Modelfest: year one results and plans for future years.

Thom Carney[*a], Christopher W. Tyler[b], Andrew B. Watson[c], Walter Makous[d], Brent Beutter[c], Chien-Chung Chen[b], Anthony M. Norcia[b], and Stanley A. Klein[e]

[a] Neurometrics Institute, 2400 Bancroft Way, Berkeley, CA 94704
[b] Smith-Kettlewell Eye Research Institute, San Francisco, CA 94115
[c] NASA Ames Research Center, Moffett Field, CA 94035
[d] Center for Visual Science, University of Rochester, Rochester, NY, 14627
[e] School of Optometry, University of California at Berkeley, Berkeley, CA 94720

ABSTRACT

A robust model of the human visual system (HVS) would have a major practical impact on the difficult technological problems of transmitting and storing digital images. Although most HVS models exhibit similarities, they may have significant differences in predicting performance. Different HVS models are rarely compared using the same set of psychophysical measurements, so their relative efficacy is unclear. The Modelfest organization was formed to solve this problem and accelerate the development of robust new models of human vision. Members of Modelfest have gathered psychophysical threshold data on the year one stimuli described at last year's SPIE meeting[1]. Modelfest is an exciting new approach to modeling involving the sharing of resources, learning from each other's modeling successes and providing a method to cross-validate proposed HVS models. The purpose of this presentation is to invite the Electronic Imaging community to participate in this effort and inform them of the developing database, which is available to all researchers interested in modeling human vision. In future years, the database will be extended to other domains such as visual masking, and temporal processing. This Modelfest progress report summarizes the stimulus definitions and data collection methods used, but focuses on the results of the phase one data collection effort. Each of the authors has provided at least one dataset from their respective laboratories. These data and data collected subsequent to the submission of this paper are posted on the WWW for further analysis and future modeling efforts.

Keywords: human vision modeling, threshold database, Modelfest, psychophysics, HVS, image compression

1. INTRODUCTION

The practical importance of having a robust computational model of human vision is perhaps nowhere more evident than at the annual SPIE meeting on Human Vision and Electronic Imaging. The rapid advances in digital transmission technologies, while impressive, cannot begin to keep up with the demand for quality images over the internet and other broadcast media. With the visual information content growing at a rate that exceeds the bandwidth of the hardware infrastructure the growing need for improved image compression methods is evident. Users of the medium demand ever higher image quality; gone are the days of thumbnail size blocky facsimiles of video. Lossless compression methods do not provide adequate bitrate savings; while lossy compression techniques can lower the bandwidth demands, they require a general model of human vision sufficient to identify where bitsaving measures will not degrade the video quality. As the demand for ever higher quality video grows, the quality of the human vision model embodied in the compression architecture must also improve. For automated evaluation of video compression technologies designed to produce high fidelity video from high-resolution source video, an advanced HVS model will be critical. The standard RMSE methods will no longer be adequate for the job. Conversely, high fidelity video compression technologies will increasingly need to incorporate advanced HVS model features to decide where bit saving can be achieved without degrading video fidelity. The requirement for a general purpose HVS model that predicts performance of the standard observer has never been more pressing. The vision science community, with many years of experience of modeling visual performance in many domains, will continue to contribute to the quest for a robust HVS model to aid in designing better compression algorithms for high fidelity video.

Over the past 35 years, the vision science community has made significant progress in understanding visual processing. Psychophysical and physiological studies have revealed a multi-stage parallel processing structure of the human visual system. Although most HVS models exhibit similarities, they also have distinct differences. The advantages and disadvantages of different model features and how they compare under different stimulus conditions are difficult to determine. HVS models are rarely compared using the same psychophysical data set[2,3], so the efficacy of different models is unclear. Interested researchers are generally left trying to reproduce the model from incomplete published descriptions when

