To be published in: Journal of the Optical Soceitey of America A, 14, 1997.

A Model of Visual Contrast Gain Control and Pattern Masking

Andrew B. Watson[1] & Joshua A. Solomon[2]

[1]NASA Ames Research Center, Moffett Field, CA 94035-1000, beau@vision.arc.nasa.gov

[2]Institute of Ophthalmology, Bath Street, London EC1V 9EL, UK, j.solomon@ucl.ac.uk

We have implemented a model of contrast gain control in human vision which incorporates a number of key features, including a contrast sensitivity function, multiple oriented band-pass channels, accelerating nonlinearities, and a divisive inhibitory gain-control pool. The parameters of this model have been optimized through a fit to the recent data that describe masking of a Gabor function by cosine and Gabor masks [Foley, J. M. (1994). Human luminance pattern mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A 11*(6), 1710-1719]. The model achieves a good fit to the data. We also demonstrate how the concept of recruitment may accommodate a variant of this model in which excitatory and inhibitory paths share a common accelerating non-linearity, but which include multiple channels tuned to different levels of contrast [Teo, P. C. & Heeger, D. J. (1994). Perceptual image distortion . *Proceedings, ICIP-94*, Austin, Texas, IEEE Computer Society Press, II, pp. 982-986].

1. INTRODUCTION

With some notable exceptions, spatial patterns are most easily seen against a uniform background; backgrounds that contain spatial patterns typically raise visual thresholds. Understanding this phenomenon of pattern masking is an important part of understanding the process of pattern detection and pattern visibility in general.

Visual masking describes a broad range of phenomena, which may be arranged in various operational or theoretical taxonomies. Here we deal exclusively static target and masking patterns that appear simultaneously and vary only in regard to their relative intensities on successive displays. In addition, the mask is a simple pattern such as a sinusoid or Gabor function. We will call this pattern masking. Considerations associated with stochastically defined masks such as visual noise remain outside the scope of the current paper.

There are two traditional explanations for pattern masking. In one, the mechanism detecting the target has a nonlinear, compressive response. The mask activates this mechanism, and pushes its response into the compressive range. The differential between responses to mask alone and target+mask is thereby reduced, and threshold elevated[1, 2]. In the second explanation, the mask inhibits the target detection mechanism, either directly or through other mechanisms. More recently, models have been proposed which incorporate both of these mechanisms within a process of contrast gain control[3, 4, 5]. These psychophysical models are largely inspired by recent analyses of the response properties of single visual neurons in primary visual cortex[6, 7, 8]. Here contrast gain control is a mechanism that serves to keep neural responses within their permissible dynamic range while retaining the information conveyed by the pattern of activity over the neural ensemble. In the normalization model of Heeger[8], each neuron has an accelerating nonlinearity but is also inhibited divisively by a pool of responses of other neurons. In the psychophysical model of Teo and Heeger that is closely based upon this cortical normalization theory, masking occurs through the inhibitory effect of this normalizing pool. Foley's model of masking also incorporates a divisive inhibitory pool.

Despite their success in predicting certain masking data, there remain reasons to consider alternative models. Foley's model was designed to predict results for a narrow range of experimental stimuli, and can make no predictions for other stimuli. To accommodate generic two-dimensional stimuli, we require that the model be image-driven, that is, it must accept images as inputs. The Teo and Heeger model is image-driven, but does not specify certain aspects, such as the contribution of different spatial frequencies to the inhibitory pool. In addition, their model places a rigid constraint on the form of the nonlinearity, which in turn obliges them to posit multiple mechanisms responsive to different ranges of contrast. The model of Wilson and Humanski is particularly concerned with temporal dynamics of the gain control process, and consequently devotes less attention to spatial details. Cannon and Fullencamp have developed an image-driven model that incorporates a gain control process, but it is designed only to predict estimates of apparent contrast[9].

Additional impetus for a general image-driven model of spatial masking arises from the enduring search by various engineering communities for a practical and accurate model of the visibility of spatial patterns. In contexts such as display design and image compression, the model must be general enough to deal with any imagery that might be displayed or compressed. Parameter-driven models designed to deal with simple patterns such as several sinusoidal gratings typically offer little guidance on what to do with complex images such as photographs. There have in fact been numerous efforts to develop image-driven models for evaluation of image quality[10], several of which incorporate pattern masking mechanisms. The model of Daly[11] is particularly complete, but does not directly predict detection thresholds. The model of Lubin[12] does predict thresholds, but like Daly's, assumes that contrast gain control occurs only within a channel.

To explore the various elements of the contrast gain control process, and to provide a general image-driven model of pattern masking that might be used in applied contexts, we have developed a model of pattern masking based on contrast gain control. We have applied this model to recent psychophysical results of Foley (1994), with generally excellent results.

2. MODEL

As discussed in the introduction, most of the existing models of masking and contrast gain control share many features. Likewise it is unclear at this time precisely which features are essential to successful prediction of human experimental data. Consequently we designed a model with a modular structure, whose building blocks correspond to a large extent with specific individual assertions about the processing of luminance contrast signals.

A. Generic Model

An overview of this generic model is pictured in Figure 1. The input to the model is a pair of images. These might be, for example, the two images presented to an observer in a two-alternative forced-choice experiment. Each image traverses an identical set of processing stages, which are illustrated for image 1. The first stage is a linear filter bank. In general, the output of this stage will be a set of filtered images, one for each filter in the bank. Borrowing a term from signal processing, we call these sub-bands[13]. The next stage is a sampling operation, which may alter the number and positions of samples in each sub-band.

At this point, the signal forks into excitatory and inhibitory paths, each of which begins with a point nonlinearity. The inhibitory stage is then subject to a second linear filtering operation, which we characterize as a pooling operation. This pooled signal then drives an inhibitory operation upon the excitatory path.

Finally, the processed representations of the two images are compared, leading to a decision.

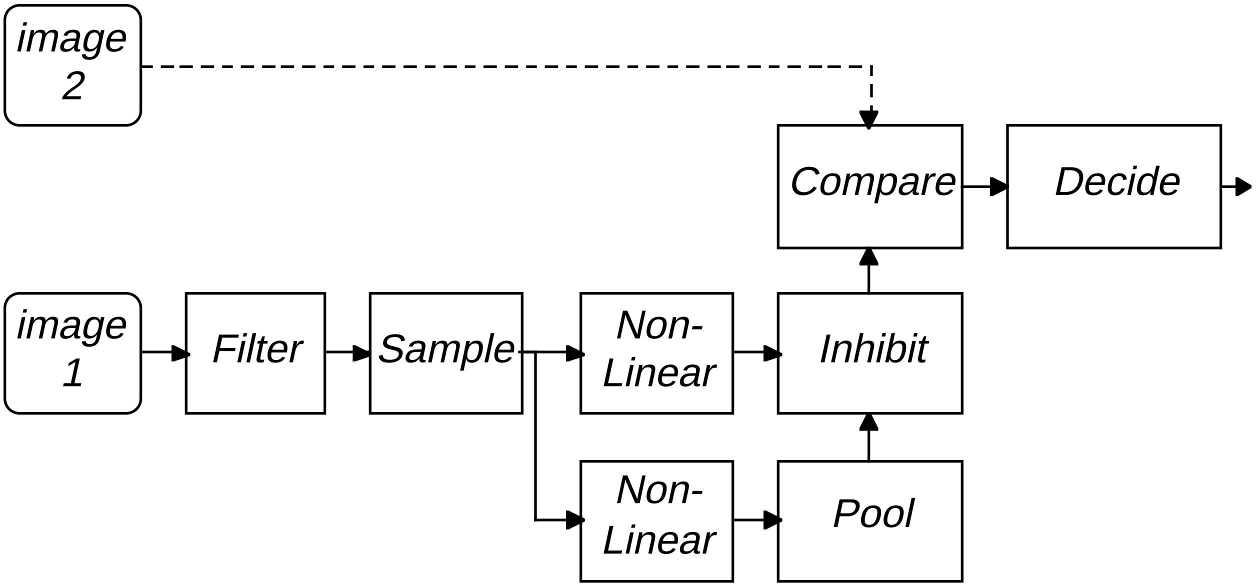

Figure 1. Outline of the generic image discrimination model incorporating contrast gain control.

This generic model can clearly accommodate a wide range of specific choices regarding filtering, nonlinearities, the inhibitory process, and the comparison mechanism. In the next section we describe the specific choices we have made, and compare them to those made by other comparable models.

B. Specific Model

Our specific model is illustrated in Figure 2. It shows the various choices we have made regarding the generic components. These are discussed in more detail in the following sections.

