Visual detection of spatial contrast patterns: Evaluation of five simple models.

Andrew B. Watson

NASA Ames Research Center, Moffett Field, CA 94035 USA

abwatson@mail.arc.nasa.gov

Abstract: The ModelFest Phase One dataset is a collection of luminance contrast thresholds for 43 two-dimensional monochromatic spatial patterns confined to an area of approximately two by two degrees. These data were collected by a collaboration among twelve laboratories, and were designed to provide a common database for calibration and testing of spatial vision models. Here I report fits of the ModelFest data with five models: Peak Contrast, Contrast Energy, Generalized Energy, a Gabor Channels model, and a Discrete Cosine Transform model. The Gabor Channels model provides the best fit, though the other, simpler models, with the exception of Peak Contrast, provide remarkably good fits as well. Though there are clear individual differences, regularities in the data suggest the possibility of constructing a standard observer for spatial vision.

OCIS codes: (330.1800) Contrast sensitivity; (330.4060) Modeling of vision

References and links

1. T. Carney, S. A. Klein, C. W. Tyler, A. D. Silverstein, B. Beutter, D. Levi, A. B. Watson, A. J. Reeves, A. M. Norcia, C.-C. Chen, W. Makous and M. P. Eckstein, "The development of an image/threshold database for designing and testing human vision models," Human Vision, Visual Processing, and Digital Display IX, Proc. SPIE, **3644**, 542-551 (1999).
2. A. B. Watson and J. A. Solomon, "ModelFest data: Fit of the Watson-Solomon model," Invest. Ophthalm. & Visual Science. **40**, S572 (1999).
3. A. B. Watson, "ModelFest Web Site, "http://vision.arc.nasa.gov/modelfest/, (1999).
4. T. Carney, "ModelFest Web Site, " http://www.neurometrics.com/projects/Modelfest/IndexModelfest.htm, (1999).
5. A. B. Watson, "Estimation of local spatial scale," J. Opt. Soc. Am. A. **4**, 1579-1582 (1987).
6. A. B. Watson, M. Taylor and R. Borthwick, "Image quality and entropy masking," Human Vision, Visual Processing, and Digital Display VIII, Proc. SPIE, **3016**, 2-12 (1997).
7. A. B. Watson, H. B. Barlow and J. G. Robson, "What does the eye see best?," Nature. **302**, 419-422 (1983).
8. K. R. K. Nielsen and B. A. Wandell, "Discrete analysis of spatial sensitivity models," J. Opt. Soc. Am. A. **5**, 743-755 (1988).
9. C. Morrone and D. Burr, "Feature detection in human vision: A phase-dependent energy model," Proc. Roy. Soc. **235**, 221-245 (1988).
10. J. Rovamo, O. Luntinen and R. Nasanen, "Modelling the dependence of contrast sensitivity on grating area and spatial frequency," Vision Res. **33**, 2773-88 (1993).
11. R. F. Quick, "A vector magnitude model of contrast detection," Kybernetik. **16**, 65-67 (1974).
12. H. Peterson, A. J. Ahumada, Jr. and A. Watson, "An Improved Detection Model for DCT Coefficient Quantization," Human Vision and Electronic Imaging, Proc. SPIE, **1913**, 191-201 (1993).
13. A. B. Watson, J. Hu, J. F. M. III and J. B. Mulligan, "Design and performance of a digital video quality metric," Human Vision, Visual Processing, and Digital Display IX, Proc. SPIE, **3644**, 168-174 (1999).
14. A. B. Watson, "DCT quantization matrices visually optimized for individual images," Human Vision, Visual Processing, and Digital Display IV, Proc. SPIE, **1913**, 202-216 (1993).
15. A. B. Watson, "DCTune Web Site, "http://vision.arc.nasa.gov/dctune/, (1999).
16. A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," J. Opt. Soc. Am. A. **14**, 2378 - 2390 (1997).
17. J. M. Foley, "Human luminance pattern mechanisms: masking experiments require a new model," J. Opt. Soc. Am. A. **11**, 1710-1719 (1994).

#14103 - $15.00 US Received October 14, 1999; Revised January 03, 2000

18. G. Sclar, J. H. R. Maunsell and P. Lennie, "Coding of image contrast in central visual pathways of the macaque monkey," Vision Res. **30**, 1-10 (1990).
19. A. J. Ahumada, Jr., "Simplified Vision Models for Image Quality Assessment," Society for Information Display International Symposium, SID Digest of Technical Papers, **27**, 397-400 (1996).
20. A. J. Ahumada, Jr. and B. L. Beard, "Image discrimination models predict detection in fixed but not random noise," J. Opt. Soc. Am. A. **14**, 2470-2475 (1997).
21. A. M. Rohaly, A. J. Ahumada, Jr. and A. B. Watson, "Object detection in natural backgrounds predicted by discrimination performance and models," Vision Res. **37**, 3225-3235 (1997).
22. G. Wyszecki and W. S. Stiles, *Color Science* (John Wiley and Sons, New York, 1982).
23. A. B. Watson and C. Ramirez, "A standard observer for spatial vision based on ModelFest data," Optical Society of America Annual Meeting, Digest of Technical Papers, **In press**, SuC6 (1999).
24. J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," Vision Res. **21**, 409-418 (1981).
25. E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D. J. Heeger, "Shiftable multi-scale transforms," IEEE Transactions on Information Theory, Special Issue on Wavelets. **38**, 587-607 (1992).

1. Introduction

ModelFest is the name of a series of workshops held at the annual meeting of the Optical Society of America whose purpose was to showcase and evaluate computational models of early human vision. More recently, ModelFest participants have collected a set of data designed to both calibrate and test vision models[1,2]. It was envisioned that the data set would be large and varied enough to adequately serve both purposes, and that the complete data set would be collected by a number of different labs, to enhance both generality and accuracy. The initial ModelFest data set consists of detection thresholds for static, achromatic patterns superimposed upon a uniform background and confined to a square area of about 2 by 2 degrees centered upon fixation. The selected stimuli consist of 43 patterns, including Gabors, Gaussians, lines, edges, multipoles, and various complex stimuli. Data were collected using standardized methods and display conditions. The complete dataset, as well as additional information, are available at several web sites[3,4]. In this report we describe fits of some simple models to the data of eight ModelFest observers. These fits provide a benchmark against which subsequent model fits may be compared.

Stimuli

Stimuli consisted of 43 monochrome images, each 256 x 256 pixels in size. The stimuli were selected by consensus of ModelFest participants. A complete list of ModelFest Phase 1 stimuli is given in Table 1. Each stimulus is identified by an index number between 1 and 43. One additional condition, similar to stimulus #35, but consisting of a new noise sample on each trial, is not considered here. Each stimulus was constructed as a set of real contrast pixels, and then scaled so that the mean (pixel contrast=0) mapped to 128 and the largest magnitude contrast mapped to 1 or 255.

Table 1. Stimulus details. Parameters sx and sy are horizontal and vertical Gaussian standard deviations; bx and by are half amplitude full bandwidths in horizontal and vertical dimensions.

Index	Type	Parameters
1	Gabor fixed size	1.12 c/d, sx=sy=0.5 deg
2	Gabor fixed size	2 c/d, sx=sy=0.5 deg
3	Gabor fixed size	2.83 c/d, sx=sy=0.5 deg
4	Gabor fixed size	4 c/d, sx=sy=0.5 deg
5	Gabor fixed size	5.66 c/d, sx=sy=0.5 deg
6	Gabor fixed size	8 c/d, sx=sy=0.5 deg
7	Gabor fixed size	11.3 c/d, sx=sy=0.5 deg
8	Gabor fixed size	16 c/d, sx=sy=0.5 deg

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

9	Gabor fixed size	22.6 c/d, sx=sy=0.5 deg
10	Gabor fixed size	30 c/d, sx=sy=0.5 deg
11	Gabor fixed cycles	2 c/d, bx=by=1 octave
12	Gabor fixed cycles	4 c/d, bx=by=1 octave
13	Gabor fixed cycles	8 c/d, bx=by=1 octave
14	Gabor fixed cycles	16 c/d, bx=by=1 octave
15	Elongated Gabor	4 c/d, sx=0.5 deg, by=0.5 octave
16	Elongated Gabor	8 c/d, sx=0.5 deg, by=0.5 octave
17	Elongated Gabor	16 c/d, sx=0.5 deg, by=0.5 octave
18	Elongated Gabor	4 c/d, bx=2 octave, by=1 octave
19	Elongated Gabor	4 c/d, sx =0.5 deg, by=1 octave
20	Elongated Gabor	4 c/d, bx=1 octave, by=2 octave
21	Elongated Gabor	4 c/d, bx=1 octave, sy=0.5 deg
22	Compound Gabor	2 & 2*sqrt2 c/d, sx=sy=0.5 deg
23	Compound Gabor	2 & 4 c/d, sx=sy=0.5 deg
24	Compound Gabor	4 & 4*sqrt2 c/d, sx=sy=0.5 deg
25	Compound Gabor	4 & 8 c/d, sx=sy=0.5 deg
26	Gaussian	sx=sy=30 min
27	Gaussian	sx=sy=8.43 min
28	Gaussian	sx=sy=2.106 min
29	Gaussian	sx=sy=1.05 min
30	Edge	sx=sy=0.5 deg
31	Line	0.5 min wide line, sx=sy=0.5 deg
32	Dipole	3 pixels wide, sx=sy=0.5 deg
33	5 Collinear Gabors	8 c/d, in phase, bx=by=1 octave, separation = 5 sx
34	5 Collinear Gabors	8 c/d, out of phase, bx=by=1 octave, separation = 5 sx
35	Binary noise	1 min pixels, sx=sy=0.5 deg
36	Oriented Gabor	4 c/d, 45 deg, bx=by=1 octave
37	Oriented Gabor	4 c/d, 0 deg, bx=by=1 octave
38	Gabor Plaid	4 c/d, 0 + 90 deg, bx=by=1 octave
39	Gabor Plaid	4 c/d, 45 + 90 deg, bx=by=1 octave
40	Disk	1/4 deg diameter
41	Bessel x Gaussian	4 c/d, sx=sy=0.5 deg
42	Checkerboard	4 c/d fundamental, sx=sy=0.5 deg
43	Natural image	San Francisco, sx=sy=0.5 deg

