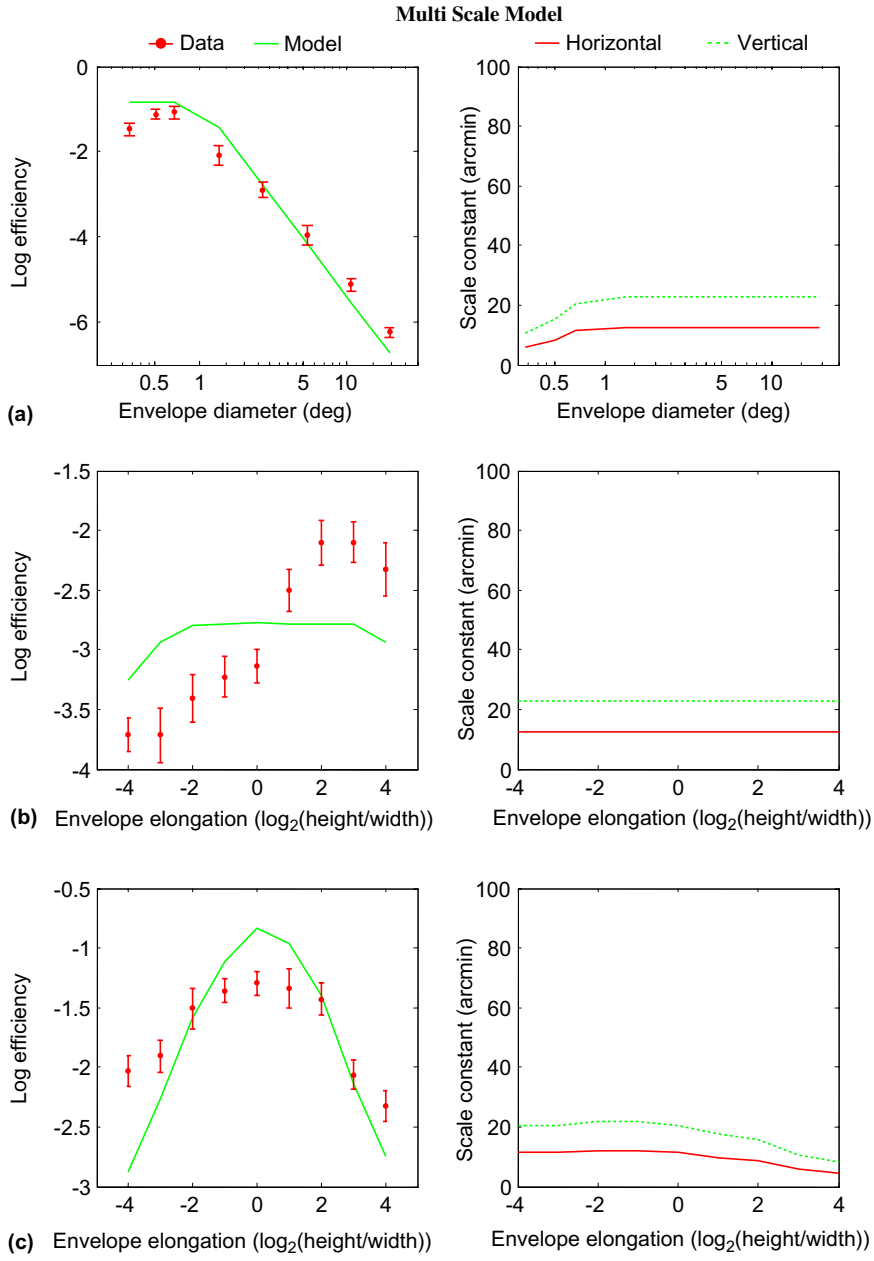

Fig. 8. Predictions of multi-scale filter selection model. Left column: efficiency predictions. Right column: selected scales. (a–c) correspond to Experiments 1–3 respectively.

filters in the neural population. A natural way to couple the length and width scale parameters is through a quadric model, in this case, a hyperbola:

$$a_1\sigma_x + a_2\sigma_y + a_3\sigma_x\sigma_y = 1$$

$$\alpha_x < \sigma_x < \beta_x$$

The second-order term in $\sigma_x\sigma_y$ constrains the hyperbolic model from selecting filters that are large in both dimensions, forcing the model to choose smaller filters, more poorly matched to the stimuli. This model has 6 free parameters, including the intrinsic filter efficiency γ.

The MLE parameters are shown in Table 1, and the corresponding scalespace curve occupied by the population of filters is shown in Fig. 7. Fig. 9 shows the predicted adaptation of filter scales. Note how an increase in vertical scale necessarily entails a decrease in horizontal scale.

Fig. 6 (second from bottom) shows that the hyperbolic scalespace model accounts much better for the data than any of the other models. Overestimation of efficiency in Experiments 1 and 2 is greatly reduced, and the model also more accurately reflects human performance for the smaller, negatively elongated stimuli (Experiment 3).