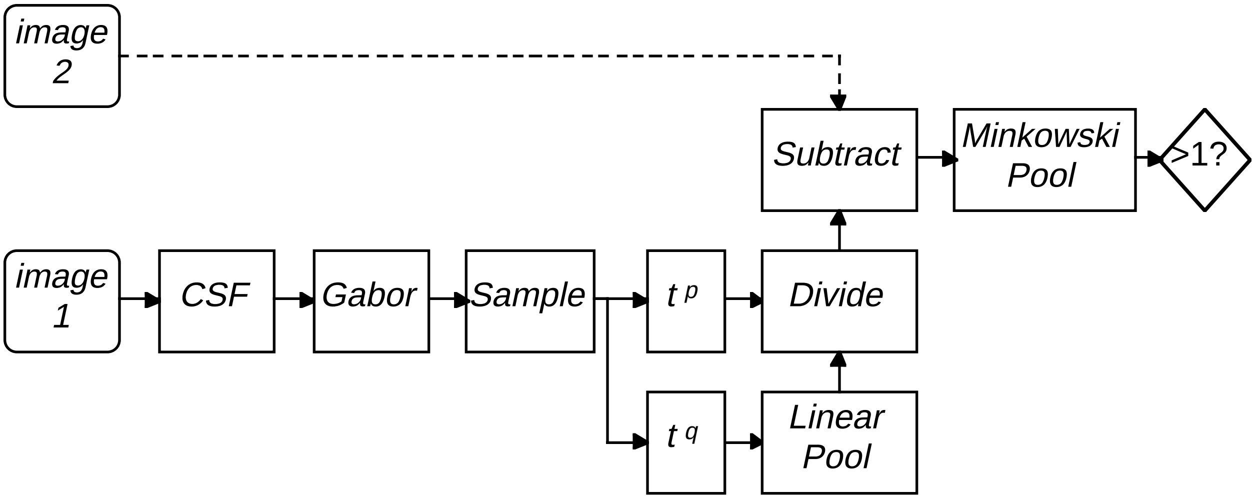

Figure 2. Outline of the specific model used in our contrast gain control model of pattern masking.

C. Example stimuli

To illustrate the behavior of the model it will be useful to consider a pair of example stimuli, as shown in Figure 3. The constituent parts are shown in Figure 3a and b: a cosine grating at a frequency of 4 cycles/image and an orientation of 45°, and a vertical Gabor function of the same frequency. These elements are combined to form the two signals of a two-alternative forced-choice detection task. In Figure 3c) is the mask alone at 25% contrast, and in d) the mask is combined with a 50% contrast Gabor target.

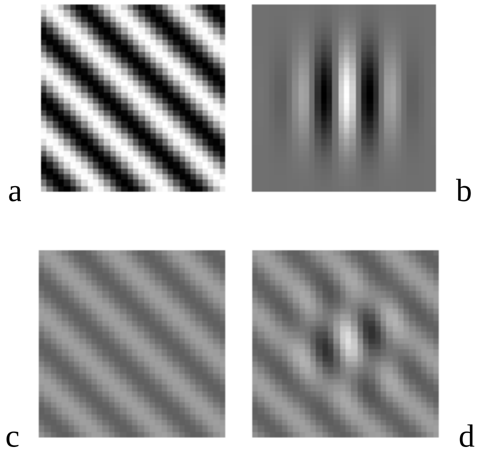

Figure 3. Example stimuli to illustrate the operation of the model: a) cosine mask at 45° orientation, b) Gabor target, c) mask at 25% contrast, d) mask at 25% contrast plus target at 50% contrast.

D. Contrast Sensitivity Filter

Variations in contrast sensitivity can be incorporated into a filter bank model in either of two general ways. The first is to set the gain of each of the filters in the bank in such a way that the ensemble produces the empirical contrast sensitivity. The second is to place a single contrast sensitivity filter (CSF) at the front end, and to calibrate the remainder of the model in such a way that it introduces no further variations in contrast sensitivity. There are many issues involved in this decision, but here we take the latter course.

The contrast sensitivity filter was designed to match in shape the CSF measured with Gabor stimuli. The data in Figure 4[14], show that a parabola in log-log coordinates is a reasonable description of this function. This log-log-parabolic filter is then implemented as a digital filter, with parameters of peak sensitivity, peak frequency, and log bandwidth.

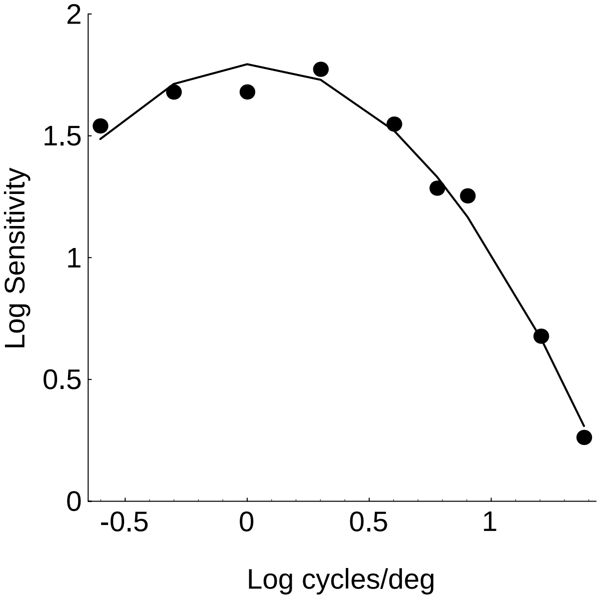

Figure 4. Contrast sensitivities for one-octave Gabor targets as a function of spatial frequency, fit by a parabola in log-log coordinates (from [14]). The parameters are peak sensitivity = 62.24, peak frequency = 1.04 cycles/degree, and log bandwidth at half height = 1.118.

E. Gabor Array

We have experimented with two different multiple-channel filter banks. The first is the set of filters defined by the Cortex Transform[15], as modified by Daly[11]. The second, from which all of the results below will be derived, is a filter set that we call the Gabor Array. It is a collection of Gabor filters that vary in spatial frequency, orientation, and phase. It is convenient to consider the filters as forming an array, in which each row corresponds to a particular spatial frequency, and each column to a particular orientation, as shown in Figure 5. The distinctive feature of this type of filter bank is the Gabor shape for each filter, and the rectangular sampling of the orientation-log frequency plane. Within these constraints, there is freedom in selection of frequency spacing and bandwidth, and orientation spacing and bandwidth. The output of the filter bank is an array of images equal in number to the number of filters, and each equal in size to the input.

In Figure 5 and succeeding figures, we illustrate the bahaviour of the model as it responds to signal and mask. In these illustrations, we show a Gabor Array with three frequency bands and four orientation bands, with one octave frequency bandwidths, and other parameters set to reasonable values. However, the actual values of model parameters used in fitting experimental data may differ and are given later in this paper.

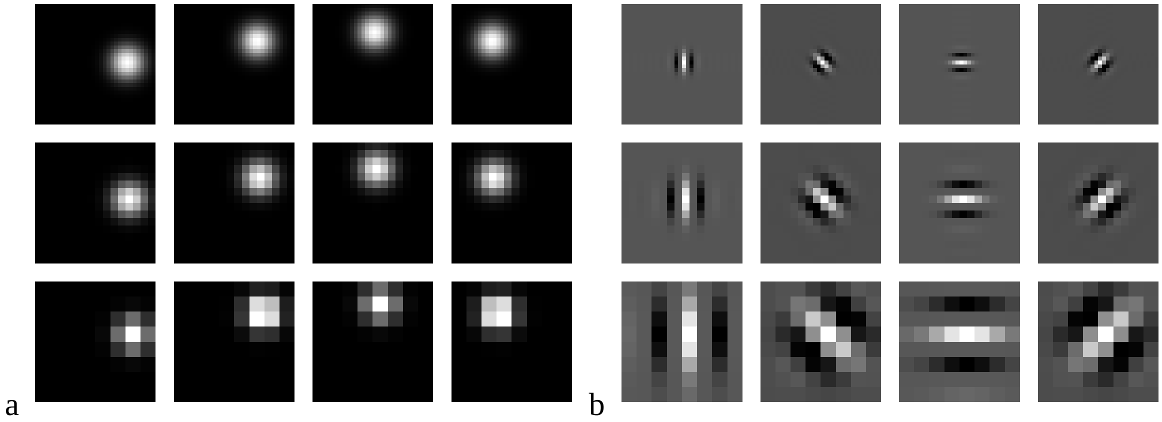

Figure 5. A Gabor Array filter bank, with three spatial frequencies and four orientations. The transfer functions (a) and even impulse responses (b) of each filter are shown, all scaled to unit amplitude.

Although we describe many of the mechanisms of this model in image-processing terms, we emphasize that the underlying physical model is that of arrays of visual neurons. Where possible, we make use of simplifying assumptions or algorithms. One of these relates to the sign of our elementary responses. Cells in primary visual cortex typically have little or no maintained discharge, and it seems likely that they signal only with positive responses. To a first approximation they appear to half-wave rectify their underlying linear responses[16]. It is conventionally assumed that positive and negative deviations of the stimulus are signaled by pairs of complementary neurons that are 180° out of phase, much like the on and off center cells at earlier levels of the visual pathway. This leads to a set of four hypothetical phases of the individual receptive fields: 0°, 90°, 180°, and 270°. Since in the absence of response noise only one of each opposed pair will respond to a given signal, we simulate this situation with just two phase receptive fields (0° and 90°), each of which produce a signed responses. It should be understood that a negative response simulates the positive response of a negative phase cell. The pictures throughout this paper which show both positive and negative deviations from zero should be viewed as the ensemble response of two sets of opposite phase neurons. Likewise mathematical expressions which depict responses indexed by phase should be understood to represent all-positive responses at one of four phases.