Original stimuli were represented as grayscale images with gray-levels between 1 and 255. When presented at contrast *c* on a background of luminance L_0, each gray-level *g* is mapped to luminance *L* according to the function

$$L(g) = L_0\left(1 + \frac{c}{127}(g - 128)\right)$$

(1

The contrast varied as a Gaussian function of time, with a standard deviation of 0.125 seconds. The precise means by which the image of a given contrast was rendered was left to the discretion of the individual labs [3,4]. Figure 1 shows the complete set of spatial stimuli. Figure 2 shows one stimulus as a QuickTime movie.

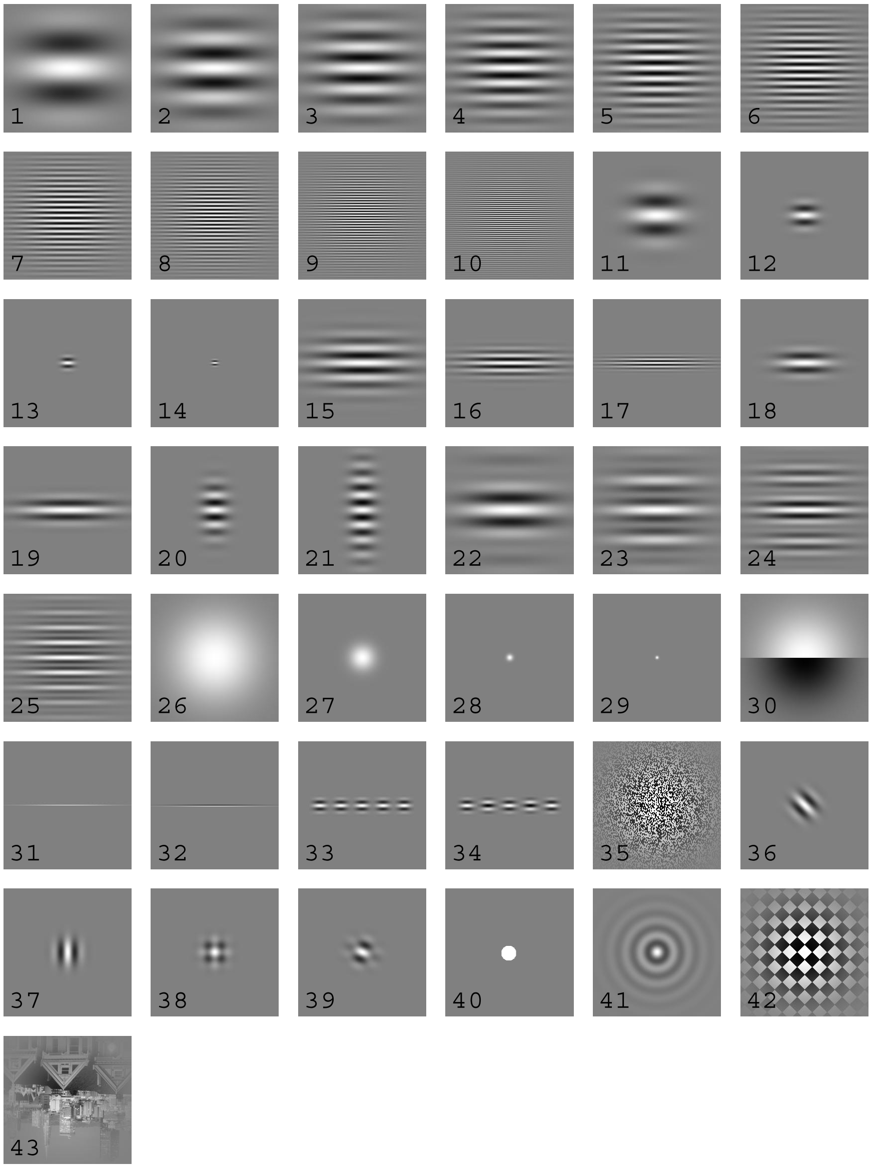

Figure 1. ModelFest stimuli.

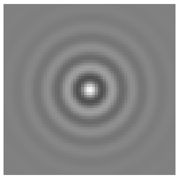

Figure 2. Movie of ModelFest stimulus #41 (240Kbytes).

Received October 14, 1999; Revised January 03, 2000

Methods

Thresholds were measured using a two-interval forced choice procedure. Feedback was provided. Each threshold was based on at least 32 trials, and each threshold was measured four times. Fixation guides, which were continuously present, consisted of four "L" shaped corner marks at the four corners of the 256x256 pixel stimulus field. The area of the screen outside the stimulus was kept at the mean luminance level. Additional details are provided in Table 2, and at the ModelFest website[3,4].

Table 2. Details of ModelFest display.

Mean luminance	30 +- 5 cd/m^2
Frame rate	>= 60 Hz
Display pixel size	1/120 degree
Image size	256 pixels square (2.13 degree)
Grayscale resolution	< 0.25 threshold
Viewing	Binocular with natural pupils

Results

Descriptive Statistics

Here we define some simple descriptive statistics that will be used to describe the data and fits. Each threshold (in dB) may be written $x_{i,j,k}$, where the indices refer to observer (k=1,…,K), stimulus (j=1,…,J), and replication (i=1,…,I). The model predictions may be written $p_{j,k,m}$ where m indexes the model. We then define three measures of error:

$$V_{j,k} = \frac{1}{I}\sum_{i=1}^{I}\left(x_{i,j,k} - \bar{x}_{j,k}\right)^2 \qquad \text{(2}$$

$$E_{j,k,m} = \left(p_{j,k,m} - \bar{x}_{j,k}\right)^2 \qquad \text{(3}$$

$$S_{j,k,m} = \frac{1}{I}\sum_{i=1}^{I}\left(x_{i,j,k} - p_{j,k,m}\right)^2 = E_{j,k,m} + V_{j,k} \qquad \text{(4}$$

The first quantity is the Maximum Likelihood estimate of the variance of each threshold estimate. The second is the squared error between the model prediction and the mean threshold. The last quantity is the average squared error between the individual thresholds and the corresponding model predictions. This is the maximum likelihood estimate of the variance of each threshold for model m.

We also define similar quantities that are averages over the stimulus subscript j :

$$V_k = \frac{1}{J}\sum_{j=1}^{J} V_{j,k} \qquad \text{(5}$$

$$E_{k,m} = \frac{1}{J}\sum_{j=1}^{J} E_{j,k,m} \qquad \text{(6}$$

$$S_{k,m} = \frac{1}{J}\sum_{j=1}^{J} S_{j,k,m} = E_{k,m} + V_k \qquad \text{(7}$$

We note that $E_{k,m}$ is the square of the RMS error between a model and the average thresholds, while V_k and $S_{k,m}$ are the estimates of variance for unconstrained and constrained models, respectively, assuming homogeneity of variance. The unconstrained model is that in which the "prediction" for each stimulus is given by the empirical mean threshold for that stimulus.

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

Mean Observer Thresholds in dB

Since the goal of this report is to describe overall fits of several models to the entire dataset, we do not separate the stimuli into various subsets according to type but rather present all thresholds together, ordered by index number. In Figure 3 we plot the mean thresholds in decibels (dB = 20 Log_{10}c) for each stimulus and observer. Error bars indicate plus and minus one standard deviation. A small version of each stimulus, slightly elongated in the vertical direction, is pictured at the top of the figure.