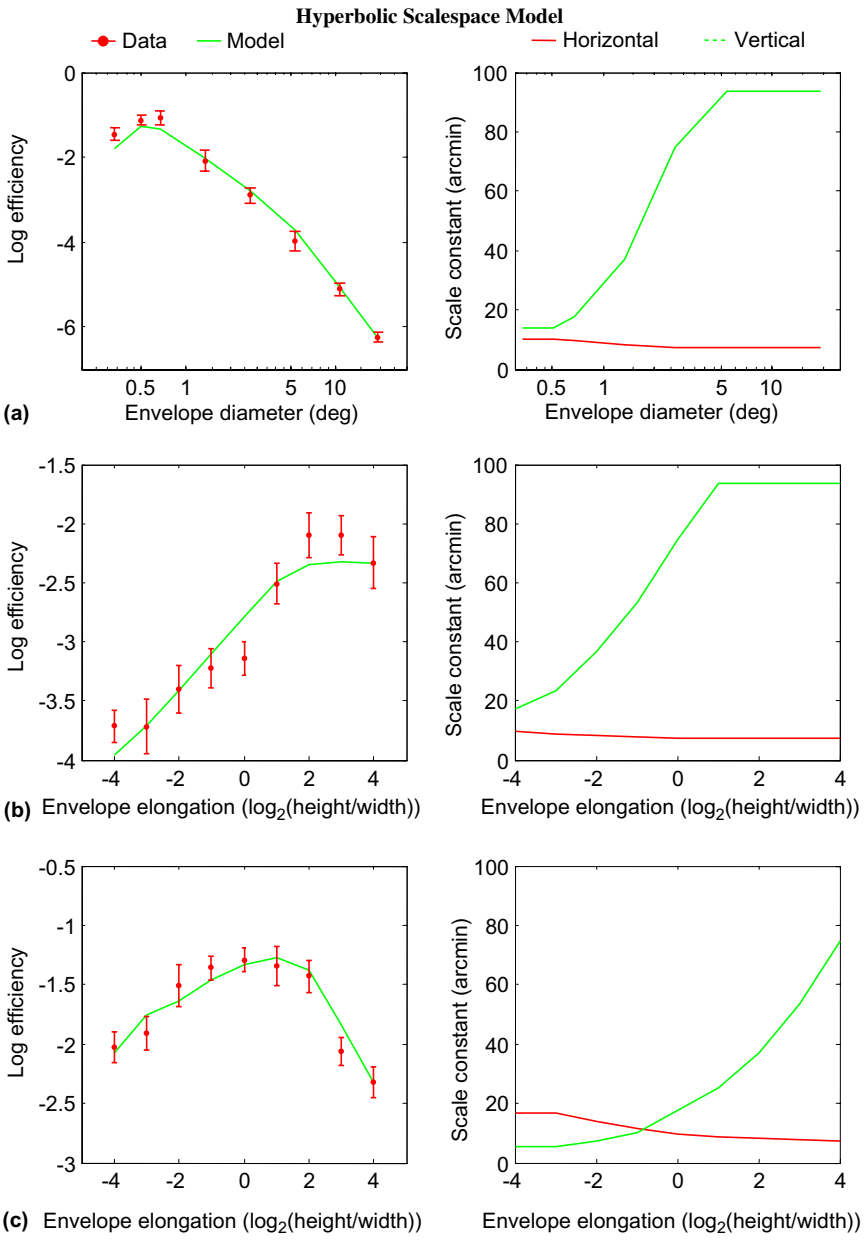

Fig. 9. Predictions of hyperbolic scalespace filter selection model. Left column: efficiency predictions. Right column: selected scales. (a)–(c) correspond to Experiments 1–3 respectively.

A perturbation analysis reveals that:

- Decreasing any of the hyperbola parameters (a_1, a_2, a_3) shifts the curve for Experiment 1 to the right, shifts the curve for Experiment 2 up, and narrows the tuning of the curve for Experiment 3. Increasing these parameters has the opposite effect.
- Decreasing α_x leads to an increase in efficiency for tall stimuli (Experiment 1, Conditions 6–8; Experiment 2, Conditions 6–9). Conversely, increasing α_x leads to a decrease in efficiency for these stimuli.
- Decreasing β_x leads to a decrease in efficiency for short, wide stimuli (Experiment 3, Conditions 1–2. Conversely, increasing β_x leads to an increase in efficiency for these stimuli.

5.5. Model 5. Nonlinear pooling model

In all four models we have considered thus far, detection was based upon selection from one or more filters centred on the stimulus (at fixation). Clearly there are many other neural filters at and around the fovea, whose receptive fields overlap the stimulus to a significant

extent. Under low-contrast (no noise) conditions, there is substantial evidence for nonlinear pooling of responses from such localized linear mechanisms over a number of stimulus parameters, including retinal space (Chen & Tyler, 1999; Graham, 1977, 1989; Legge & Foley, 1980; Stromeyer & Klein, 1975). Nonlinear pooling has often been modeled as probability summation in a high threshold model of signal detection (Graham, 1977), in which it is assumed that a signal is detected if the noisy response of one or more mechanisms exceeds a criterion threshold. Quick (1974) suggested a particular asymmetric noise distribution which leads to a simple form for the probability of detection over N mechanisms: $p(\text{detection}) = 1 - 2^{-\sum_{i=1}^{N} R_i^k}$, where R_i is the *expected* response of the ith mechanism to the stimulus. The model has been used in a number of studies, with typical values of k in the range of 4–5.

On the other hand, there is substantial evidence that pooling shuts down (Kersten, 1984; Legge & Foley, 1980) for high-contrast stimuli. Thus, despite the evidence for nonlinear pooling in low-contrast conditions, prior data suggests that these mechanisms are unlikely to be acting for our high-noise stimuli. Nevertheless, it is sensible to assess whether the standard multi-channel (multi-scale) model could explain our data if nonlinear pooling effects are incorporated.