A second obvious approximation in our simulation of neural populations is that ours is a *sampled* model in which a set of neurons at representative points in space, frequency, orientation and phase are used to approximate the presumably more continuous distribution in nature.

The Gabor Array filter bank is implemented as a set of analytic filters, which produce a complex output from a real input[17, 18]. The real and imaginary parts of this complex response represent the responses of even (0° and 180°) and odd (90° and 270°) filters, respectively. Likewise positive values represent responses of 0° and 90° phase neurons, while negative values represent responses 180° and 270° phase neurons.

F. Sampling

To this point, the model has multiplied the dimensionality of the signal by a factor equal to the number of filters. While there is some reason to believe that the primary visual cortex may oversample its retinal input, and while such oversampling may have advantages in terms of shift invariance, there are also powerful advantages to down-sampling the lower frequency channels. First, this allows a “pyramid” style of representation, in which each sub-band is sampled in proportion to its characteristic spatial wavelength. Second, downsampling in proportion to wavelength greatly reduces the amount of subsequent computation required.

We have experimented with both down-sampled and un-sampled representations, and have generally found that within the limits of our experiments, and provided that the filters are appropriately normalized (see below), we obtain very similar results. We have therefore worked mainly with the sampled variant since it is computationally much more efficient.

Specifically, for our octave-spaced frequency channels, we have down-sampled each sub-band in each spatial dimension by a factor of 2^L, where L is the level which is 0 for the highest frequency band. Our Gabor filters have a center frequency of N 2^{-L-1} for each level, where N is the Nyquist frequency of the input. Down-sampling by of 2^L will preserve information provided the response at that level has no energy above N 2^{-L}, that is, one octave above the filter frequency. This is approximately true for Gabor filter bandwidths of less than about 1.5 octaves.

Figure 6 shows the down-sampled real response array to the mask alone (a) and the mask+target (b) example stimuli. Each row contains response images for one spatial frequency, while each column corresponds to a particular orientation. The arrangement is the same as for the filters in Figure 5(b). Note that due to downsampling, the number of samples in the third row is sixteen times fewer than in the first row. As expected, the mask-alone responses appear primarily in the sub-band at the corresponding frequency and orientation, while the mask+target responses also show activity in the sub-band tuned to the target. As noted above, the filters produce both real and imaginary response images; here and in subsequent figures we show only the real part.

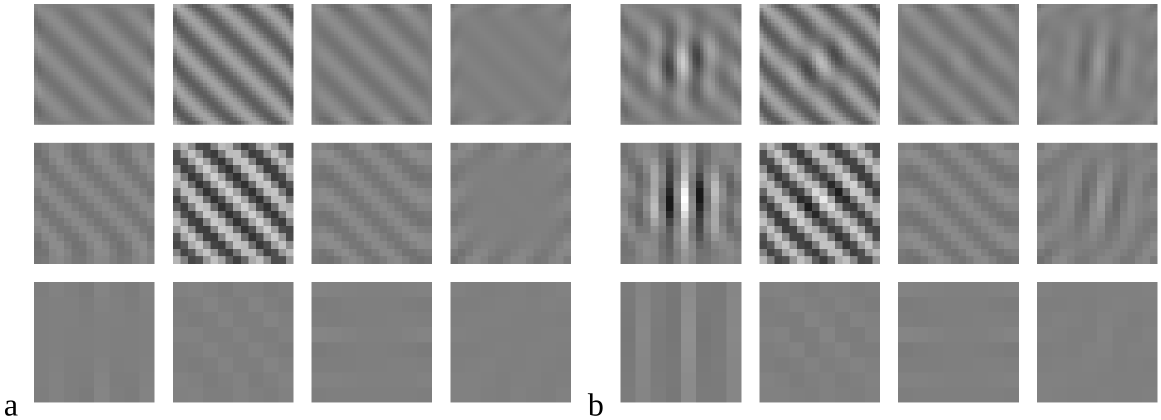

Figure 6 Responses of the Gabor filter bank to mask alone (a) and mask+target (b).

G. Excitatory Non-linearity

Each scalar sample in the excitatory path undergoes a power law nonlinearity with an exponent of p . In the model of Foley, this is typically a value between 2 and 3, while in the model of Teo & Heeger it has a value of 2. Note that even and odd samples separately undergo this nonlinearity. In addition, the nonlinearity is applied to the unsigned response magnitude, to which the sign is then re-attached. This conforms to the sign-preservation premise described above.

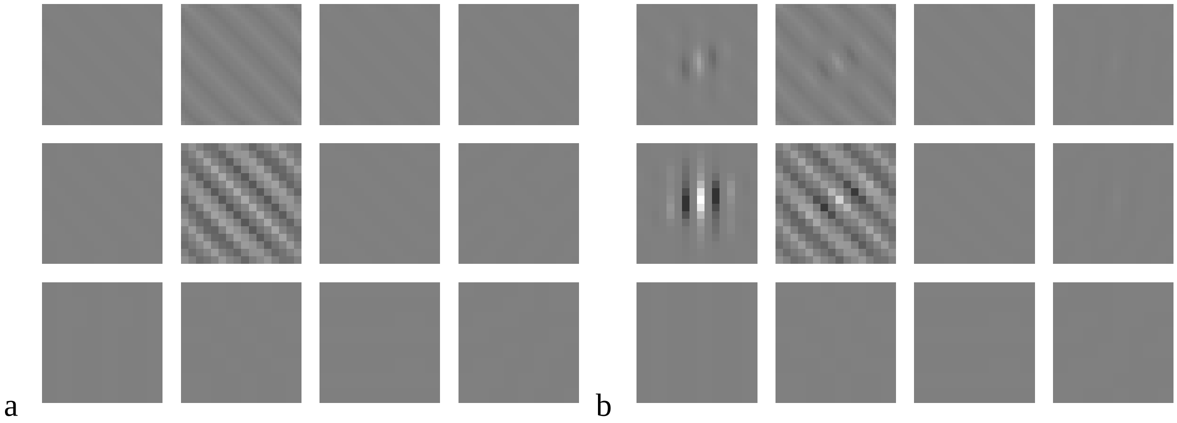

Figure 7. Responses following the excitatory nonlinearity to a) mask, b) mask+target.

Figure 7 shows responses to mask-alone and mask+target following the excitatory non-linearity. Comparison of Figure 6 and Figure 7 shows that the excitatory nonlinearity tends to suppress small responses, and amplify large ones.

H. Inhibitory Non-linearity

The inhibitory nonlinearity is identical to that in the excitatory path, except for a possibly different exponent q. We have generally investigated values of q = 2. In Foley's model this exponent is less than p, and usually around 2. In Teo & Heeger's model it is 2.

I. Inhibitory Pooling

The pooling operation in the inhibitory path linearly combines signals over the five dimensions of phase (0°, 90°, 180°, 270°), frequency, orientation, and space (x and y). In its most general form, this combination would be specified by a five-dimensional kernel that is itself a function of five dimensions. We make the following simplifying assumptions. We assume complete summation over phase. This agrees with Teo & Heeger's model, while Foley's model does not address this point. We also assume that pooling is shift invariant over level (log frequency), orientation, and space. These two assumptions allow us to implement the pooling by way of a convolution operation in the transform domain. Specifically, let the filter responses be written $t_{\bar{u},\bar{x},\phi}$, where $\bar{u} = (L, \theta)$ specifies the level and orientation of the sub-band, $\bar{x}$ specifies location within the sub-band, and ϕ is phase. Then the pooled response can be written as a convolution with a pooling kernel *H*,

$$t^{q}_{\bar{u},\bar{x},\phi} * H_{\bar{u},\bar{x},\phi} \tag{1}$$

Because it is a convolution, it can be implemented by way of multiplication in the frequency domain. This is "circular" convolution, in which the borders of the two operands are implicitly connected at their edges (toroidal boundaries). While this is a natural assumption for the periodic orientation dimension, it will cause wrap-around errors in spatial and frequency dimensions unless buffer regions are inserted, as we have done. In the phase dimension, we have always assumed perfect summation, so the calculation reduces to a four-dimensional convolution. In the sampled case, the sub-band size varies with frequency, so that it must be expressed as a set of separate three-dimensional convolutions at each spatial frequency. In this sampled case we have also, for simplicity, assumed no pooling over spatial frequency.

As a further simplification, we have considered only separable Gaussian kernels, which may be specified entirely by width in orientation, space, and (in the un-sampled case) frequency. An example pooling kernel is shown in Figure 8. This example specifies almost complete pooling over orientation, but very little pooling over space. Because each row is a separate three-dimensional kernel for one spatial frequency, pooling over frequency is not represented. The three-dimensional Fourier transforms of these kernels, which are used in the convolution, are also shown.