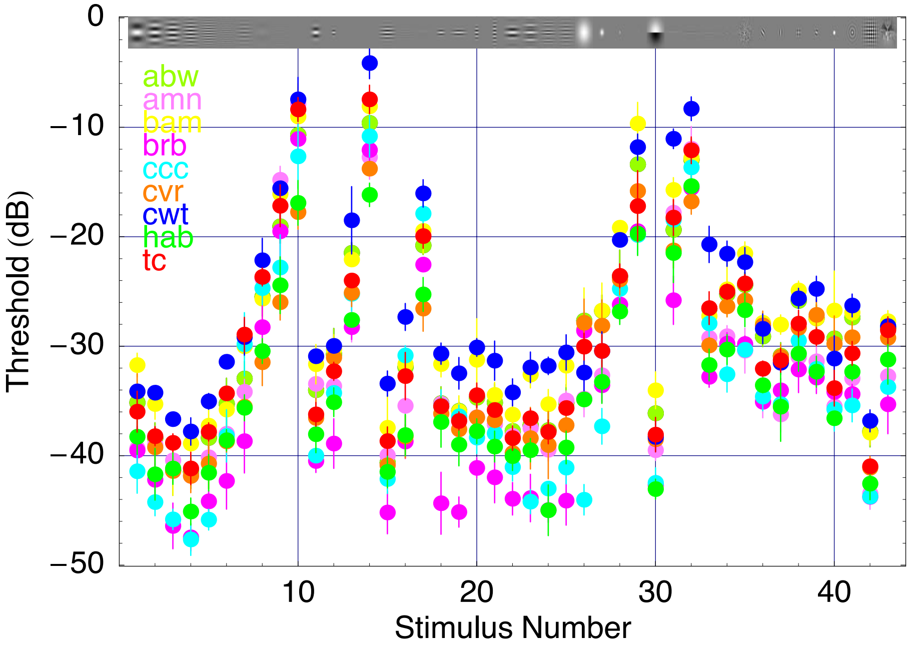

Figure 3. Threshold versus stimulus number for each observer.

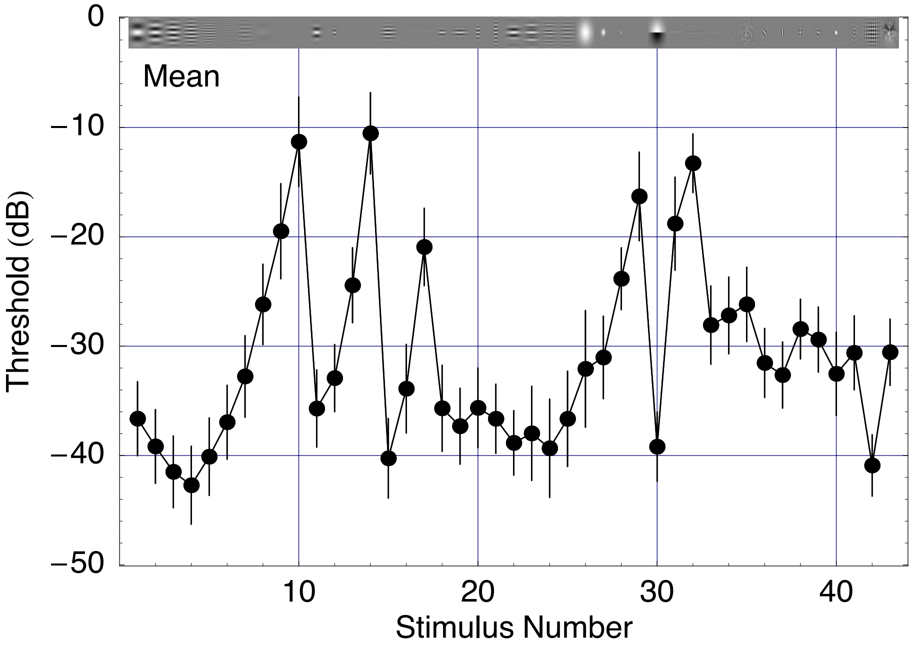

Figure 4. Mean thresholds versus stimulus index.

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

The mean threshold for the eight observers, is shown in Figure 4. The mean thresholds range between -44 and -10 dB. The first ten stimuli, which correspond to Gabor functions of fixed size with frequencies varying in steps of approximately half an octave, yield thresholds that depict a conventional contrast sensitivity function, and resemble comparable data in the literature[5]. Figure 5 shows the mean within-observer ($\frac{1}{K}\sum_{k}\sqrt{V_{j,k}}$) and overall standard deviations ($\sqrt{\frac{1}{IK}\sum_{i}\sum_{k}\left(x_{i,j,k}-\bar{x}_{j}\right)^{2}}$) as a function of stimulus index.

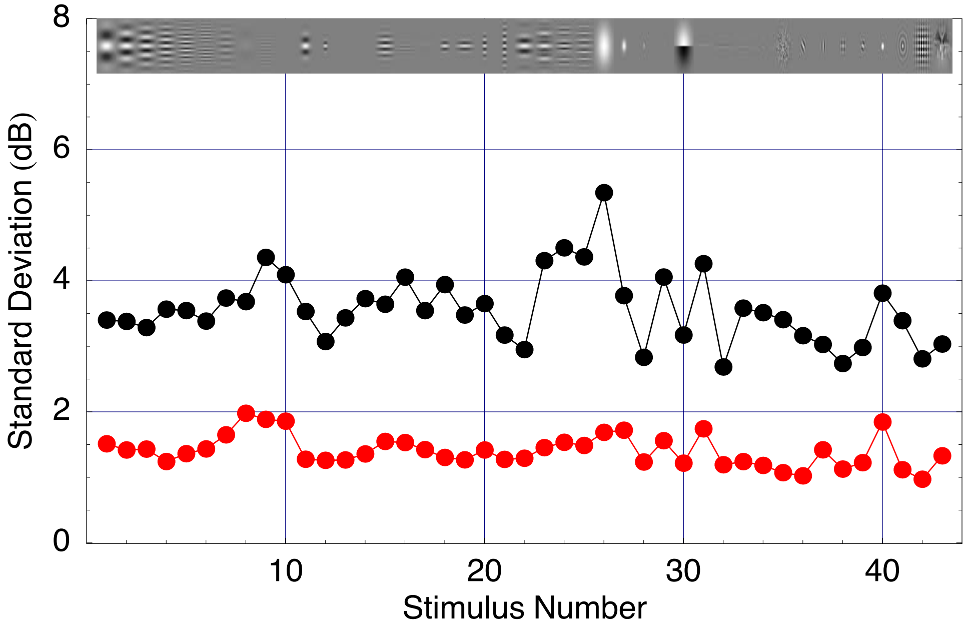

Figure 5. Threshold variability versus stimulus number. Red points indicate mean within-observer standard deviation; black points are standard deviation across observers.

Contrast Energy and Barlow Units

Elsewhere[6] we have defined and advocated the use of a unit of stimulus strength which takes into account the spatial and temporal extent of the stimulus, not merely its peak intensity. This unit, the Barlow, is defined as the contrast energy of a stimulus times 10^{-6}. Contrast energy is defined as the integral over space and time of the square of the contrast waveform of a stimulus. The contrast waveform is the luminance waveform, minus a defined mean luminance, and divided by that mean luminance. The factor of 10^{-6} is introduced so that the stimulus seen best by human observers (with least contrast energy) has an intensity of about 1 Barlow[7]. One virtue of the Barlow unit is that it is proportional to detection efficiency, in an ideal observer sense. We have also introduced a logarithmic version of the Barlow, called the deciBarlow, abbreviated dBB, which is defined as dBB = 20 Log_{10}(Barlow).

In Figure 6 we show the mean observer thresholds expressed as dBB, and in Figure 7 we show the mean over observers. It can be seen that the mean thresholds for some stimuli approach the value of 0 dBB. These most efficiently detected stimuli are typically small targets such as Gabor functions consisting of a few cycles of about 4 cycles/degree.

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

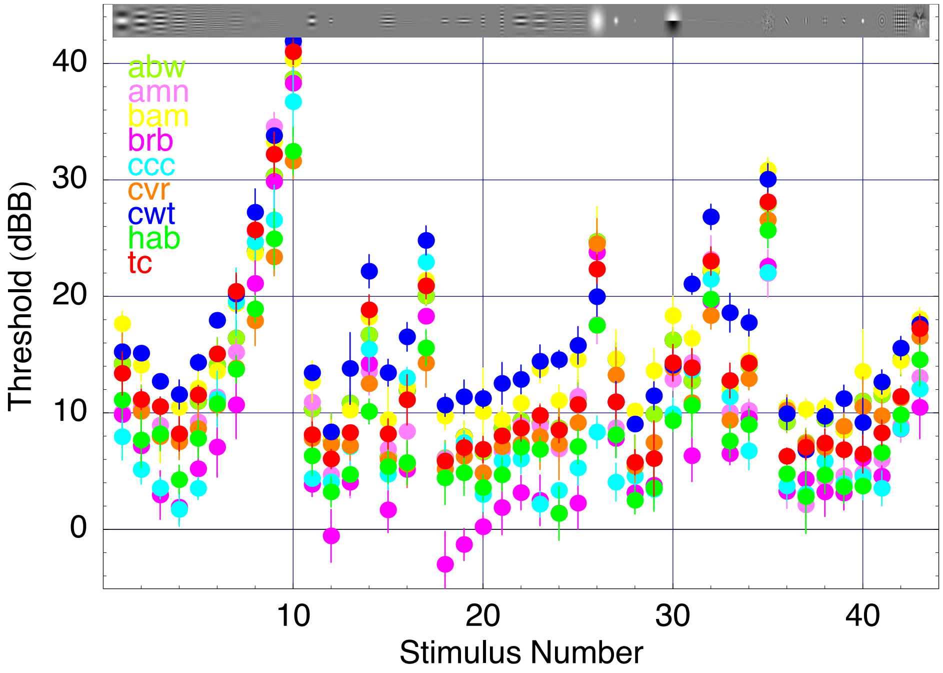

Figure 6. Mean and standard deviation of observer thresholds in dBB.