The effects of nonlinear pooling are normally assessed under "Quick pooling" assumptions, i.e., that filter outputs are statistically independent, and psychometric functions assume a particular shape (Graham, 1977; Quick, 1974). Due to the high noise input and nonorthogonality of the relevant filters in our experiments, these conventional methods cannot be applied in a straightforward manner. We therefore decided to simulate nonlinear pooling models using Minkowski summation (Watson, 2000). We used QUEST (Watson & Pelli, 1983) to estimate a threshold for each parameterization of the model and for each condition, and then employed a stochastic sampling technique to estimate optimal model parameters. [4]

[4] Due to noise in estimated model thresholds, standard optimization methods such as Nelder–Mead Simplex fail for this application. In our stochastic sampling technique, we first sample the parameter space over a coarse grid and compute the corresponding squared deviation of the model from the human data. Second, we fit a quadratic to the resulting objective function, computing 95% confidence intervals for the quadratic parameters. Third, we randomly sample this 95% confidence region until we find parameters yielding a convex function. Finally, we determine the minimum of this function, and select the corresponding parameters for evaluation of the model on the next iteration. This procedure tends to focus sampling near the actual minimum of the objective function. After a pre-specified number of iterations (between 500–10,000), the quadratic approximation to the objective function is typically convex, and we select the parameters corresponding to its minimum as our maximum likelihood estimate.

The computational cost of this simulation technique prohibits an optimization over all filter locations, orientations, scales, shapes and bandwidths. We therefore restricted our simulation to the vertically-oriented broadband filters we assumed for Models 1–4 (Eq. (1)). However, rather than restricting these filters to be centred only at fixation, we assumed them to be distributed along the vertical stimulus edge. All filters were constrained to be the same shape ε. Since contrast sensitivity in noise is known to be relatively constant up to roughly 16 deg eccentricity (Rovamo, Franssila, & Nasanen, 1992), we assumed the same intrinsic efficiency γ for all filters.

We incorporated diversity in scale by assuming 1, 3 or 5 discrete channels. For the multi-channel versions of the model, scale was sampled logarithmically, and we optimized the minimum scale and the scalespace sampling rate to fit the human data. Filters sampled the vertical edge at fixation and at regular intervals along the vertical edge. Sampling was assumed proportional to the vertical scale constant of the filters, and we evaluated 3 spatial sampling rates: $\sigma_y/2$, σ_y and $2\sigma_y$.

For a given parameterization of the model, each trial yields a vector $\mathbf{r}$ of responses. This vector is first split into half-wave rectified components:

$$\mathbf{p} = [p_1, p_2, \ldots, p_m] \text{ where } p_i = \max(0, r_i)$$

and

$$\mathbf{n} = [n_1, n_2, \ldots, n_m] \text{ where } n_i = -\min(0, r_i)$$

A decision is then made based on the L_m norm of the half-wave rectified components (Minkowski summation):

$$p > n \rightarrow \text{edge is light/dark}$$

$$p < n \rightarrow \text{edge is dark/light}$$

where

$$p = \left(\sum p_i^m\right)^{1/m} \quad \text{and} \quad n = \left(\sum n_i^m\right)^{1/m}$$

The Minkowski exponent m used in this nonlinear pooling model should not be confused with the exponent k used in the Quick pooling model. In the Quick pooling model of probability summation, detection is always determined by maximum selection over multiple channels (i.e. by whether any of the channels exceeds a criterion threshold), and the exponent k is a parameter of an assumed internal noise distribution, which partially determines the steepness of the psychometric function. Note that the Quick pooling model is formed by nonlinear summation over *expected* responses R_i of multiple channels.

In contrast, in our nonlinear pooling model, the exponent m determines how *noisy* responses p_i and n_i to a particular stimulus are combined over channels to form scalar responses p and n. In our case, noise is

normally distributed and in the stimulus. An exponent of $m = 1$ corresponds to linear summation, and in the limit as $m \to \infty$, the Minkowski summation becomes maximum selection: $p = \max_i p_i$ and $n = \max_i n_i$. In this limit, our nonlinear pooling model becomes the analogue of the probability summation model for detection, in that both are governed by maximum selection over multiple channels.

We optimized the one-channel version of the model over three continuous parameters: the scale constants of the filter σ_x, σ_y and the Minkowski exponent m. The three- and five-channel models were optimized over four parameters: the scale constants σ_x, σ_y of the smallest filter, the scalespace sampling rate $\partial\sigma_x$, and the Minkowski exponent m.

We found that neither the number of channels nor the spatial sampling rate had any effect on the fit of the model to the human data. Moreover, the Minkowski exponent m consistently converged to very large values (35–75). In this range, model behaviour is indistinguishable from a maximum selection scheme.

Based on these results, we constructed a simplified nonlinear pooling model for further evaluation. This model consisted of a single channel, with a vertical sampling rate of $2\sigma_y$ (for computational efficiency), with a maximum selection combination rule (probability summation). This simplified model has only 3 free parameters (σ_x, σ_y and γ). Fig. 7 shows the MLE scale parameters for the model, and Fig. 6 (bottom row) shows the fit of the model to the human data. It can be seen that incorporating a nonlinear pooling mechanism greatly improves the fit of the single channel and multiscale models to the data, to be comparable with the 2D scalespace model. However, the model does not fit the data as well as the hyperbolic scalespace model: it underperforms the data for large windows elongated along the edge (Experiment 2), and predicts an overly narrow tuning to small, circular stimuli (Experiment 3).

A perturbation analysis reveals that:

- Decreasing σ_x shifts the curve for Experiment 1 left, shifts the curve for Experiment 2 down, and shifts the curve for Experiment 3 right. Increasing σ_x has the opposite effect.
- Decreasing σ_y shifts all the curves left. Increasing σ_y has the opposite effect.