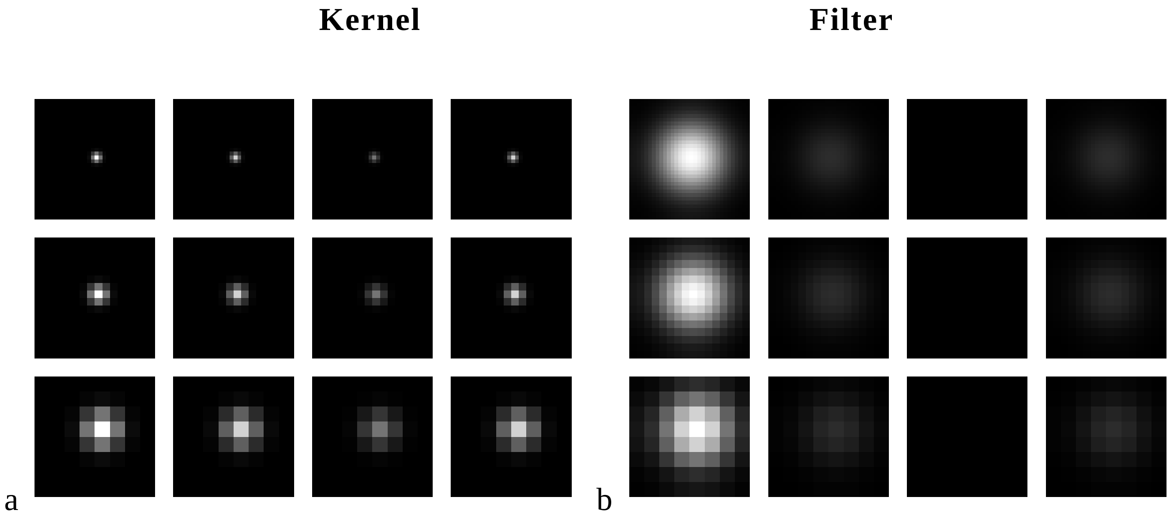

Figure 8. Example set of three inhibitory pooling kernels, one for each level (a), and their corresponding 3D Fourier transforms (b).

Foley’s model employs a combination rule in which filter responses to similarly oriented target and mask components are combined linearly before the nonlinearity, but responses to differently oriented target and mask components are combined after the nonlinearity.

The output of the inhibitory stage has the same dimensions as that of the excitatory stage, and is intended to represent the aggregate inhibitory signal that will control the gain of each neuron. This output is illustrated in Figure 9 for the two example stimuli. Note that the complete summation over phase has produced a de-modulated, all positive signal, and that the inhibition extends broadly over orientation but is confined largely to one band of spatial frequency.

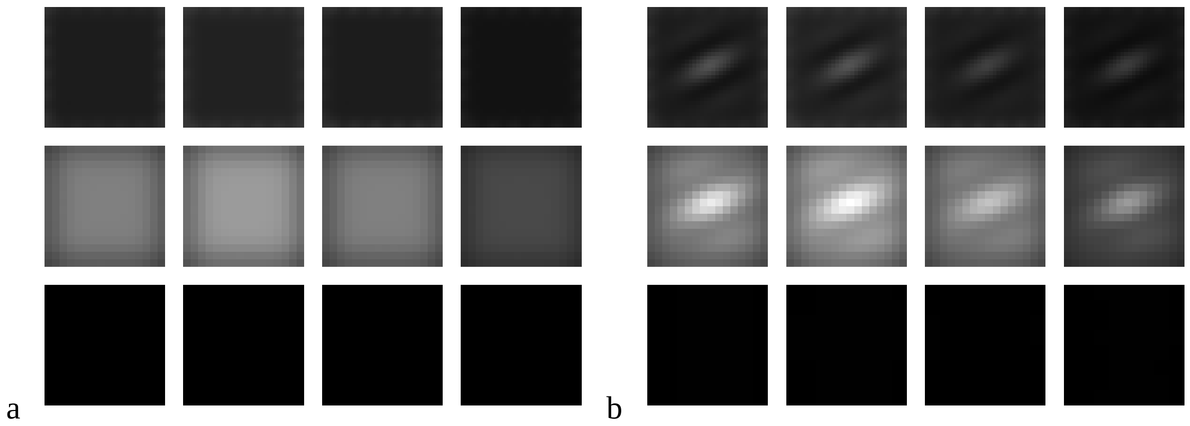

Figure 9. Response of inhibitory path to the mask-alone (a) and mask+target (b).

J. Divisive Gain Control

After pooling, the inhibitory path controls the gain of the excitatory path through a divisive operation,

$$r_{\bar{u},\bar{x},\phi} = \frac{t^{p}_{\bar{u},\bar{x},\phi}}{b^{q} + t^{q}_{\bar{u},\bar{x},\phi} * H_{\bar{u},\bar{x},\phi}} \qquad (2)$$

The gain-control expression contains a positive constant *b*, which defines the point at which saturation begins and also prevents division by zero. A similar divisive formulation is common to most models of contrast gain control[2, 3, 5, 8]. Some further comments on the parameterization of this expression are provided in the appendix.

This divisive operation is applied on a sample-by-sample basis. Figure 10 shows the results for our example stimuli.

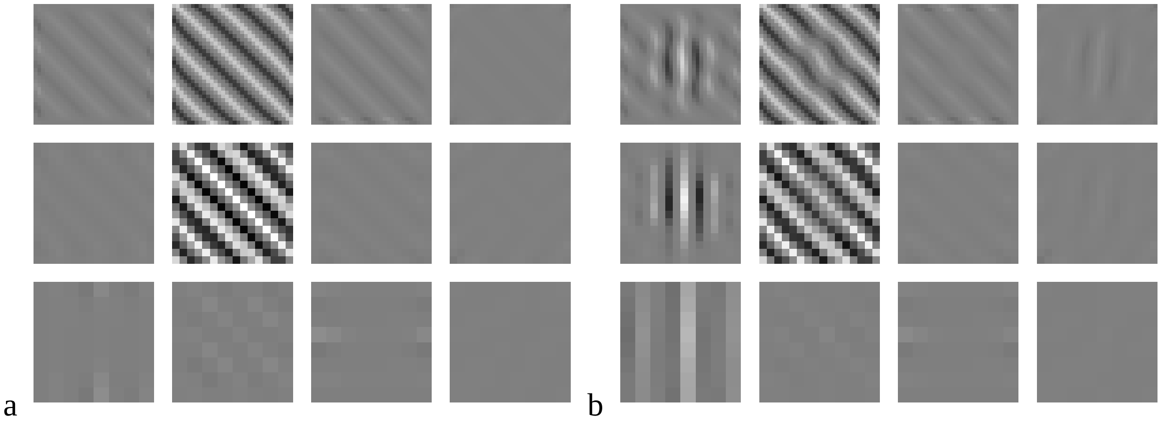

Figure 10. Normalized responses to mask-alone (a) and mask+target (b).

K. Comparison

At the comparison stage, the normalized responses to the two stimuli are subtracted. This step, which is common to most models of masking, is consistent with simple ideal observer theory, but is one of the steps most susceptible to criticism. It assumes, for example, that the observer has a photographic memory for the two images.

L. Decision

We adopt a simple probability summation rule at the decision stage. A Minkowski summation (Holder Norm) with exponent β is applied to the response differences,

$$d = \left[\sum \left| {}_1 r_{\bar{u},\bar{x},\phi} - {}_2 r_{\bar{u},\bar{x},\phi} \right|^{\beta} \right]^{1/\beta} . \tag{3}$$

The differences are assumed to be at threshold, and the images discriminated, when $d>1$. Figure 11 shows the real part of the differential response, after it is rectified and raised to the power β=4. This result is shown both for the example target Gabor of 50% contrast that has been used in the previous figures, but also for a target contrast of 17% approximately the threshold value found for this configuration of target and mask[5]. Though there are also sizable responses at other points, it is clear that the largest response is, as expected, at the center of the Gabor target. This is particularly true for the near-threshold responses in Figure 11b.

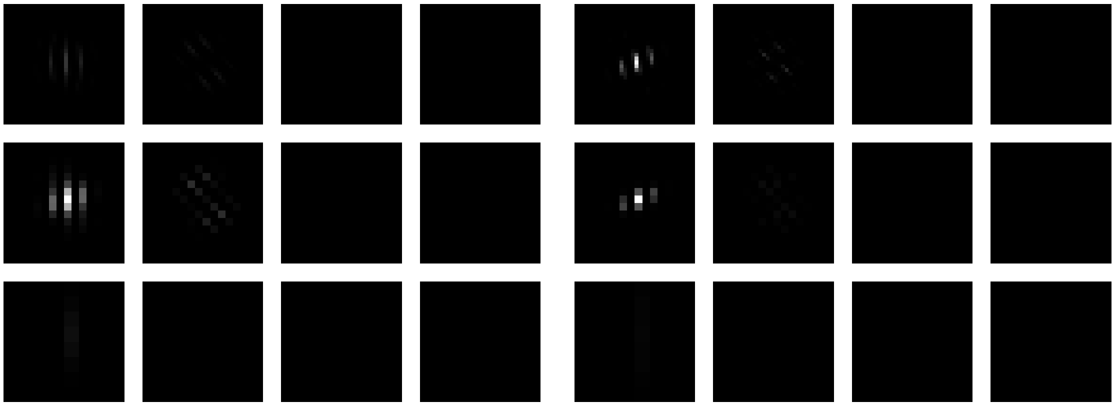

Figure 11. Differential response raised to the power β=4 for the example stimuli: a) target contrast = 50%, b) target contrast = 17%. For clarity, both images are displayed at full contrast; in a) the largest value is 195.3, in b) it is 1.3.