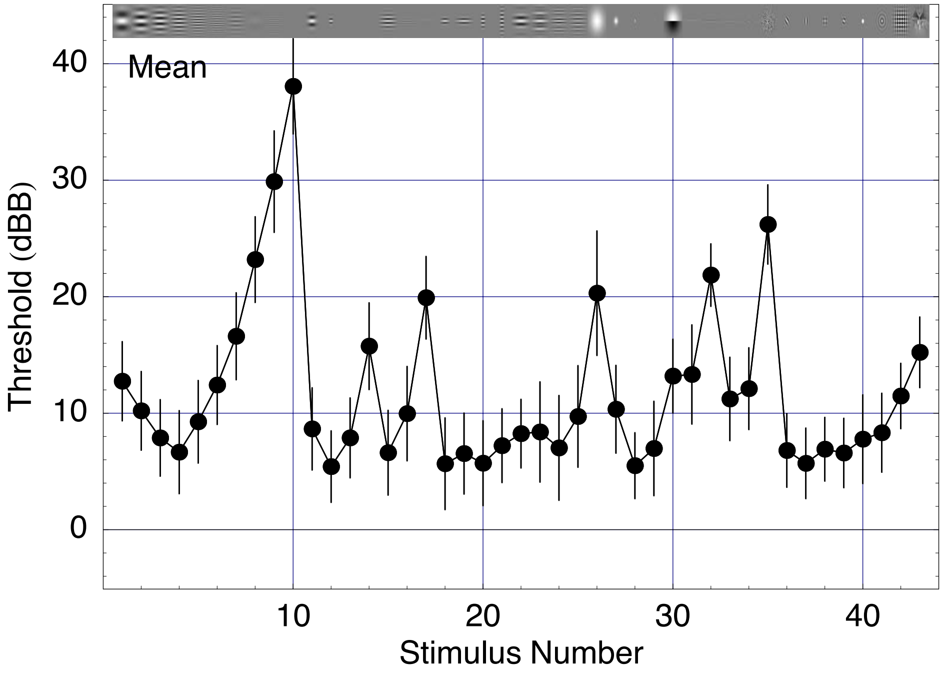

Figure 7. Mean and standard deviation of threshold for each stimulus in dBB.

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

Models: General Structure

In this report we consider five models: Peak Contrast (PC), Contrast Energy (CE), Generalized Energy (GE), Gabor Channels (GC), and Discrete Cosine Transform (DCT). In the case of the DCT model, we consider three variants with block sizes of 8, 16, and 32 pixels.

All models considered consist of four general stages: conversion from luminance to contrast, spatial filtering by a contrast sensitivity function (CSF), a linear (channel) transform, and pooling of transform coefficients to yield a single number that is assumed to be constant at detection threshold. We first consider those stages common to all models.

Conversion to contrast

The convention of Equation 1) shows how to convert the gray-level of each pixel to luminance, given a mean luminance and a contrast. We define the contrast of a pixel to be its luminance, less the mean luminance, divided by that mean. Thus for each pixel, conversion to contrast is achieved by subtracting the nominal mean of 128 , and dividing by 127.

Spatial Contrast Sensitivity Function Filter

In each of the models, the spatial filter serves to control sensitivity to various spatial frequencies, and is thus analogous to a contrast sensitivity function (CSF). The same type of spatial filter was used in each of the three models, though its parameters were allowed to differ from model to model. Because we are at this point indifferent to the particular form of the filter, we have a used a form which adheres closely to the data itself. This is a filter constructed in one dimension by linear interpolation between sample values in a linear-frequency, log-gain space. The frequency coordinates of the sample values were the spatial frequencies of the fixed-size Gabor functions (stimuli #1-10), plus frequencies of 0 and 120: 0, 1.12, 2, 2.83, 4, 5.66, 8, 11.3, 16, 22.6, 30, 120 cycles/degree. The gain values were set initially to the inverse of the corresponding thresholds for an observer. During the optimization, the values were allowed to vary freely from that starting point. In two dimensions, the filter is obtained as a surface of revolution of the one-dimensional filter. We call this type of filter the *interpolation filter*.

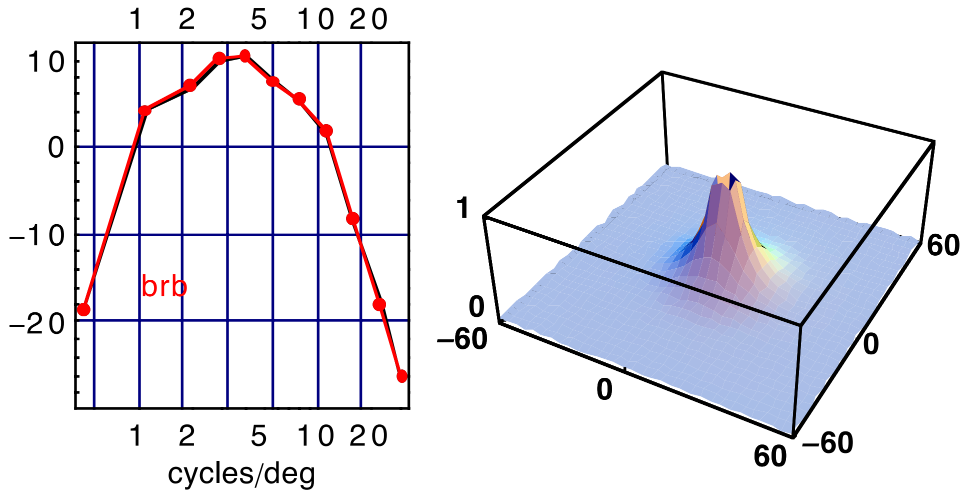

Figure 8. Interpolation filter example. A) One-dimensional filter. B) Two dimensional filter.

Received October 14, 1999; Revised January 03, 2000

Linear Channel Transform

For the Peak Contrast, Contrast Energy and Generalized Energy models, the channel transform was an identity transform, that is, no transform was performed and the transform coefficients are the filtered contrast pixels. For the DCT models, the channel transform was the blocked Discrete Cosine Transform, with a block size of 8, 16, or 32 pixels. For the Gabor Channel model, the channel transform consisted of a bank of linear channel filters varying in frequency and orientation. Both DCT and GC linear transforms are described in greater detail below.

Pooling

The final step in each model is the pooling of all transform coefficients using a Minkowski metric,

$$R = \left[\sum_i \left| r_i \right|^\beta \right]^{1/\beta} \qquad (8$$

where r_i are the individual coefficients and β is the pooling exponent. To compute contrast thresholds for individual stimuli, we compute R for a unit contrast stimulus, and then compute threshold contrast as the contrast that would yield a value of R=1, namely 1/R. The value of R=1 is arbitrary, because model responses are in arbitrary units.

For the Peak Contrast model, the pooling operation consists of selecting the single pixel with the largest absolute value. This is equivalent to a Minkowski metric with $\beta = \infty$. For the Energy model, $\beta = 2$. For the other models, β was a parameter estimated by optimizing the fit to the data. In general, the Minkowski exponent controls the *efficiency* of summation over transform coefficients. For example, complete (linear) summation is achieved with β=1, while β=∞ corresponds to no summation at all.

Models: Details

Peak Contrast

The Peak Contrast model consists of conversion to contrast, spatial filtering, and selection of the single pixel with the largest absolute value. We include this model primarily to demonstrate how poorly it performs, though it is has occasionally been entertained as a model of visual sensitivity. For this model, the free parameters are the eleven gain values of the interpolation filter.

Contrast Energy

The Contrast Energy model consists of conversion to contrast, spatial filtering, and pooling by squaring ($\beta = 2$) and summation over image pixels. The contrast energy model is motivated in part by its status as an ideal observer in the event that detection is limited by internal noise. In addition, energy models have been widely employed in human vision[7-10]. For this model, the free parameters are the eleven gain values of the interpolation filter.

Generalized Energy

The Generalized Energy model consists of conversion to contrast, spatial filtering, and Minkowski pooling with an arbitrary exponent β. It is identical to the Contrast Energy model, except that the pooling exponent is free to vary instead of being fixed at 2. In vision theory, Minkowski pooling has often been interpreted as a consequence of probability summation[11]. The generalized energy model may also be interpreted as an ideal observer acting upon coefficients after a point non-linearity. For this model, the free parameters are the eleven gain values of the interpolation filter, and the exponent β.

Received October 14, 1999; Revised January 03, 2000

Discrete Cosine Transform

This model employs the Discrete Cosine Transform (DCT) at the linear transform stage. The DCT is a Fourier-like transform that is widely used in image and video compression. It has also been used as a model of spatial transformations in early human vision[12-15]. In such models, it is typically adopted because in addition to transforming images into a hybrid space-frequency representation, it is also a very simple transform, for which fast algorithms are known, and which has in addition the properties of orthogonality, invertibility, and energy preservation. In the DCT model, the linear transform is followed by Minkowski pooling with exponent β estimated from the data. For this model, the free parameters are the eleven gain values of the interpolation filter and the exponent β. The block size may be considered a thirteenth parameter, although we consider only three values (8, 16, 32).