5.6. Quantitative comparison of models

Of the five models we have evaluated only three (2D scalespace, hyperbolic scalespace, nonlinear pooling) are qualitatively consistent with the human data. We thus subjected only these three models to detailed quantitative comparison. We based this evaluation on the sum of squared deviations of each model from the human data, and used Aikake's Information Criterion (AIC) for model evaluation and selection (Akaike, 1974; Burnham & Anderson, 1998). AIC is an information-theoretic measure for comparing heterogeneous models differing in complexity.

Table 2 shows that of the three models, the hyperbolic scalespace model has the smallest total deviation from the human data. Akaike weights, which can be loosely interpreted as probabilities of model correctness given that one of the models under evaluation must be correct (Burnham & Anderson, 1998), indicate that the hyperbolic scalespace model is by far the most likely to be correct.

Bootstrapping can be combined with AIC to yield a more familiar, 'frequentist' model selection analysis (Burnham & Anderson, 1998). In a typical analysis, data are bootstrapped and MLE model parameters are re-estimated for each bootstrapped data sample and model under evaluation. Next, the AIC is computed for each bootstrapped model, thus taking into account model complexity. Then, for each bootstrapped data sample, the re-estimated model yielding the lowest AIC is selected. Finally, selection frequencies for each model under evaluation are tabulated.

The resulting distributions of AIC estimates for the 2D and hyperbolic scalespace models over 1000 bootstrapped samples are shown by the dotted curves in Fig. 10(a). While the curves overlap substantially, the AIC estimates are highly correlated. Fig. 10(b) shows the AIC difference between the two models: the hyperbolic scalespace model is selected as more probable in 99.1% of trials.

Unfortunately, the computational cost of simulating the nonlinear pooling model prohibits re-estimation of MLE model parameters, and we can only evaluate all three models together for the bootstrapped samples using the MLE model parameters for the actual dataset. However, as can be seen from Fig. 10(a), re-estimation

Table 2
Quantitative model comparison

	2D scalespace	Hyperbolic scalespace	Nonlinear pooling
Free parameters	4	6	3
Sum of squared errors	215	195	213
Akaike's information criterion	−318	−359	−324
Akaike weight	0.000	1.000	0.000
Bootrapped model selection probabilities	1.1%	96.0%	2.9%

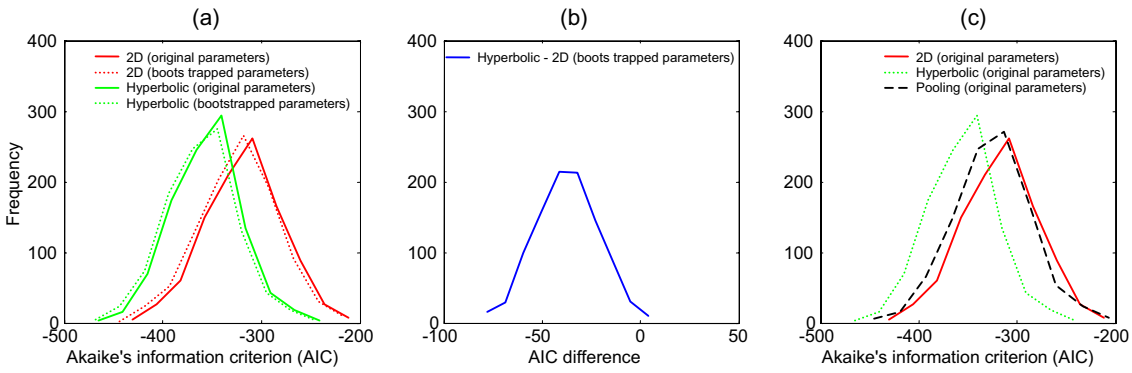

Fig. 10. Bootstrapped Akaike's information criterion for model selection. (a) Bootstrapped AIC values based on original and re-estimated MLE model parameters for 2D and hyperbolic scalespace selection models. (b) AIC difference between hyperbolic and 2D scalespace models, based on re-estimated MLE model parameters. (c) AIC values based on original MLE model parameters for 2D scalespace, hyperbolic scalespace and nonlinear pooling models.

of MLE parameters has little effect on the AIC distributions for 2D and hyperbolic scalespace models. If the same applies to the nonlinear pooling model, this approximation will have little effect on the results of our analysis.

Fig. 10(c) shows the AIC distributions of the three models using the original MLE model parameters, and Table 2 lists model selection frequencies. The hyperbolic scalespace model is selected as most probable for 96% of samples.

6. Discussion

6.1. Neural basis

Fig. 11 shows the filter scales predicted by our models, together with the optimal scales for detecting each stimulus, i.e., the scales for those filters most closely matching our stimuli. It can be seen that all models predict filters distributed over a relatively small region of scalespace, much smaller than the range spanned by our stimuli. This suggests that physiological constraints limit the range of filters available for edge detection. Is this range of scales quantitatively consistent with known receptive field properties of simple cells in early visual cortex?

To answer this question, we examined prior data from monkey (Baizer et al., 1977; DeValois, Albrecht et al., 1982; Devalois, Yund et al., 1982; Levitt et al., 1994; Parker & Hawken, 1988; Schiller et al., 1976), with the understanding that, given poorly-understood inter-species differences and sampling error, we can at most hope to find an approximate correspondence to our models based on human psychophysical data. Since the single channel and multi-scale selection models are qualitatively inconsistent with the human data, we restrict our attention to the 2D scalespace, hyperbolic scalespace and nonlinear pooling models.