In *Human Vision and Electronic Imaging V*, Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors, Proceedings of SPIE Vol. 3959 (2000) • 0277-786X/00/$15.00

trying to make comparisons. Partially in response to this situation, a workshop to promote vision modeling was organized for the 1997 annual OSA meeting. About 40 attendees participated in the workshop and began setting the framework for the Modelfest group. At subsequent ARVO and OSA meetings[4, 5], as well as through extensive internet communications, the group has focused on the goal of providing an extensive public stimulus database to be used for testing and developing HVS models. The threshold database would provide researchers with a 'standard observer' for spatio-temporal vision. The plan, which is now coming to fruition, was to create a database that included visual stimuli and corresponding psychophysical threshold data, from laboratories across the country. The first Modelfest data collection group, organized in 1998, decided to limit the first phase of data collection to monochromatic spatial patterns. The 44 stimuli deemed critical for developing and challenging vision models were soon available on the WEB[1,6,7] with threshold data available the following year. New Modelfest data groups are now forming to go beyond static achromatic spatial targets. Stimuli designed to challenge models in the areas of contrast masking, and temporal modulations are being developed. Membership in a data collection group is open to all those willing to collect a dataset once the group decides on the appropriate stimuli. Once a large, readily accessible database of stimuli with psychophysical thresholds exists, the developers of general purpose HVS models will be compelled to provide performance data using the database images so researchers can properly evaluate the model. It will soon become easier to determine which model innovations actually improve performance. Modelfest is a dramatic change from how HVS modeling has progressed in the past. This new approach offers researchers a simple way of comparing models and learning from each other's innovations and mistakes. This promises to facilitate the development of comprehensive HVS models consistent with physiological data as well as special purpose applied models for use in commercial applications. Several laboratories have already, or are about to, report model fits to the data [8,9,10,11,12,13]. Here we provide a progress report and some preliminary analysis of subsets of the data.

2. METHODS

This section presents an overview of the methods and stimuli. For a more detailed presentation of the methods and stimuli used by the Modelfest group see Carney et. al.[1] or visit one of our WWW sites:
http://neurometrics.com/projects/Modelfest/IndexModelfest.htm and http://vision.arc.nasa.gov/modelfest/ .
The final stimulus set was slightly different from that described in our previous paper. The dipole stimulus was changed from 2 to 3 pixels wide, (a mean luminance pixel was inserted between the bright line and dark line) because the two pixel stimulus was too weak for assessing threshold in many cases. The fixation pattern was also changed as described below.

2.1 Display and psychophysical methods

The list below includes the important required display and subject viewing conditions and psychophysical methods:

- Display mean luminance: 30 ±5 cd/m^2
- Display frame rate: ≥ 60 hz.
- Display pixel size: 0.5 min. square.
- Display gray scale resolution: 1/4 or less of the stimulus threshold (d'=1).
- Stimulus temporal waveform: 500 msec Gaussian (125 msec sd)
- Subject Viewing: Binocular viewing with natural pupils.
- Presentation: 2AFC or rating scale
- Data collection/analysis: Objective methods. (staircase, Quest, method of constant stimuli)
- Trial placement: Most trials must be located near the final threshold (d' = 1-2)
- Repetitions: Thresholds based on a minimum of 4 blocked runs
- Threshold Level: equivalent to 84% correct on 2AFC.
- Fixation: Narrow high-contrast "L" shaped pattern located at the stimulus corners.

2.2 Stimulus specification

The stimuli were 44 static 256 by 256 (0.5 min) pixel grayscale images. Most stimulus patterns were multiplied by a radial Gaussian envelope with a s.d. of 0.5 deg so the edges the patterns were approximately the same luminance as the surrounding field. To facilitate transferring images between laboratories and computer operating systems the images were stored as compressed industry standard TIFF files. The distributed images were at maximum contrast and could span the 8-bit range (minus one) from 1 to 255, with the mean luminance of the display surrounding the stimulus pattern at 128. When possible the predominant modulation in the image was oriented vertically to minimize display adjacent pixel luminance interactions. A Gaussian function (sd = 125 msec) of 500 msec duration temporally modulated the stimulus on each presentation to limit transients.

Table 1, adapted from one of our WEB sites[1], characterizes the stimuli, names of the TIFF files and shows the mean subject threshold, as of January 2000. The first two columns provide the condition number and stimulus description. Column 3 indicates the base spatial frequency or other unit of size of the stimulus. Columns 4 & 5 indicate the standard deviation of the Gaussian envelope in the x & y directions, respectively. Column 6 shows the vertical size in octaves, half amplitude full

bandwidth. The conversion from degrees (column 5) to octaves is: octave = 0.56 * deg/freq. Column 9 shows the threshold and standard error of the mean (n=9, except stimulus #44, n = 6).