Gabor Channels

For the Gabor Channels model, the linear transform was an array of Gabor filters[16]. The details of the filters are given in Table 3. Pyramid sampling means that each output image from each channel was down-sampled in both dimensions to a resolution of twice the channel frequency. For a given value of β, the channel gains were adjusted so that the ensemble had an approximately flat contrast sensitivity function, so that all variation in sensitivity with spatial frequency is done by the interpolation filter. Among other advantages, this allows the parameters of the interpolation filter to be initialized to the inverse thresholds for the first ten Gabor stimuli (#1-10), and allows the recovered Interpolation filter to be regarded as the CSF of the systems as a whole.

Table 3. Gabor Channels model parameters.

Number of frequencies	11
Number of orientations	4
Number of phases	2 (odd and even)
Bandwidth	1.4 octaves
Highest frequency	30 cycles/degree
Lowest frequency	0.9375 cycles/degree
Frequency spacing	1/2 octave
Orientation spacing	45 degrees
Pyramid sampling	yes

In the Gabor Channels model, the linear transform is followed by Minkowski pooling with exponent β estimated from the data. For this model, the free parameters are the eleven gain values of the interpolation filter, and the exponent β.

Model Fits

Each of the models was fit separately to the data for each observer. Parameters were optimized so as to minimize $E_{k,m}$ (or equivalently, $S_{k,m}$). For each fit, we indicate in Table 4 and Figure 9 the residual RMS error in dB,

$$RMS_{k,m} = \sqrt{\tfrac{1}{J}\sum_{j=1}^{J}\left(\bar{x}_{j,k} - p_{j,k,m}\right)^2} = \sqrt{E_{k,m}}\,. \qquad (9$$

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

Table 4. RMS error for each model and observer.

	abw	amn	bam	brb	ccc	cvr	cwt	hab	tc	All
PC	5.22	3.91	5.34	5.56	5.14	5.51	4.73	4.71	4.38	4.97
CE	2.22	2.38	2.15	2.39	2.11	2.43	2.16	2.69	2.45	2.34
GE	1.85	1.08	1.85	2.24	1.83	2.13	1.46	1.91	1.50	1.79
GC	1.46	1.44	1.43	1.98	2.11	1.70	1.46	1.57	1.35	1.63
DCT8	1.86	1.05	1.77	2.24	1.85	2.11	1.38	1.84	1.50	1.77
DCT16	1.77	1.40	1.60	2.06	1.95	2.00	1.22	1.77	1.52	1.72
DCT32	1.90	2.26	1.71	2.00	2.09	2.03	1.78	2.34	2.16	2.04

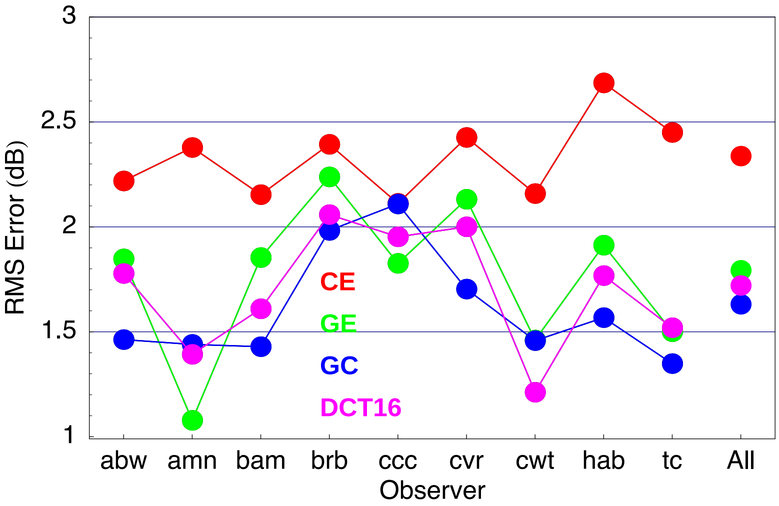

Figure 9. RMS error for four models and nine observers, and the group.

We show our best individual fit in Figure 10.

Received October 14, 1999; Revised January 03, 2000

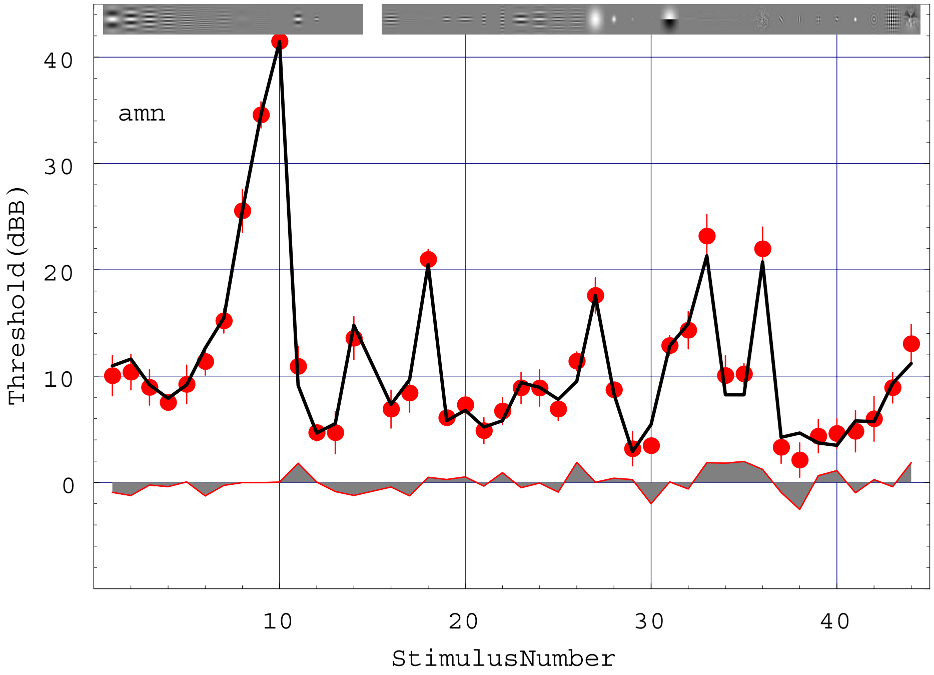

Figure 10. Fit of the DCT8 model to the data of observer amn. The red points and error bars show means and standard deviations from four replications. The model prediction is shown by the black line. The filled trace at the bottom shows the prediction error. The RMS error is 1.06 dB. This is the best fit among all models and observers.

Contrast Sensitivity Parameters

All of the models made use of the same Interpolation filter, and the parameters of this filter were estimated in the fitting procedure. Figure 11 shows the estimated parameters from the eight observers for the Contrast Energy model. Each parameter is a gain at a particular spatial frequency, and is plotted at the that frequency (except for the frequency 0 cycles/degree, which is plotted at 0.5 cycles/degree). The other models yielded results similar in form.). The point for observer ccc at the lowest spatial frequency is clearly anomalous. This observer was very sensitive to the large Gaussian stimuli (see Figure 3 and Figure 6), for unknown reasons.

We are interested in analytic formulae for this filter. The heavy black line shows the best fitting version of a parabola in the log-sensitivity, log frequency space. This function, which has been used previously in applied contexts[12] appears to be a reasonable fit to the filter parameters. A convenient form for the parabola is $0.275 - 1.536\ (x - 0.472)^2$ which shows that it peaks at 2.97 cycles/degree (0.472 log cycles/degree

Received October 14, 1999; Revised January 03, 2000

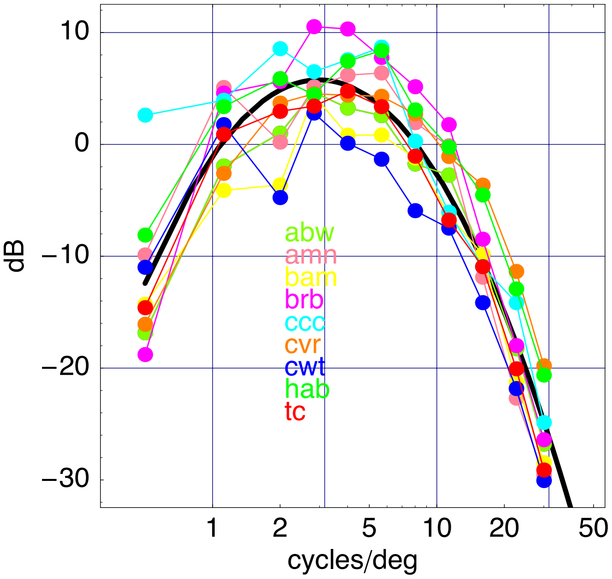

Figure 11. Interpolation filter parameters for the Contrast Energy model.

Pooling Exponents

In GE, GC, and DCT models the exponent β is free to vary. In Figure 12 we show the estimates of β for each observer and the mean. The mean estimates lie between 2.5 and 4, within the range expected from probability summation. They are also consistent with values assumed in non-linear transducer models[16,17], as well as with the power-law behavior of certain visual neurons[18].