Table 3 shows that these three models predict filters with peak spatial frequency tuning in the range 0.8–1.8 cpd. This is in the lower range of peak spatial frequencies for simple cells in V1 (DeValois, Albrecht et al., 1982; Parker & Hawken, 1988). However, DeValois, Albrecht et al. (1982) reported that the distribution of peak spatial frequencies depends on the spatial frequency bandwidth of the receptive field. The bandwidth of the Gaussian first-derivative filter employed in our model is 2.6 octaves. DeValois, Albrecht et al. (1982) found that simple cells with spatial frequency bandwidths in this range were clustered at relatively low peak spatial frequencies. Thus the range of filter widths predicted by all three models is consistent with the spatial frequency characteristics of broadband simple cells in primary visual cortex. [5] These values also fall within the range of peak spatial frequencies measured for simple cells in V2 (Levitt et al., 1994).

The upper bound on the filter length parameter σ_y for the 2D scalespace model was 59.8 arcmin, just outside the receptive field length distribution from a sample of 56 monkey V1 simple cells reported by Parker and Hawken (Parker & Hawken, 1988). This diversity in receptive field lengths is well-approximated by a log-normal distribution. Based on this model, one would expect to find a receptive field of length 59.8 arcmin or greater in 1 out of every 64 cells. This figure therefore constitutes a very reasonable estimate of the upper bound for the population. However, the upper bound of

[5] Neither our data nor our models are inconsistent with the existence of more narrowband neurons tuned to higher spatial frequencies. These neurons are simply not well-suited to detecting edges, and hence are unlikely to determine human performance in our experiments.

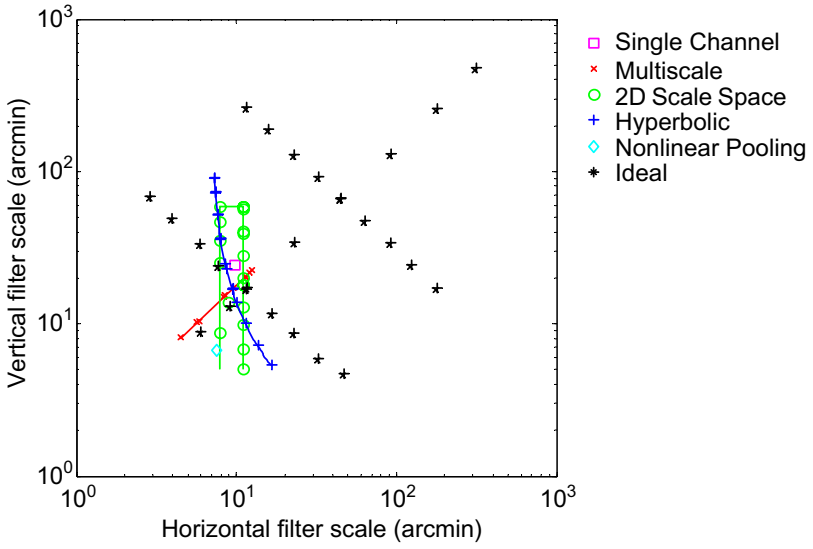

Fig. 11. Optimal filter scales for detection of stimuli employed in this paper, overlaid on filter scales predicted by each model.

Table 3
Predicted receptive field properties

	2D scalespace	Hyperbolic scalespace	Nonlinear pooling
Peak spatial frequency (cpd)	1.2–1.7	0.8–1.8	1.7
Length (arcmin)	<5.1–59.8	5.5–93.5	7.0
Elongation	0.5–7.5	0.3–12.7	0.9
Orientation bandwidth, full-width at half-height (deg)	27–152	16–154	123
Orientation bandwidth, full-width at $1/\sqrt{2}$-height (deg)	16–132	10–141	95

93.5 arcmin found for the hyperbolic scalespace model is less consistent with the Parker and Hawken data, suggesting a neural locus in an extrastriate area such as V2, where receptive fields are larger (Baizer et al., 1977; Levitt et al., 1994).

We found that imposing a lower bound α_y on the length scale constant did not improve the fit of the 2D scalespace model to the data, and increasing α_y above 5.1 arcmin increased the error of the fit. Thus the 2D scalespace model suggests that the neural population includes filters with length scale constants as low as 5.1 arcmin. The hyperbolic scalespace model predicts a lower bound of 5.5 arcmin. Only three of the 56 V1 simple cells examined by Parker and Hawken had length scale constants below 5 arcmin. Thus the lower bounds predicted by our models are in rough agreement with the physiological data.

Schiller et al. (1976) measured orientation bandwidth at $1/\sqrt{2}$ height for simple cells in V1, reporting a range of 10 deg to more than 120 deg. Devalois, Yund et al. (1982) and Parker & Hawken (1988) measured orientation bandwidth at half-height for simple cells in V1, and Levitt et al. (1994) used the same measure for simple cells in V2. Devalois, Yund et al. reported orientation bandwidths ranging from 6 to 360 deg. Parker and Hawken reported orientation bandwidths ranging from less than 10 deg to more than 90 deg. However, they excluded 7 of 105 cells from analysis because their orientation tuning was too weak to be measured. It is possible that these cells correspond to the higher range of orientation bandwidths reported by De Valois et al. Levitt et al. reported a range of 26 deg to more than 180 deg for simple cells in V2.

The 2D scalespace and hyperbolic scalespace models predict an almost full use of this range of orientation bandwidths for edge detection (Table 3), with the exception of the most weakly-tuned neurons (full-width at half-height bandwidths greater than about 153 deg). It is possible that these weakly orientation-tuned neurons are poorly matched to our stimuli in other ways (e.g., spatial frequency bandwidth). The nonlinear pooling model predicts a single detection filter toward the upper end of this range.