This decision rule has a number of possible interpretations. The first interpretation is that of probability summation among independent high-threshold mechanisms. Each neuron independently determines whether on a given trial its response in the two intervals differed by a criterion amount, and if this criterion is met in any neuron, the correct interval is selected. If the exponent is 2, then this rule is equivalent to an ideal observer of a signal known exactly, applied to the normalized responses.

M. Channel Filter Normalization

We have elected to place all of the variation in sensitivity with spatial frequency in a CSF filter that precedes the Gabor filter bank. In particular, we normalize each level of the Gabor filter array in such a way that the CSF filter specifies the approximate sensitivity to one-octave Gabor signals at each frequency of the array.

This normalization is accomplished in the following way. Consider the Gabor filter at u={f,0} (the orientation is not important, but we will assume that a filter at orientation 0 exists). We write the response of this filter to a unit amplitude one octave Gabor signal at frequency f as G_f . Then with no mask and assuming that the inhibitory signal is negligible, we have

$$d = b^{-q} \left[\sum \left| G_f \right|^{\beta} \right]^{1/\beta} \tag{4}$$

The set of filters at frequency f is then scaled by d^{-1}. In the absence of a CSF filter, the threshold amplitude for the Gabor would then be 1; in the presence of the CSF filter it will have a threshold equal to the inverse of the value of the CSF at the corresponding frequency. This approximation neglects the contribution of all channels but the one at the frequency and orientation of the Gabor, which in most cases will be minor.

3. SIMULATIONS

A. Computational Methods

The model described above was implemented in the Mathematica programming language[19].

The model contains a number of parameters that affect the size of the computation. Most importantly, these are *D*, the width in degrees of the simulated square area of visual field, and *F*, the spatial frequency in cycles/degree of the highest channel simulated. Together, these parameters specify the sampling density of the images. Thus *P*, the image width in samples, is usually equal to 4 *D F*. Most of the data we simulate here concern small targets that can be largely contained within an area of 2 degrees on a side. In addition, they all have spectra centered at 2 cycles/degree. Therefore most of the simulations we show are for *F*=4 cycles/degree, *D*=2 degrees. Thus the stimuli, and the response image in one sub-band at the highest frequency, are represented by images 32 pixels on a side. We have experimented with larger sizes, and found little change. Another important factor is *O*, the number of orientation channels simulated. A final factor is *C*, the number of spatial frequency channels simulated. Once the highest frequency is chosen, adding lower frequencies spaced an octave apart adds relatively little computational effort (about 30%) in the down-sampled case. In the down-sampled case, the total number of samples in the response array is approximately $2^6 \, O \, F^2 D^2 /3$. For example, with a field width of 2 degrees, three frequencies (1, 2, and 4 cycle/degree) and eight orientations, the number of complex response samples is 10,752. For reference, we provide in Table 1 a summary of model parameters, variables, and simulation control parameters.

Model Parameters	
a	CSF peak amplitude
f_0	CSF peak frequency
w	CSF log bandwidth
p	excitatory exponent
q	inhibitory exponent
b	saturation constant
s_x	pooling width in x or y
s_θ	pooling width in orientation
s_f	pooling width in frequency
k	octave bandwidth of Gabor filters
β	Minkowski exponent
Model Variables	
$t_{\bar{u},\bar{x},\phi}$	Gabor filter response
$r_{\bar{u},\bar{x},\phi}$	normalized response
d	decision variable, analogous to d’
Simulation Control Parameters	
O	number of orientations simulated
D	width in degrees of the square simulated area
P	width in pixels of the square simulated area
F	spatial frequency in cycles/degree of highest channel
C	number of octave-spaced frequency channels

Table 1. Model Notation.

B. Effect of Mask Orientation

Foley and Boynton[5, 20] collected data for detection of a 2 cycle/degree, 0° orientation Gabor target added to a 2 cycle/degree cosine mask at one of several orientations, or to a combined Gabor and cosine mask. These stimuli are illustrated in Figure 12.

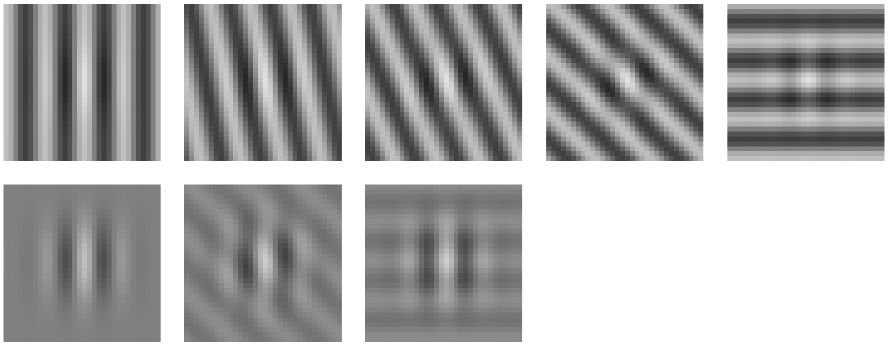

Figure 12. Stimuli of Foley and Boynton. The first row shows a Gabor target added to cosine masks at orientations of 0°, 11.25°, 22.5, 45° and 90°. Second row shows the same Gabor target added to an identical Gabor mask or a Gabor mask plus a cosine mask at 45° or 90°.

In the experiments depicted by the first row in Figure 12, Foley and Boynton varied the contrast of the cosine mask at each orientation over the range from -46 to -10 dB in steps of 4 dB. The first panel of the second row illustrates a contrast discrimination experiment in which a Gabor target was detected upon a Gabor mask, whose contrast was likewise varied over the range from -46 to -10 dB in steps of 4 dB. In the final two panels of the second row, the contrast discrimination experiment was repeated in the presence of an additional cosine mask with fixed contrast of -20 dB and an orientation of either 45° or 90°. For each experiment depicted by a single image in Figure 12, the resulting set of data set may be represented by a plot of target threshold vs. mask contrast, both in dB.

Foley and Boynton presented results from two observers, KMF and JYS. Each complete data set contains of 88 thresholds, consisting of ten masked thresholds and one absolute threshold from each of the panels in Figure 12. We have fit our model to all 88 points of each observer. All model parameters were optimized separately for each observer, except for $q = 2$, and $s_f = 0$. Simulation control parameters *were:* $O = 8$, $D = 2$ degrees, $P = 32$, $F = 4$ cycles/degree, $C = 3$.

The experimental data (points) and simulation results (curves) are shown in Figure 13 and Figure 14. Individual panels correspond to the separate experiments illustrated in Figure 12. There are systematic departures, but given the size and complexity of the data set, the fit is quite good. The RMS errors of the fits are 1.67 and 2.0 dB for KMF and JYS respectively. The optimized parameters are given in Table 2. Foley noted some key effects in the data: 1) the facilitation for like-oriented target and mask, 2) the diminished but persistent masking as mask orientation differs from the target, and 3) the persistence of facilitation in the presence of the fixed cosine mask. All of these effects, as well as many of the detailed gyrations of the data are captured by the model.

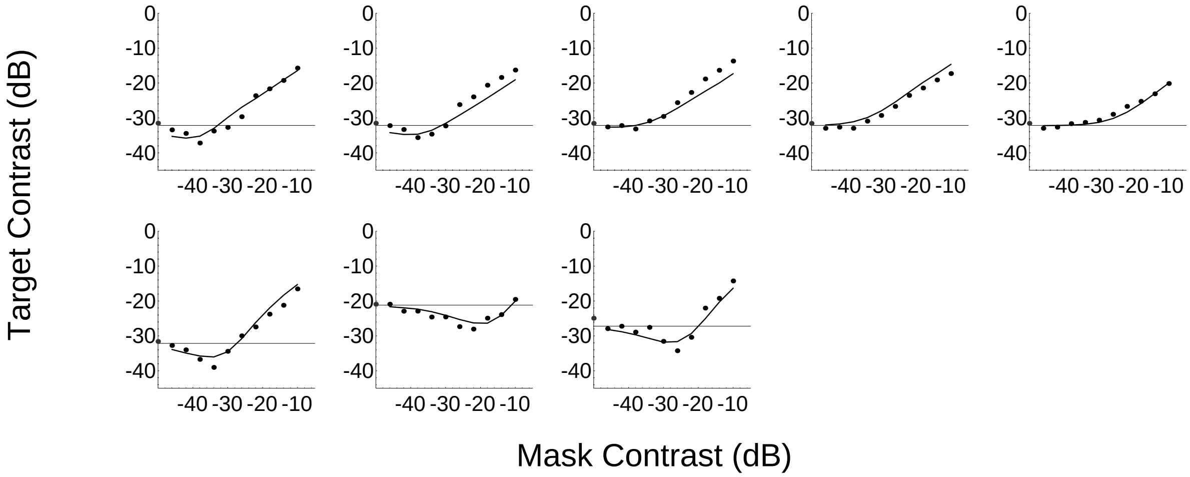

Figure 13. Data and simulations for observer KMF. In each panel contains an absolute threshold (no mask) plotted on the vertical axis; the corresponding model prediction is indicated by a horizontal line.