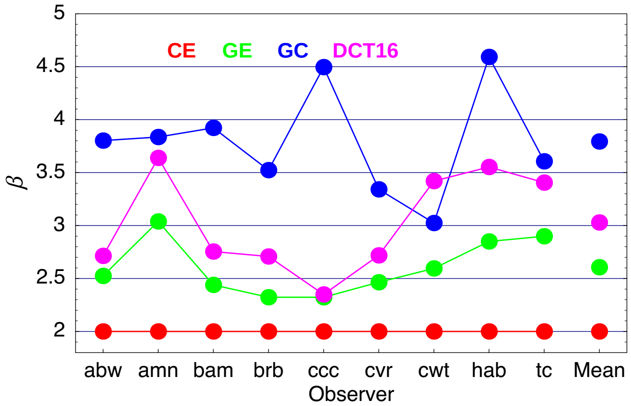

Figure 12. Estimated pooling exponent β for five models. Eight observers and the mean are shown.

Discussion

Peak Contrast Model

Not surprisingly, the Peak Contrast model provides a very poor fit to the data. Even though the Interpolation filter parameters are optimized for this model, the RMS error is over 5 dB. The mean residual error for each stimulus is shown in Figure 13. Because the Peak Contrast model takes no account of the area of the stimulus, the actual thresholds for the smallest stimuli (#14 and #29) are much larger than predicted.

Received October 14, 1999; Revised January 03, 2000

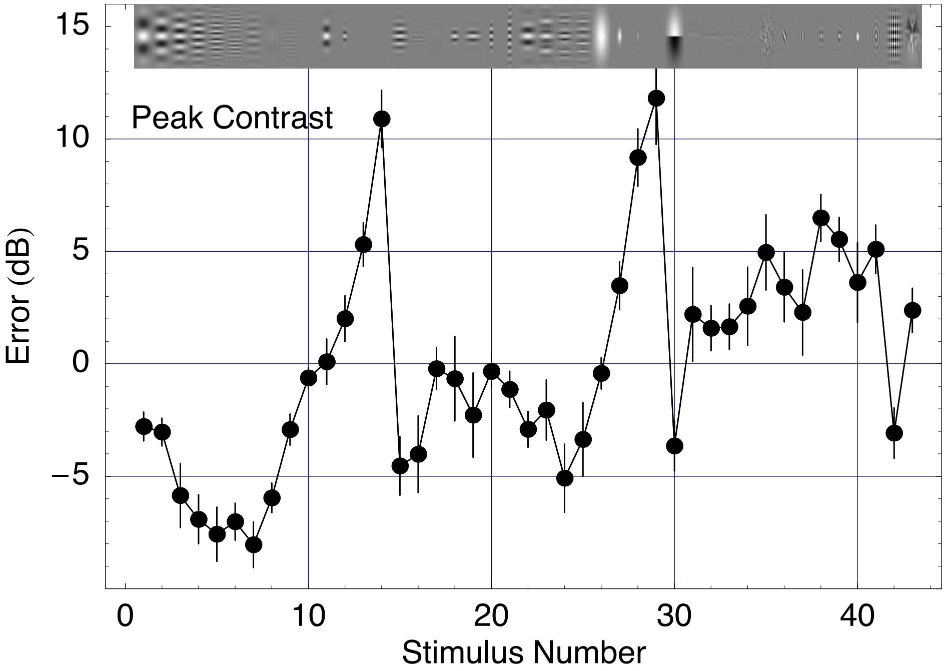

Figure 13. Mean error and standard deviation for each stimulus for the Peak Contrast model.

Contrast Energy Model

The Contrast Energy Model provides a remarkably good fit to the data. We must distinguish between models that serve a practical purpose, and which seek to predict mean performance with reasonable accuracy, and those models whose primary purpose is to test detailed theoretical assertions about visual structure and function, perhaps in a single individual. Certainly from the practical point of view, the Energy model is attractive since it is very simple to compute, it has a plausible basis in signal detection theory, and its errors of prediction are dwarfed by the differences among observers.

Figure 14 shows the mean error of the Contrast Energy model for each stimulus, averaged over observers. The mean errors are bounded by -4 and +7 dB. Some systematic departures from the model are evident. For stimuli #11-14 (Gabors with fixed numbers of cycles) the actual thresholds are progressively lower than the predictions as the frequency increases. A plausible explanation for this effect is that as frequency increases, stimulus area decreases. Since actual visual sensitivity decreases with eccentricity, while the model is spatially homogeneous, predicted thresholds for smaller targets should be too high. The same explanation can be offered for the Gaussians with decreasing standard deviations (stimuli #27-29), and possibly the line and dipole (#31 and #32) though these are small in only one of the two dimensions. Relatively reduced sensitivity for large targets could also be produced by inefficient summation over space, as occurs in the Generalized Energy model, discussed below.

Received October 14, 1999; Revised January 03, 2000

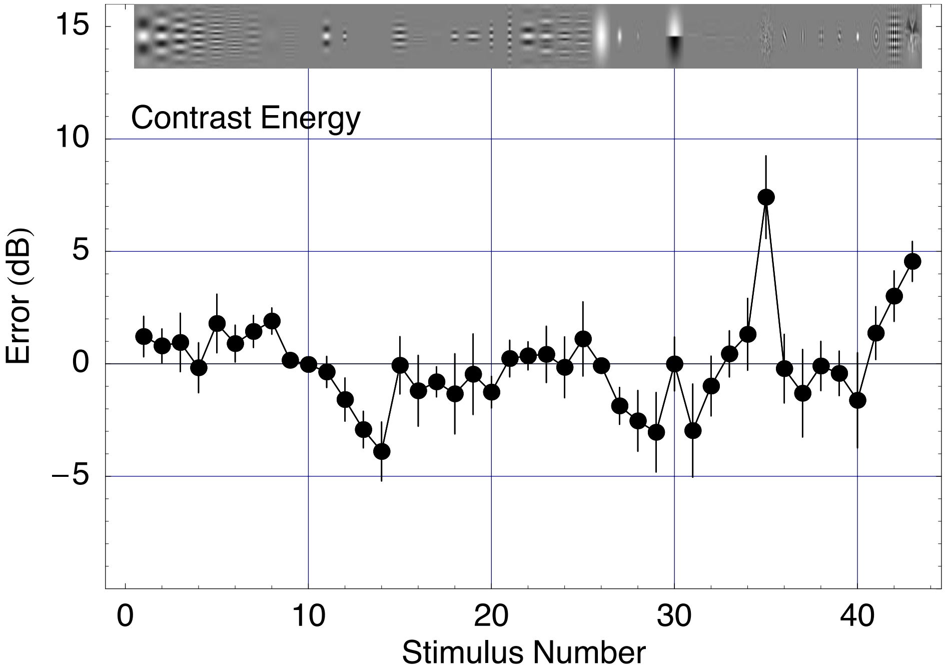

Figure 14. Mean error for each stimulus for the Contrast Energy Model.

The largest prediction errors in the positive direction (actual threshold > predicted) are for the noise sample (#35) and the last three stimuli, Bessel x Gaussian (#41), Checkerboard (#42) and the Natural Image (#43). All four are large stimuli, using the large standard Gaussian aperture, and could thus suffer from the size effects mentioned above. But this would make them no less visible than stimuli #1-10 (fixed size Gabors) which have the same size. One other property that they share is that they are broad-band, that is, they contain spatial frequencies distributed broadly over the two-dimensional frequency domain. Channel models, which partition this domain into bands and summate inefficiently between them, could therefore be expected to account better for these four stimuli.

Generalized Energy Model

In Figure 15 we plot the mean error versus stimulus for the Generalized Energy model. The fit of this model is remarkably good. Although it lacks channels or complicated processing of any kind, only two of the stimuli depart from the model predictions by more than 2 dB. In comparison to the Contrast Energy model, many of the larger negative excursions are greatly reduced, especially those for the smaller Gabors and Gaussians (#12-14, #27-29). This confirms the point made earlier that for centrally located targets, inefficient summation can mimic spatial inhomogeneity. The positive excursions for broad-band targets, especially at #35, remain, although they are attenuated. This makes sense, since the Generalized Energy model sums inefficiently over space, but cannot sum inefficiently over frequency since it lacks channels. The generalized energy model is similar to models proposed by Ahumada and colleagues[19-21], who have also pointed out that their performance may rival that of channel models at greatly reduced computational cost.

Received October 14, 1999; Revised January 03, 2000

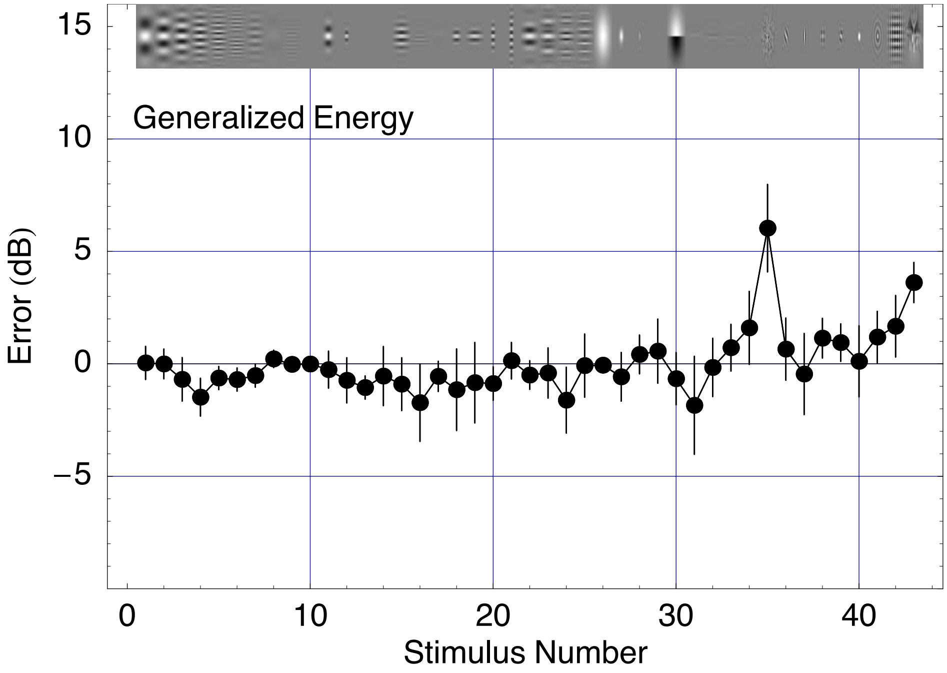

Figure 15. Mean error for each stimulus for the Generalized Energy model.