In summary, all three models are roughly consistent with known receptive field properties of neurons in V1 and V2. The two filter selection models predict a broad diversity in receptive field shape and orientation tuning roughly consistent with physiological data, but only relatively modest variation in receptive field width (scale) for broadband (2.6 octave) receptive fields. The

parameters of the 2D Scalespace model are consistent with a neural locus for edge detection in primary visual cortex (V1). However the longer receptive fields predicted by the hyperbolic scalespace model suggest an extrastriate mechanism. The nonlinear pooling model exploits only one of the range of mechanisms provided by the early visual cortex; the parameters of this mechanism fall within the range measured in both V1 and V2.

6.2. Why a hyperbola?

Since three of our models are at least qualitatively consistent with the human data, it is too early to declare any one of them correct: future experiments designed to more clearly discriminate between them are needed. Nevertheless, since our quantitative analysis suggests that the hyperbolic scalespace model is the most probable, it is natural to wonder why the filter population would lie on a hyperbolic curve in two-dimensional scalespace. One possibility is that the advantage of receptive field diversity for detection is tempered by the anatomical and physiological cost of synthesizing large linear receptive fields. This cost is likely related to the number of photoreceptors over which the receptive field integrates, i.e., to the area of the receptive field, which is proportional to the product of the two receptive field scales $\sigma_x\sigma_y$. The hyperbolic model may thus reflect a tradeoff, permitting receptive field diversity while controlling the total input cost of the representation through the second order term in $\sigma_x\sigma_y$.

6.3. What does the eye see best?

The few detection studies that have systematically investigated detection performance as a function of stimulus shape have employed Gabor stimuli in low-contrast conditions (Polat & Tyler, 1999; Watson et al., 1983). Watson et al. measured contrast thresholds for drifting Gabor patches over a range of parameters. They found the optimal Gabor stimulus to be a circular patch with a carrier frequency between 6 and 8 cpd and a diameter of roughly 3 cycles, drifting at 4 Hz, and presented for a duration of 160 ms.

In the present study, we restricted our attention to stationary, broadband stimuli, embedded in high contrast noise, presented for 153 ms. Measuring width at half-height, Watson et al.'s optimal stimulus had a diameter of roughly 0.4 deg. In our first experiment, we found optimal efficiency for a circular stimulus with a diameter between 0.3 and 1.3 deg. The hyperbolic scalespace model predicts optimal efficiency for a stimulus roughly 0.5 deg in diameter. While Watson et al. report that a circular stimulus is optimal, our third experiment suggests that efficiency is quite broadly tuned for stimulus shape: efficiencies for stimuli ranging in elongation from 1:4 to 4:1 are not significantly different.

Polat and Tyler also measured detection performance for Gabor stimuli of roughly the same width (0.5 deg at half-height). They estimated the elongation of the most sensitive mechanisms to lie between 1.8:1 and 3.9:1. However, while the decline in sensitivity for elongations greater than 4:1 is fairly clear from their data, the decline for elongations less than 4:1 is less clear.

In summary, despite differences in methods and stimuli, our study is in rough agreement with the results of Watson et al. regarding the optimal stimulus size. On the issue of shape, results are more equivocal. While Watson et al., reported optimal detection for circular stimuli and Polat and Tyler reported elongated stimuli to be optimal, our own results suggest broad tuning over stimulus shape. Differences in stimuli and methodology may account for these discrepancies: more comprehensive experiments under consistent conditions will be required to determine this.

6.4. Prior models

Most previous models for visual detection assume some diversity in the underlying linear detection mechanisms, but typically this diversity is restricted to one-dimensional spatial frequency tuning properties, either over a number of discrete mechanisms (Watson, 2000; Watson et al., 1983; Wilson & Bergen, 1979; Wilson et al., 1983) or a continuum (Kersten, 1984; Watt & Morgan, 1984). One exception is the study of Geisler & Albrecht (1997), who measured the response of neurons in macaque primary visual cortex to drifting sinusoidal gratings, and estimated tuning to a number of stimulus properties. They modeled the sensitivity of a neuron to each stimulus parameter as a random variable and estimated the generating distributions from the neural response data.

Our filter selection model is in the spirit of Geisler & Albrecht (1997), Kersten (1984) and Watt & Morgan (1984) in that it assumes a population of mechanisms over a continuum in some parameter space. It differs principally in our attention to the joint distribution of filter length and width tuning parameters.

Geisler and Albrecht used their model to compare neural and behavioural discrimination of contrast and spatial frequency. Although they estimated a diversity of orientation tuning bandwidths, they did not compare their model to behavioural orientation discrimination. Since they used only extended grating stimuli and modeled only the marginal distributions for individual stimulus parameters, their study cannot be used to infer the two-dimensional shape of the underlying mechanisms.

Kersten (1984) measured detection thresholds for vertical grating stimuli in noise over various spatial frequencies. The stimuli were fixed in height, and

Gaussian-windowed in the horizontal direction. Kersten found very little pooling across gratings in noise, with efficiency peaking when only 1 cycle of the carrier was visible. He proposed a cross-correlator model based on filter selection from a family of 1.7 octave filters over a range of peak spatial frequencies. He suggested that simple cells in striate cortex might serve as the detectors in this model.

Our filter selection theory is very similar in form to Kersten's cross-correlator model. Since Kersten used gratings of fixed height, his model is necessarily one-dimensional, whereas we are explicitly interested in the two-dimensional shape of these filters. Kersten inferred a 1.7 octave bandwidth from his grating data, whereas we assume a priori a 2.6 octave bandwidth filter, roughly matched to our stimuli. Both values fall in the range of bandwidths reported for simple cells in V1 (DeValois, Albrecht et al., 1982) and V2 (Levitt et al., 1994).