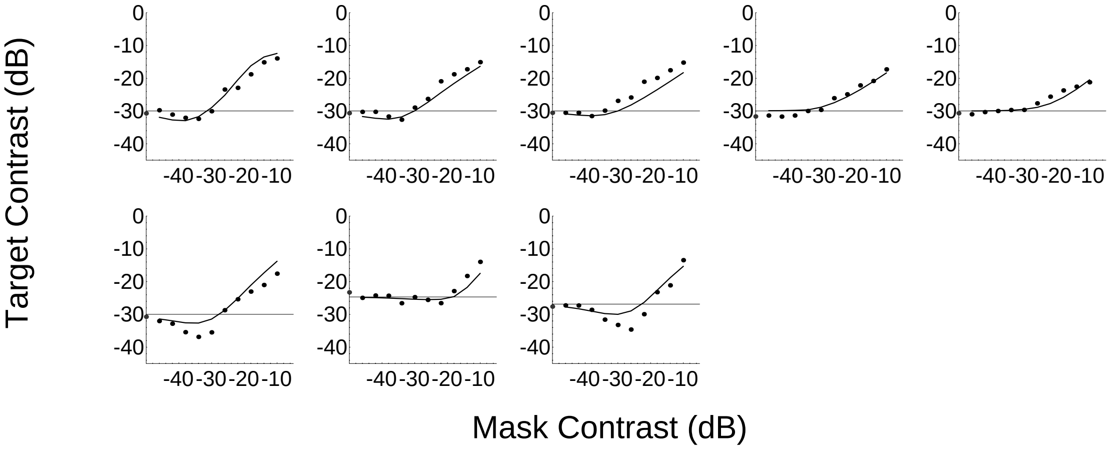

Figure 14. Data and simulations for observer JYS.

Parameter		Observer	
		KMF	JYS
CSF peak amplitude	a	39.58	42.88
CSF peak frequency	f_0	2.036	1.003
CSF $\log_{10}$ bandwidth	w	1.12	1.12
excitatory exponent	p	2.323	2.297
inhibitory exponent *	q	2.	2.
saturation constant	b	0.0203	0.0785
pooling width in x or y	s_x	1.55	0.53
pooling width in orientation	s_θ	88.74°	79.74°
pooling width in frequency *	s_f	0	0
octave bandwidth of Gabor filters	k	0.893	1.487
Minkowski exponent	β	5.414	4.87
rms error of fit		1.668	1.995

Table 2. Estimated model parameters and rms error for the two observers of Foley and Boynton. Parameters with asterisks were fixed.

C. Effect of Mask Phase

In another experiment, Foley and Boynton[20] compared the masking effects of cosine masks of 0° and 90° phase. In their model, as in our own, the inhibitory path sums over all phases, while the excitatory path does not. Thus facilitation, which depends upon the accelerating non-linearity in the excitatory path, should be phase dependent, while the masking should be less affected. Their data are shown in Figure 15. In fact, the data provide only modest support for the absence of facilitation, but the model, which shows no facilitation, does provides a good fit to the data.

The simulations in Figure 15 were produced by optimizing model parameters for these data, using the parameters from observer KMF as starting point. The initial values themselves provided a reasonable fit, but since the phase data were from a third observer (CCC), a further optimization was deemed appropriate. The final parameters were very similar to those for observer KMF listed in Table 2.

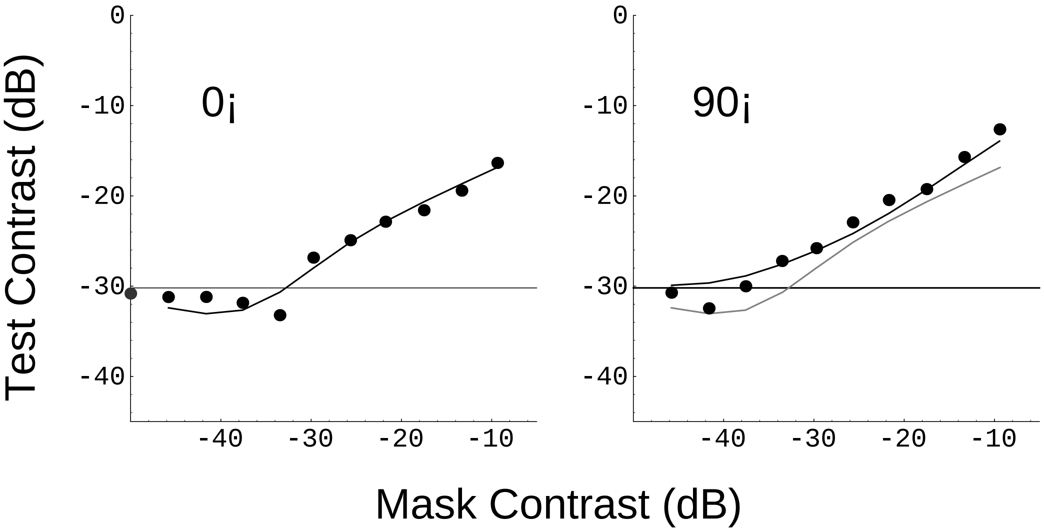

Figure 15. Effect of mask phase. Target was a 2 cycle/degree Gabor function; mask was a cosine at either 0° or 90° phase relative to the center of the Gabor. Curves are fits of the model. The 0° prediction is reproduced in light gray in the right hand panel for comparison.

4. DISCUSSION

We have constructed an image-based model of contrast gain control whose essential elements are, in order, 1) a contrast sensitivity filter, 2) linear, oriented, frequency-selective channels, 3) excitatory and inhibitory paths with different point non-linearities, 4) linear pooling within the inhibitory path over the dimensions of frequency, orientation, phase, and space, 5) division of the excitatory path by the inhibitory path, and 6) Minkowski pooling over the differential response to the pair of images to be discriminated. This model is similar in many respects to models developed by Foley[5] and Teo and Heeger[3].

The model provides a good fit to masking data in which the orientation of the mask is varied, in which contrast discrimination is accomplished in the presence of a fixed mask, and in which the phase of the mask is varied.

A. Quality of Fit

The rms error of the fit for the two observers is 1.67 and 2.0 dB. This compares to Foley's Model3 which he reports to yield rms fits of 1.23, and 1.48 dB. However, the latter model, despite its lesser generality, contains 12 estimated parameters, whereas ours contains only 9. (Indeed several of these parameters, notably f_0 and s_x , have little influence on the predictions.) Another relevant comparison is the rms error between the two observers: 2.14 dB. We propose here a "Prediction Turing Test," whereby we ask whether the predictions of the model could be identified as such when cast into a hat along with the data of several observers (to truly pass this test, some noise must be added to our predictions). In this case, the optimized fit of the model is in each case closer to the data than the data of the two observers are to one another.

B. Estimated Model Parameters

The estimated values of the excitatory pooling exponent *p*, 2.323 for KMF and 2.298 for JYS, are comparable to values of 2.55 and 2.72 estimated by Foley[5]. However, we fixed the value of the inhibitory exponent *q* at 2, while Foley allowed it to vary, estimating values of 2.18 and 2.32. Thus estimates of *p*-*q*, which largely determine the log-log slope of the masking data, are quite similar for the two models.

The estimated pooling width (Gaussian scale) in orientation is 89° for KMF and 80° for JYS. This very broad orientation pooling is consistent with estimates obtained by Foley and Boynton.

The estimates of spatial pooling width are 1.55 (KMF), and 0.53 (JYS), expressed in units of wavelength of the channel frequency, though the predictions are not very sensitive to this parameter. This suggests very localized pooling over space. Other recent experiments that directly address the spatial extent of inhibitory pooling, however, also suggest very local pooling[21, 22]. Estimates of spatial pooling of contrast gain that are based on apparent contrast, rather than detection thresholds, appear to yield much larger extents. For example, D'Zmura and Singer[23] estimate a Gaussian scale of 6.3 degrees for channels centered at 1.45 cycles/degree, while our scale at 2 cycles/degree is around 0.5 degree. It should be noted that Foley and Boynton's experiments, upon which we have based our simulations, were not designed to explore spatial pooling. This is clearly a subject for further investigation.