Gabor Channels Model

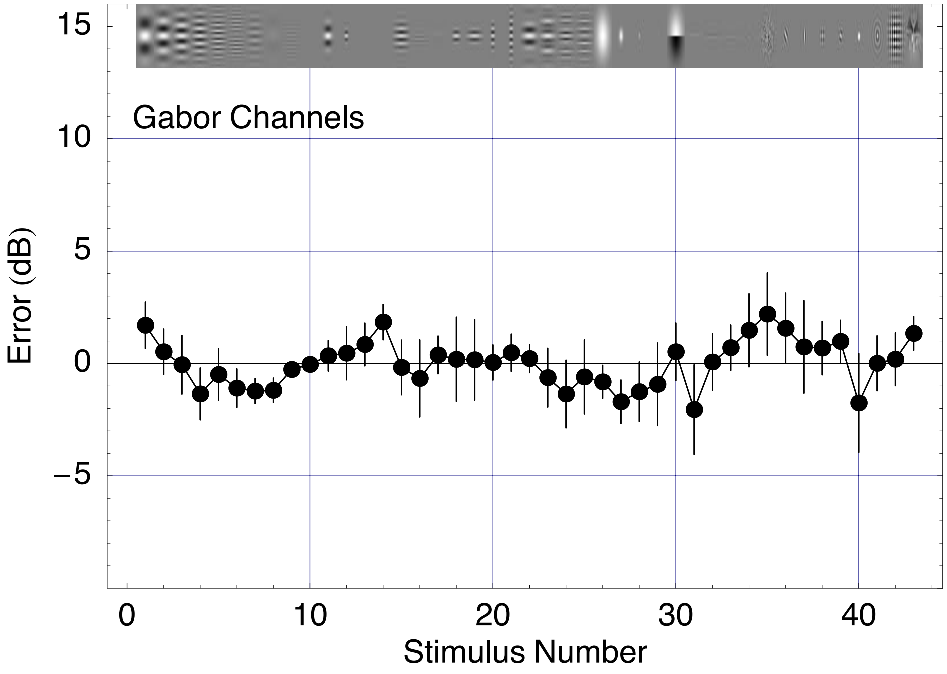

Figure 16. Mean error versus stimulus for the Gabor Channels model.

Received October 14, 1999; Revised January 03, 2000

In Figure 16 we plot the mean RMS error versus stimulus for the Gabor Channels model. On average, the Gabor Channels model provides the best fit to the data, though it is only slightly better than the Generalized Energy or DCT models. The maximum RMS error is about 2 dB. As expected, the large prediction errors for broad-band stimuli are eliminated, presumably due to inefficient summation over separate bands of frequency. The curious progression of errors for the set of fixed size (#1-10) and fixed cycles (#11-14) Gabors, may be explained in the following way. The smallest Gabors (which have higher spatial frequencies) may be difficult to fixate, leading to higher than predicted thresholds. The estimated Interpolation filter coefficients at higher spatial frequencies are thereby reduced, yielding predicted thresholds for the large fixed size Gabors that are greater than observed.

One troublesome feature of the Gabor Channels model is that it has no low-pass channel centered at 0 cycles/degree. Instead, the stimuli that may be dominated by their 0 cycles/degree component (Gaussians and disc) must be detected by the lowest frequency channels which peak at about 0.94 cycles/degree. This is a common failing among "channel" models, and in the future we will test whether addition of a low-pass channel improves the fit of the model.

DCT Models

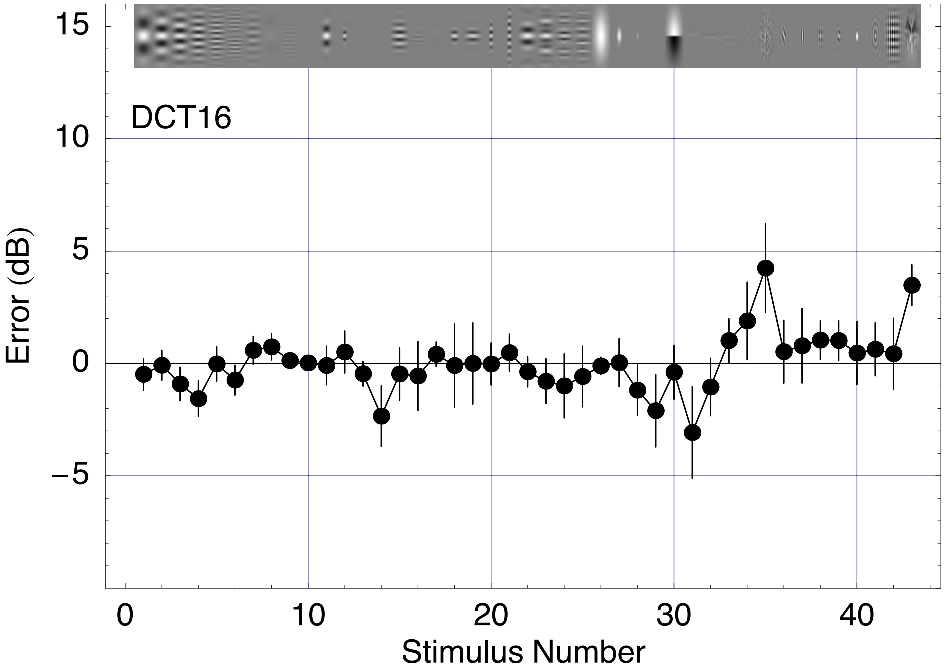

Figure 17. Mean error versus stimulus number for the DCT model with a block size of 16 pixels

The set of three DCT models, which vary only in block size, were considered to see whether a simple unitary transform could substitute for the more complex Gabor filter array of the Channel model. In general, the DCT models do not show much improvement over the Generalized Energy model. Indeed, with a block size of 8 pixels (1/15 degree), the GE and DCT predictions are nearly identical. This is no doubt because the lowest frequency "channel" in the 8x8 pixel DCT is at 7.5 cycles/degree; thus all the partition into separate bands of frequency occurs at relatively high frequencies, where there is little sensitivity and little stimulus energy. This is why we considered block sizes of 16 and 32 pixels. These reduce the lowest frequency "channel" to 3.75 and 1.875 cycles/degree respectively, but also enlarge the "receptive field' of each channel, narrow its bandwidth, and reduce its sampling

Received October 14, 1999; Revised January 03, 2000

density. The three block sizes yield RMS errors of 1.756, 1.713, 1.996. The mean error of the best of these (DCT16) is shown in Figure 17.

Alternate Measures of Model Fit

Maximum Absolute Average Error

Although RMS error is a simple intuitive measure of the goodness of fit of the models, it depends upon the selection of stimuli used in the experiment. For example, we have found that many models do well for narrow-band stimuli (e.g. Gabors), but not for broad-band stimuli (e.g. noise). Thus the RMS error would be quite different, and the relative performance of the models might be quite different, if we had used many broadband stimuli and few narrow-band stimuli. This problem cannot be entirely avoided, since our set of stimuli is a miniscule sample from a very large set. But one partial solution is to quantify model performance by the maximum in the average over observers of the RMS error for each stimulus. For example, examination of Figure 17 shows that RMS error for the DCT16 model, averaged over observers, has a maximum absolute value of about 4 (which occurs for the noise stimulus). Using this metric, the relative performance of the models is shown in Figure 18. This metric separates the performance of GE, DCT, and GC models, compared to simple RMS error. This is because the Gabor Channel model deals well with the noise stimulus, which is a challenge for all the other models.

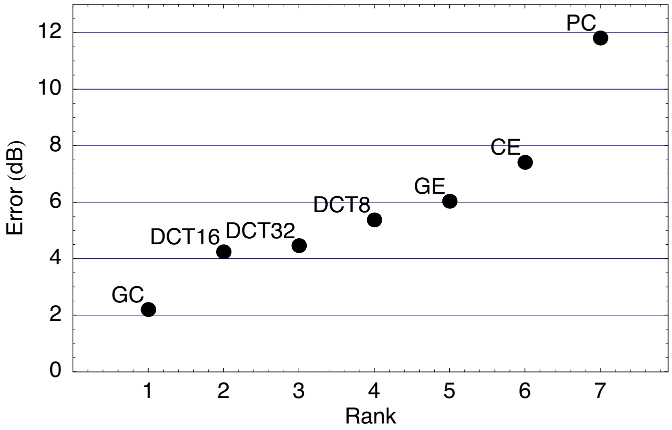

Figure 18. Maximum absolute average error for seven models.