While Kersten posited mechanisms ranging broadly in scale (peak spatial frequency tuning of 0.5–32 cpd), we infer mechanisms with modest scale variation, but varying broadly in shape. Kersten also implicitly assumed these multiple scale mechanisms to have substantially different intrinsic efficiencies, whereas we assume all mechanisms to have the same intrinsic efficiency. Given the difference in the stimuli, it is likely that the exact nature of the filters underlying detection of Kersten's grating patches are different than those underlying edge detection. However it is also possible that diversity in filter shape plays a role in Kersten's grating detection results as well.

6.5. Pooling and nonlinear mechanisms

In their 1997 study, Geisler and Albrecht contrasted two models relating neural sensitivities to behavioural performance: the first based upon the envelope of the most sensitive neurons, and the second based on optimal pooling. Their first model corresponds roughly to our selection models, whereas their optimal pooling model corresponds to our pooling model with an exponent of $m = 1$.

While Geisler and Albrecht found the two models to be equally consistent with the data, in the present work we found we were better able to account for the psychophysical data by a filter selection theory involving a family of filters of various shapes than by a nonlinear pooling theory involving filters of fixed shape. Given the strong agreement between the human data and the predictions of the hyperbolic scalespace model, it is possible that the high amplitude stimulus noise was effective in shutting down pooling mechanisms (Kersten, 1984; Legge & Foley, 1980). However, it is still possible that some nonlinear pooling model over a more diverse set of linear filters could explain the data just as well. Our evaluation of nonlinear pooling models suggests a large Minkowski exponent m, corresponding to maximum selection (true probability summation).

Chen & Tyler (1999) have demonstrated phase-insensitivity for low-contrast detection of colinearly-arranged Gabor patches, suggestive of a nonlinearity (at least a rectification) in the detection mechanism. It is not known whether similar results would be obtained in the high contrast conditions of our own experiments.

6.6. Implications for computer vision algorithms

The first computer vision algorithms for edge detection employed detection filters of fixed shape and scale (Canny, 1986; Hueckel, 1973; Roberts, 1965). More recent filter selection algorithms have demonstrated the advantage of using filters over a range of scales (Elder & Zucker, 1998; Lindeberg, 1998). The results of the present study suggest that the human visual system may rely more on diversity in filter shape than scale. One obstacle to exploiting this idea in computer vision algorithms is computational cost. Modern edge detection algorithms typically rely upon steerable, separable filters for computational efficiency (Freeman & Adelson, 1991; Perona, 1995). Derivative filters based on isotropic Gaussian kernels are both steerable and separable, but filters based on elongated kernels generally are not. Thus practical edge detection algorithms employing filters ranging in shape may require the invention of more efficient methods for convolution with nonseparable, nonsteerable kernels.

The performance of edge detection algorithms has also been improved by application of more global, nonlinear methods for computing support along image curves, e.g., relaxation labeling (Hummel & Zucker, 1983; Zucker, Hummel, & Rosenfeld, 1977), hysteresis with thresholding (Canny, 1986). These methods correspond roughly to the nonlinear pooling model we have evaluated. While we did find that inclusion of a nonlinear pooling mechanism improved the correspondence of our single-channel and multi-scale models to the human data, our best account of the data relies on a simple, local filter selection strategy, over a hyperbolic curve in two-dimensional scalespace.

6.7. Limitations of the present study

Our nonlinear pooling model assumed filters located along the vertical edge of the stimulus; we ignored filters centred on or near the elliptical boundary of the stimulus. There are several reasons to believe that the effect of these filters would be negligible:

1. The contrast of the vertical edge is double the contrast of the elliptical boundary. This means that the responses of the filters tuned to the vertical edge will

be roughly twice as large as the responses of filters tuned to the elliptical boundary.
2. The curvature of the elliptical boundary will further reduce the response of these filters.
3. Our analysis of nonlinear pooling models suggests a very high exponent—effectively maximum selection (true probability summation). In such a model, only responses of comparable magnitude have any effect on detection performance.

Contrast sensitivity for a 3 cpd grating in noise is known to be constant up to roughly 16 deg retinal eccentricity (Rovamo et al., 1992). However, it is possible that the range of filter scales in the neural population may increase with eccentricity. Incorporating this variation into our nonlinear pooling model might improve the fit of the model slightly for the positive elongation conditions of Experiment 2, where a larger filter at eccentric locations would boost efficiency toward human levels. However, the main deviation of the model from the data occurs for Experiment 3, and we would expect no effect of this modification for these relatively small stimuli, for which the maximum eccentricity of the elliptical window is less than 2.7 deg.

We have restricted our attention in this study to a very simple stimulus: a vertical luminance step edge within an elliptical window. Our interest in edges stems from their importance in signaling events of interest in natural visual scenes: boundaries of objects and shadows, changes in surface reflectance and attitude. By restricting the nature of the stimulus, we have been able to focus on the issue of shape and scale diversity in edge detection mechanisms. However, our results do not necessarily generalize to other forms of simple stimuli such as blurred edges, curved contours, or to more complex visual scenes.

The human visual system is sensitive to visual edges over modalities other than luminance, including colour, texture, stereoscopic depth, and motion. The present study pertains only to the detection of luminance edges; our results and models do not necessarily generalize to these other modalities.