Because the experiments of Foley and Boynton that we simulate employ cosine or Gabor stimuli of 2 cycles/degree, they do not provide much evidence regarding the size of the pool in the frequency dimension. In our simulations we have therefore assumed no inhibitory pooling over frequency. Indeed, simulation results change very little if channels at 1 and 4 cycles/degree are omitted altogether.

C. Size and Resolution of Simulations

Predictions of psychophysical performance which rely upon simulations of arrays on neurons (so-called “neural images”), raise questions regarding the required density of spatial samples, and of the size of the portion of the visual field that is simulated. Often these questions are finessed by simulating a very large area at a very high density. Here we have used a more economical approach, simulating as small a region as possible (usually a 2 degree square) at as low a resolution as possible (usually 4 samples/cycle for each channel). It is comforting to know that even smaller simulations yield, with the same parameters, nearly identical results. For example, since the results in Figure 13 depend primarily upon the channel at 2 cycle/degree, one can obtain very similar predictions from an 8 x 8 pixel simulation covering a 1 degree square. We believe that, regardless of the computational environment, such intelligent economy will prove essential as researchers attempt to simulate additional dimensions (e.g. time) additional visual functions, and additional visual areas of the brain.

D. Pattern vs. Noise Masking

In this paper we have tried with some success to account for one set of masking effects with a contrast gain control model. Contrast gain control, however, is but one of several means by which masks may reduce the visibility of a target. For example, masking by intense dynamic white noise will certainly elevate thresholds, but the conventional interpretation of this elevation is that the noise adds variance to the decision variable. If the noise is static but white, a statistical interpretation is still likely to be proffered. As the noise bandwidth narrows, interpretations are more likely to be in terms of contrast gain control. It seems likely that many “pattern masking” situations will involve both deterministic contrast gain control and so-called “noise masking” effects. Some of the masking effects exhibited here may therefore be due to “noise masking.” This is another area in need of clarification.

E. Relation to other models

An important distinction between our model and that of Foley is the set of elementary component responses in which it is expressed. In our model, they are the putative responses of neurons. In Foley’s, they are responses to the particular stimuli employed in his experiments. The virtue of Foley’s scheme is that the number of quantities computed is on the order of the number of stimuli, while in our scheme it is on the order of the number of neurons, which is typically a much larger number. The virtue of our scheme is that it can compute the outcome for an arbitrary pair of stimuli, while Foley’s cannot generalize beyond the particular stimuli employed. A second virtue of our scheme is that it can embody specific assumptions about the internal machinery, for example an assertion that in the inhibitory path there is complete pooling between odd and even phase receptive field pairs, while such a concept does not even arise in a parameter-driven model.

The most prominent differences between our model and that of Teo and Heeger are the latter’s use of 1) multiple contrast channels, and 2) exponents of 2 in both excitatory and inhibitory paths. These two features are related. In general, use of an exponent of 2 in both paths leads to rapid saturation, so that single neurons cannot respond over the full range of contrast. In our model (and that of Foley), this is dealt with by having an excitatory exponent slightly larger than the inhibitory exponent. In Heeger’s model, it is handled by having several channels (usually four) to handle different parts of the total contrast range.

F. Recruitment

These two approaches can be reconciled through the mechanism of recruitment. Consider a set of rapidly saturating neurons, that are all identical save that their input gains vary over a broad range (like the multiple contrast channels in Heeger’s model). A recruitment mechanism which linearly sums their responses can behave very much like a single non-saturating neuron. Figure 16 shows as a thick solid line the example of summing the responses of four saturating mechanisms whose input gains vary in steps of 0.5 log unit (thin solid lines). For comparison, a single non-saturating neuron with a similar response is also shown as a dashed line. The sum clearly yields a non-saturating mechanism, though it is not identical to the single non-saturation neuron. Alternatively, the responses of a set of rapidly saturating neurons could be combined nonlinearly, say with a Minkowski exponent of 4, reflecting probability summation. We have not studied the mathematics of this problem in detail, but it is clear that if a larger number of saturating mechanisms is allowed, or if their output gains are allowed to vary, or if the exponent of the saturating mechanisms is allowed to vary, then an essentially perfect match can be obtained. Figure 16B shows an example with varying output gains. This means that our model simulations are consistent with an alternate recruitment model incorporating linear summation over multiple saturating mechanisms that vary in input gain.

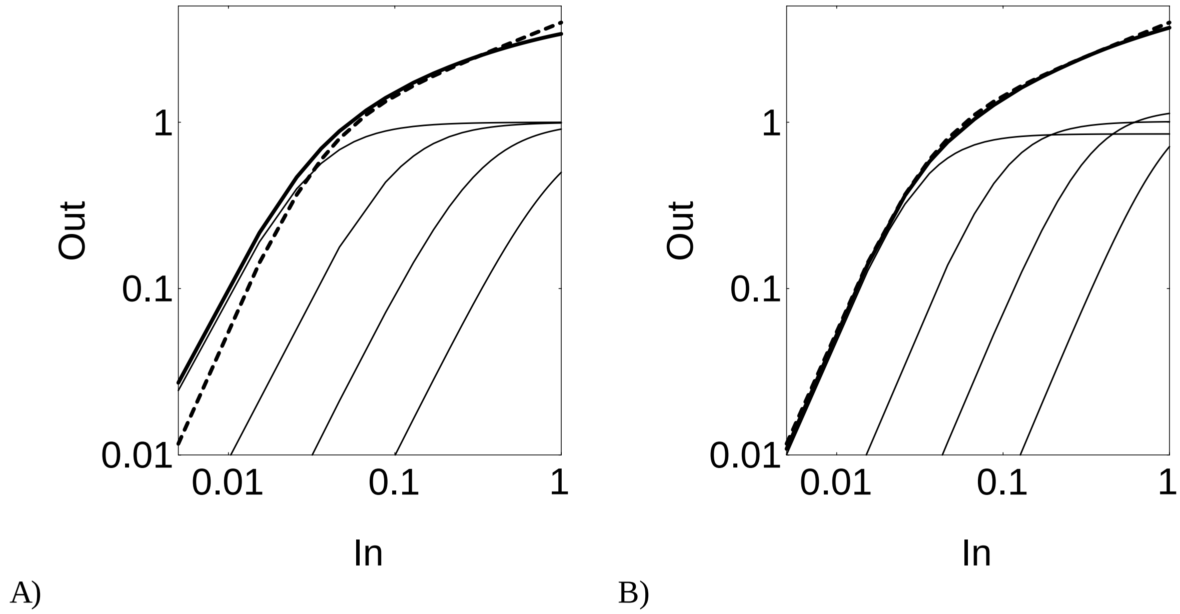

Figure 16. Thin lines show the responses of a set of saturating neurons with *p* = *q*, and input gains ranging between 0 and 1.5 log units in 0.5 log unit steps. The heavy solid line is the sum of these responses. The dashed line is the response of a single, non-saturating neuron with *p* = 2.4, *q* = 2, *a* = 1, *b* = 1, *g* = 0.0316. A) *p* = *q* = 2, *a* = 1, *b* = 1; B) *p* = *q* = 2.4, *a* = {0.850606, 1.01095, 1.20151, 1.428}, *b* = 1.

G. Effect of Duration

The data in Figure 13 and 15 are for targets and masks of 33 msec duration. More recently Foley has reported data for a duration of 100 msec[24], which show a reduced spread of masking over orientation. Such a change could be accommodated within our model by a change reduction in the parameter s_θ, but such variations should remind us that there are many detailed aspects of contrast gain control that have yet to be understood. Furthermore, in a practical application such as evaluation of still image quality, a longer duration is probably more appropriate to mimic typical viewer behavior.

H. Objections to the “Standard Model” of Masking

Recently, traditional models of masking, e.g. [2], which base performance on the differential response to mask and target+mask of a single neuron, have been questioned[25]. Nachmias noted that when discriminating a 10 cycle/degree cosine mask from that mask plus a 2 cycle/degree target, the overall contrast in each interval of the forced-choice trial could be independently perturbed, with little effect on performance. While the main target of Nachmias’ critique was the single neuron postulate, this observation also apparently causes problems for the model presented here. Adding different contrasts to the two intervals will generally increase our decision variable; indeed it will typically be non-zero even when target contrast is zero.

However, it should be understood that our’s is a model of discrimination, not identification. In effect, we are asking whether the two intervals are different, not what they look like. If in Nachmias’ experiment the observers were instructed to report whether the intervals were different, then the contrast perturbation would presumably have had a large effect.

But the fact that performance is little changed by this rather radical change in the task does suggest that the observer’s discrimination algorithm may be sufficiently sophisticated that it can operate under either set of conditions, and such sophisticated models may be required when more elaborate psychophysical tasks are explored.