Chi-Square Statistics

Examination of Figure 5 suggests that thresholds for the 43 stimuli have approximately equal variance. Under this homogeneity assumption, it is possible to construct a Chi-Square statistic for the fit of each model for each observer, and for the fit for the complete group of observers. For an individual observer, this statistic is

$$\chi^2_{k,m} = IJ \ln\left(1 + \frac{S_{k,m}}{V_k}\right). \qquad (10$$

The degrees of freedom is equal to the difference in the number of parameters in the unconstrained model (the 43 means plus one variance) and in the constrained model (11 for PC and CE, 12 for GE, DCT, GC).

For a single model, combining the results for all observers, the statistic is

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

$$\chi_m^2 = \sum_{k=1}^{K} \chi_{k,m}^2 \quad . \qquad (11$$

Below we provide a table of these statistics. In the case of the combined statistic, the Chi-Square with large degrees of freedom can be approximated by a Standard Normal, which is provided in the last column of the table. In all cases, the models are rejected at the 0.05 level. This is not atypical in tests of this sort, which assess whether the deviations from the model can be attributed to chance alone. Clearly the remaining deviations, even for the best model, while small, are not due to chance. At this stage we have made no effort to minimize the number of model parameters (we use 11 parameters for the csf filter) which makes the test particularly challenging.

Table 5. Chi-Square statistics for each model and observer, and for each model.

	abw	amn	bam	brb	ccc	cvr	cwt	hab	tc	*df*	All	*df*	N(0,1)
PC	458	348	388	414	427	445	456	387	443	33	3770	297	62.5
CE	211	212	150	181	180	210	227	228	269	33	1871	297	36.8
GE	169	70	122	167	150	180	140	151	152	32	1301	288	27.0
GC	123	110	82	142	180	134	140	114	131	32	1158	288	24.2
DCT8	171	67	115	167	152	177	129	143	152	32	1274	288	26.5
DCT16	160	105	99	149	163	166	108	135	155	32	1242	288	25.9
DCT32	175	200	108	143	178	169	181	195	236	32	1586	288	32.3

General Discussion

The ModelFest dataset, because it has been collected from a substantial number of observers, by a number of experimenters, and using a rather large and diverse set of stimuli, is a particularly useful test bed for models of spatial vision. One may use the dataset with at least lessened concern for vicissitudes of subject, lab, experimenter, or stimulus. Such concerns are not eliminated, of course. It is clear that the number of stimuli is still small, and a different selection might favor one model or another.

Within these constraints, this fitting exercise has provided a number of important insights. The first is that *all* of the models considered, with the exception of Peak Contrast, provided a reasonable fit to the data. In the context of the broad range of stimuli and the considerable size of individual differences, the residual errors were impressively small. As noted above, we may distinguish between practical and theoretical models, and in that sense any of the models considered (except Peak Contrast) here could serve the practical role.

A second important observation is that all of the models with $2 < \beta < 4$ performed substantially better than the simple Contrast Energy model in which $\beta = 2$. This modest increase in the inefficiency of summation, which may have many possible causes, appears a quite robust feature of the best fitting models. The models which share this feature (GE, GC, and DCT) differ little in the quality of their fits.

This leads to a further intriguing result. Much of the theoretical and experimental work in spatial vision in the last thirty years has focussed upon spatial channels; on their existence and on their detailed shape and number. However in this exercise, while the Gabor Channel model does provide the best fit, it is not much better than a model with rather crude channels (DCT16), or with no channels at all (GE). Thus while channels may be strongly implied by other psychophysical results, their effects here are modest, and evinced mainly by broadband stimuli (e.g. #35, noise).

Another insight gained is that certain stimuli proved dramatically harder to fit, or dramatically more effective at distinguishing among models. In particular, examination of

#14103 - $15.00 US
Received October 14, 1999; Revised January 03, 2000

Figure 19 shows that stimuli #35 (noise), #43 (natural image), and #31 (line) were troublesome for all models, while #35 and #14 (smallest Gabor) were the best at distinguishing among models.

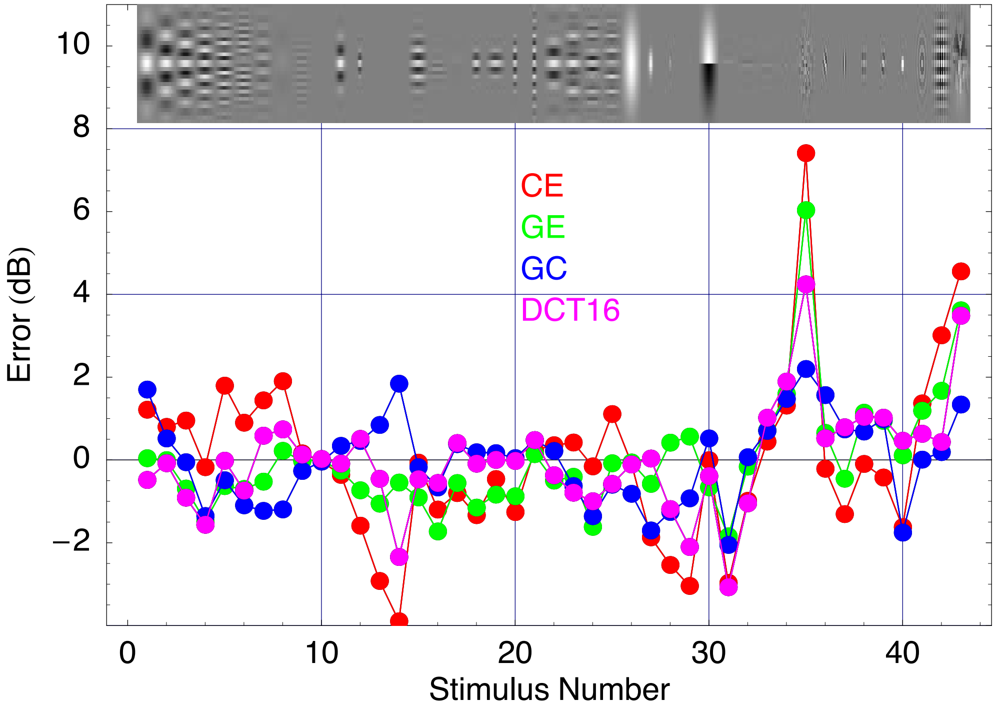

Figure 19. Mean error versus stimulus number for four models.

Standard Observers

A *standard observer* is a set of tabular data or a simple model designed to simulate the psychophysical performance of a specified population of observers. In color vision, standard observers have proven useful in both theoretical and practical applications[22]. No comparable standard observer exists for spatial vision. The relatively good fit of simple models to the ModelFest dataset, and the relatively consistent behavior of the model parameters (for example Figure 11), encourage us to consider the use of one of these models as the basis of a standard observer for spatial vision[23]. To be more useful, however, this standard observer should be augmented with a treatment of spatial contrast masking, which is largely absent from the present data, but which may be the focus of a future ModelFest experiment.

Future Models

Our purpose here has been to provide an initial survey of the performance of a small number of simple models on the ModelFest dataset. Here we offer a few comments on what might be profitable directions for future modeling of these data.

Perhaps the most significant attribute of threshold spatial vision that is absent from the present set of models is spatial inhomogeneity. All of the present models assume homogeneous sensitivity over the 2 degree square field, while in fact sensitivity, especially at the highest spatial frequencies, is known to vary markedly over this retinal extent. For example, sensitivity to 12 cycles/degree may decline by more than 6 dB over 1 degree of eccentricity[24].

The channel model considered here was designed somewhat arbitrarily to provide an initial estimate of the fit of such models. Channel models have many variants, and we may

Received October 14, 1999; Revised January 03, 2000

expect some other version to fit better than the one considered here. As noted above, the Gabor Channel model does not deal systematically with sensitivity to 0 cycles/degree. One solution to this defect might be to add a 0 cycle/degree channel to the Gabor Channel model, but because it is an ad hoc addition, rules governing its bandwidth and gain normalization are not obvious. Another approach would be a Wavelet model, in which both low-pass and band-pass filters are generated in a systematic way. Another similar approach would be to use "shiftable" filters[25].

In Figure 11 we showed that the estimated shape of the interpolation filter was similar for all observers, and could be approximated by a log parabola. A model with this form of simplified csf filter (with only 3 or 4 parameters instead of the 11 used by the interpolation filter) is another promising direction for further study, and a possible first step on the road towards a spatial standard observer[23].

Conclusions

We have fit the ModelFest Phase One dataset with five simple models. All models except for Peak Contrast provided reasonable fits, relative to the variability among observers. Of the remaining four models, the worst was the Contrast Energy model and the best was the Gabor Channel model. Generalized Energy and DCT models performed almost as well as the Gabor Channel model.

Acknowledgements

We thank Brent Beuter, Tina Beard, Albert Ahumada, Jeffrey Mulligan and Ruth Rosenholtz for thoughtful reviews of the manuscript, and the ModelFest group for solidarity and perseverance. This work was supported by NASA Grants 548-51-12-4110 and 199-06-12-39.

Received October 14, 1999; Revised January 03, 2000