7. Conclusions

Our results indicate that the front-end filters underlying human edge detection are relatively small, consistent with receptive field sizes of simple cells in areas V1 and V2. Detailed modelling suggests a spatially local optimal selection mechanism over a diverse population of filters lying on a hyperbolic curve in two-dimensional scalespace. In contrast to the standard multi-channel model of visual processing, these filters are found to be much more diverse in their shape (orientation bandwidth) than their scale (peak spatial frequency tuning). An alternate model involving nonlinear spatial pooling (probability summation) over a less diverse filter population is qualitatively consistent with the data, but quantitatively less consistent than the local filter selection model.

Acknowledgements

This research was supported by grants from NSERC, GEOIDE, CRESTech and the PREA. We thank Mark Georgeson and two anonymous reviewers for their very helpful comments.

References

Akaike, H. (1974). New look at the statistical model identification. *IEEE Transactions on Automatic Control, AC19*(6), 716–723.

Baizer, J. S., Robinson, D. L., & Dow, B. M. (1977). Visual responses of Area 18 neurons in awake, behaving monkey. *Journal of Neurophysiology, 40*(5), 1024–1037.

Barlow, H. B. (1962). A method of determining the overall quantum efficiency of visual discriminations. *Journal of Physiology, 160*, 155–168.

Bonneh, Y., & Sagi, D. (1999). Contrast integration across space. *Vision Research, 39*(16), 2597–2602.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and multimodel inference*. New York: Springer.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(6), 679–698.

Chen, C. C., & Tyler, C. W. (1999). Spatial pattern summation is phase-insensitive in the fovea but not in the periphery. *Spatial Vision, 12*(3), 267–285.

DeValois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

Devalois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22*(5), 531–544.

Elder, J. H., & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(7), 699–716.

Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*(9), 891–906.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience, 14*(5), 897–919.

Graham, N. (1972). Spatial frequency channels in the human visual system: Effects of luminance and pattern drift rate. *Vision Research, 12*(1), 53–68.

Graham, N. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research, 17*(5), 637–652.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: Comparison of single-channel and multiple-channels models. *Vision Research, 11*(3), 251–259.

Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research, 18*(7), 815–825.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields and functional architecture in the cat's visual cortex. *Journal of Neuroscience, 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Neuroscience, 195*, 215–243.

Hueckel, M. H. (1973). A local visual operator which recognizes edges and lines. *Journal of Assisted Computers and Machines, 20*(4), 634–647.

Hummel, R. A., & Zucker, S. W. (1983). On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*(3), 267–287.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research, 24*(12), 1977–1990.

Koenderink, J. (1984). The structure of images. *Biological Cybernetics, 50*, 363–370.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of Optical Society of America, 70*(12), 1458–1471.

Levitt, J. B., Kiper, D. C., & Movshon, J. A. (1994). Receptive fields and functional architecture of macaque V2. *Journal of Neurophysiology, 71*(6), 2517–2542.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision, 30*(2), 117–154.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of Royal Society of London B, 207*, 187–217.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology—London, 283*(October), 79–99.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in receptive-fields of simple cells in cat's striate cortex. *Journal of Physiology—London, 283*(October), 53–77.

Parker, A. J., & Hawken, M. J. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journal of Optical Society of America A, 5*, 598–605.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. B. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge, UK: Cambridge University Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of Optical Society of America A, 16*(3), 647–653.

Perona, P. (1995). Deformable kernels for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*(5), 488–499.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. Trans. IRE PGIT, 4, 171–212.

Polat, U., & Tyler, C. W. (1999). What pattern the eye sees best. *Vision Research, 39*, 887–895.

Quick, R. F. (1974). Vector-magnitude model of contrast detection. *Kybernetik, 16*(2), 65–67.

Roberts, L. (1965). Machine perception of 3-dimensional solids. In J. Tippett (Ed.), *Optical and electro-optical information processing*. Cambridge, MA: MIT Press.

Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *Journal of Optical Society of America, 38*(2), 196–208.

Rovamo, J., Franssila, R., & Nasanen, R. (1992). Contrast sensitivity as a aunction of spatial-frequency, viewing distance and eccentricity with and without spatial noise. *Vision Research, 32*(4), 631–637.

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex 2. Orientation specificity and ocular dominance. *Journal of Neurophysiology, 39*(6), 1320–1333.

Stromeyer, C. F., & Klein, S. (1975). Evidence against narrow-band spatial frequency channels in human vision: Detectability of frequency modulated gratings. *Vision Research, 15*(8-9), 899–910.

Verghese, P., & Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature, 381*(6578), 161–163.

Watson, A. B. (1982). Summation of grating patches indicates many types of detector at one retinal location. *Vision Research, 22*, 17–25.

Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express, 6*(1), 12–33.

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature, 302*(5907), 419–422.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*(2), 113–120.

Watt, R. J., & Morgan, M. J. (1984). Spatial filters and the localization of luminance changes in human vision. *Vision Research, 24*(10), 1387–1397.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research, 19*, 19–32.

Wilson, H. R., & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of Optical Society America A, 1*(1), 124–131.

Wilson, H. R., McFarlane, D. K., et al. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research, 23*(9), 873–882.

Young, R. A. (1991). Oh say can you see? The physiology of vision. In B. E. Rogowitz, M. H. Brill, J. P. Allebach (Ed.), *SPIE proceedings: Human vision, visual processing, and digital display II*. Bellingham, WA: SPIE—The International Society for Optical Engineering, p. 1453.

Zucker, S. W., Hummel, R., & Rosenfeld, A. (1977). An application of relaxation labeling to line and curve enhancement. *IEEE Transactions on Computers, 26*, 394–403.