5. SUMMARY

We have implemented an image-driven model of pattern discrimination that incorporates a contrast gain-control mechanism. The gain control is achieved by division of the excitatory signal from each neuron by an inhibitory signal that is a linear combination of responses of neurons within a neighborhood in space, frequency, orientation, and phase. The model is designed to allow simple adjustment of the size of this neighborhood. Excitatory and inhibitory paths are each subject to possibly different accelerating nonlinearities. We have found that this model, with optimized parameters, provides a good account of pattern masking data. In particular, the model accounts for the effect of cosine mask orientation on thresholds for a Gabor target, the effect of a combined mask consisting of both a Gabor and a cosine, and of the effect of cosine mask phase.

6. APPENDIX: PARAMETERIZATION OF NEURAL RESPONSE

A few brief observations on the parameterization of the basic transducer function may be useful, given the variety of schemes evident in the literature. We begin with an expression that includes parameters *c* (signal contrast) g_i (input gain of neuron *i* to signal), *a* (output gain), *b* (saturation constant), *p* and *q* (excitatory and inhibitory exponents), and w_i (weight of contribution to inhibitory pool of neuron *i*).

$$\frac{a\left(g_k c\right)^p}{\sum_i w_i \left(g_i c\right)^q + b^q} \tag{5}$$

If we let

$$v = \left[\sum_i w_i \left(g_i / g_k\right)^q\right]^{1/q} \tag{6}$$

For simplicity also write *g* for *gk* (excitatory input gain). Then the neural response simplifies to

$$\frac{a(gc)^p}{(gcv)^p + b^q} \tag{7}$$

The relationships among the parameters may be seen most easily when the transducer is plotted in log-log coordinates. There it is essentially two straight lines, with slopes of p and p-q. The corner between the two segments (the zero of the third derivative) is at *c = b/(g v)*. Figure 17 shows example responses for several values of *b*.

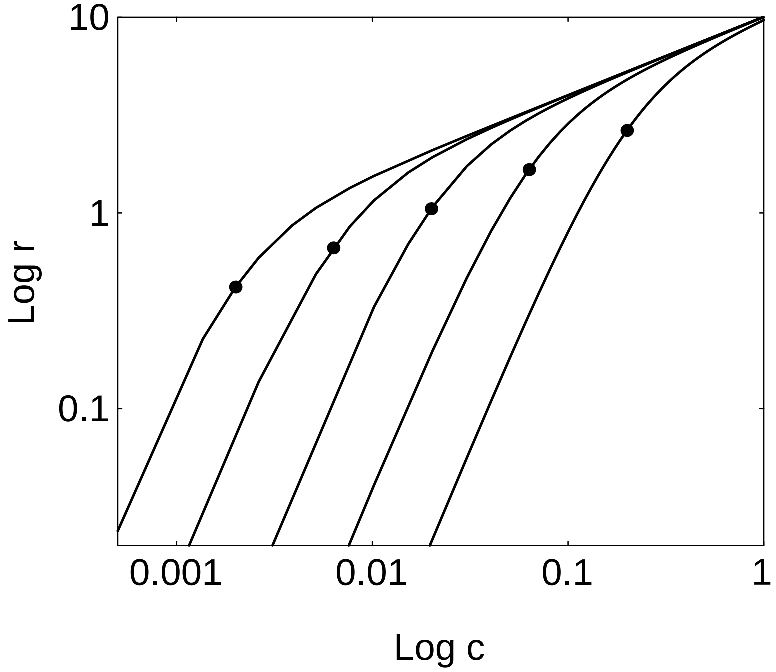

Figure 17. Neural response functions for several values of *b* (0.01, 0.0316, 0.1, 0.316, 1.). The corner of each curve, which occurs at *c= b/(gv)*, is plotted as a point. Other parameters are *v* = 0.5, *a* = 1, *g* = 10, *p* = 2.4, *q* = 2.

The actions of the three parameters *a*, *g*, and *b* on the log-log curve are as follows: *a* produces a purely vertical shift; *g* produces a purely horizontal shift, and *b* produces both a horizontal shift and a vertical shift. Because the curve does not change shape, but merely shifts horizontally and vertically. It is easy to see that there are only two degrees of freedom, and how the parameters may be interchanged.

For example, to eliminate *g*, we set

$$\begin{aligned} a' &= a\,g^{p-q} \\ b' &= b/g \end{aligned} \tag{8}$$

To eliminate *b* (i.e. b=1), we set

$$\begin{aligned} a' &= a\,b^{p-q} \\ g' &= g/b \end{aligned} \tag{9}$$

To eliminate *a*, we set

$$\begin{aligned} g' &= g\, a^{1/(p-q)} \\ b' &= b\, a^{1/(p-q)} \end{aligned} \tag{10}$$

Note that this last procedure will not work when *p* = *q*. In that case, the parameter *a* is required, but either *b* or *g* can be eliminated. In optimization of model parameters, it is important to eliminate these redundancies. We have usually selected a parameterization in which *a*=1, while *b* and *g* are free to vary.

7. ACKNOWLEDGMENTS

This work was supported by NASA Grant 199-06-12-39 to Andrew Watson, and a National Research Council Fellowship to J. Solomon. We thank Albert Ahumada and Miguel Eckstein for useful discussions.

8. REFERENCES

1. J. Nachmias and R. Sansbury. "Grating contrast: discrimination may be better than detection," Vision Research. **14** , 1039-1042 (1974).

2. G. E. Legge and J. M. Foley. "Contrast masking in human vision," Journal of the Optical Society of America. **70**(12), 1458-1471 (1980).

3. P. C. Teo and D. J. Heeger. "Perceptual image distortion," SPIE Proceedings. **2179**, 127-139 (1994).

4. H. R. Wilson and R. Humanski. "Spatial frequency adaptation and contrast gain control," Vision Research. **33**, 1133-1149 (1993).

5. J. M. Foley. "Human luminance pattern mechanisms: masking experiments require a new model," Journal of the Optical Society of America A. **11**(6), 1710-1719 (1994).

6. G. C. DeAngelis, J. G. Robson, I. Ohzawa and R. D. Freeman. "Organization of suppression in receptive fields of neurons in cat visual cortex," Journal of Neurophysiology. **68**(1), 144-163 (1992).

7. W. S. Geisler and D. G. Albrecht. "Cortical neurons: Isolation of contrast gain control," Vision Research. **32**(8), 1409-1410 (1992).

8. D. J. Heeger. "Normalization of cell responses in cat striate cortex," Visual Neuroscience. **9**, 181-198 (1992).

9. M. W. Cannon and S. C. Fullenkamp. "A transducer model for contrast perception," Vision Research. **31**, 983-998 (1991).

10. A. Ahumada, Jr. "Computational Image Quality Metrics: A Review," Society for Information Display International Symposium, Digest of Technical Papers. **24**, 305-308 (1993).

11. S. Daly. "The visible differences predictor: an algorithm for the assessment of image fidelity quality," in *Digital images and human vision*, A. B. Watson, ed. (MIT Press, Cambridge, MA, 1993).

12. J. Lubin. "The use of psychophysical data and models in the analysis of display system performance," in *Digital images and human vision*, A. B. Watson, ed. (MIT Press, Cambridge, MA, 1993).

13. J. W. Woods. *Subband image coding* (Kluwer Academic Publishers, Norwell, MA, 1991).

14. J. M. Valeton and A. B. Watson. "Contrast detection does not have a local spatial scale," Investigative Ophthalmology & Visual Science. **31**(4), 428 (1990).

15. A. B. Watson. "The cortex transform: Rapid computation of simulated neural images," Computer Vision, Graphics, and Image Processing. **39**(3), 311-327 (1987).

16. D. J. Heeger. "Half-squaring in responses of cat simple cells," Visual Neuroscience. **9**, 427-443 (1992).

17. A. B. Watson. "Efficiency of an image code based on human vision," Journal of the Optical Society of America A. **4**(12), 2401-2417 (1987).

18. A. B. Watson and A. J. Ahumada, Jr. "Model of human visual-motion sensing," Journal of the Optical Society of America A. **2**(2), 322-342 (1985).

19. S. Wolfram. *The Mathematica Book* (Wolfram Media/Cambridge University Press, Champaign, IL, 1996).

20. J. M. Foley and G. M. Boynton. "A new model of human luminance pattern vision mechanisms: Analysis of the effects of pattern orientation, spatial phase and temporal frequency," SPIE Proceedings. **2054**, (1994).

21. R. J. Snowden. "The effect of contrast surrounds on contrast centres: merely normal masking?," Investigative Ophthalmology & Visual Science. **36**(4 (Supplement)), S438 (1995).

22. J. A. Solomon and A. B. Watson. "Spatial and spatial frequency spreads of masking: measurements and a contrast gain-control model," Perception. **24**(Supplement), 37 (Abstract) (1995).

23. M. D'Zmura and B. Singer. "Spatial pooling of contrast gain control," Journal of the Optical Society of America A. **13**(11), 2135-2140 (1996).

24. J. M. Foley. "Simultaneous pattern masking: How come threshold elevation bandwidth decreases as stimulus duration increases?," Investigative Ophthalmology and Visual Science. **37**(3 Supplement), S912 (1996).

25. J. Nachmias. "Masked detection of gratings: the standard model revisited," Vision Research. **33**(10), 1359-1365 (1993).