Cond.	Stimulus Description	Spatial Scale	Horiz. Size (deg.)	Vert. Size (deg.)	Vert. Size (oct.)	Area relative to #12.	Stimulus Filename	Threshold -log10(cont)
1	Gabor fixed size	1.12 c/d	0.5	0.5	1	12.8	GaborPatch1.tif	1.83±0.05
2	Gabor fixed size	2 c/d	0.5	0.5	0.56	12.8	GaborPatch2.tif	1.96±0.05
3	Gabor fixed size	2.83 c/d	0.5	0.5	0.396	12.8	GaborPatch3.tif	2.07±0.05
4	Gabor fixed size	4 c/d	0.5	0.5	0.28	12.8	GaborPatch4.tif	2.13±0.06
5	Gabor fixed size	5.66 c/d	0.5	0.5	0.198	12.8	GaborPatch5.tif	2.01±0.05
6	Gabor fixed size	8 c/d	0.5	0.5	0.14	12.8	GaborPatch6.tif	1.85±0.05
7	Gabor fixed size	11.3 c/d	0.5	0.5	0.099	12.8	GaborPatch7.tif	1.64±0.06
8	Gabor fixed size	16 c/d	0.5	0.5	0.07	12.8	GaborPatch8.tif	1.31±0.05
9	Gabor fixed size	22.6 c/d	0.5	0.5	0.050	12.8	GaborPatch9.tif	0.97±0.07
10	Gabor fixed size	30 c/d	0.5	0.5	0.037	12.8	GaborPatch10.tif	0.57±0.06
11	Gabor fixed cycles	2 c/d	0.28	0.28	1	4.0	GaborPatch11.tif	1.79±0.06
12	Gabor fixed cycles	4 c/d	0.14	0.14	1	1.0	GaborPatch12.tif	1.65±0.05
13	Gabor fixed cycles	8 c/d	0.07	0.07	1	0.25	GaborPatch13.tif	1.22±0.05
14	Gabor fixed cycles	16 c/d	0.035	0.035	1	0.0625	GaborPatch14.tif	0.53±0.06
15	Elongated Gabor	4 c/d	0.5	0.28	0.5	7.143	ElongatedGabor15.tif	2.00±0.06
16	Elongated Gabor	8 c/d	0.5	0.14	0.5	3.571	ElongatedGabor16.tif	1.69±0.07
17	Elongated Gabor	16 c/d	0.5	0.07	0.5	1.788	ElongatedGabor17.tif	1.05±0.06
18	Baguette Summation	4 c/d	0.28	0.14	1	2.00	ElongatedGabor18.tif	1.78±0.06
19	Baguette Summation	4 c/d	0.5	0.14	1	3.571	Baguette19.tif	1.87±0.06
20	Baguette Summation	4 c/d	0.14	0.28	0.5	2.000	Baguette20.tif	1.78±0.06
21	Baguette Summation	4 c/d	0.14	0.5	0.28	3.57	Baguette21.tif	1.83±0.05
22	Subthreshold Sum	2 & $2\sqrt{2}$ c/d	0.5	0.5		12.8	Subthreshold22.tif	1.94±0.05
23	Subthreshold Sum	2 & 4 c/d	0.5	0.5		12.8	Subthreshold23.tif	1.90±0.07
24	Subthreshold Sum	4 & $4\sqrt{2}$ c/d	0.5	0.5		12.8	Subthreshold24.tif	1.97±0.07
25	Subthreshold Sum	4 & 8 c/d	0.5	0.5		12.8	Subthreshold25.tif	1.83±0.07
26	Gaussian		0.5	0.5		12.8	Gaussians26.tif	1.60±0.09
27	Gaussian		0.14	0.14		1.00	Gaussians27.tif	1.55±0.06
28	Gaussian		0.0352	0.0352		0.063	Gaussians28.tif	1.19±0.04
29	Gaussian		0.0175	0.0175		0.016	Gaussians29.tif	0.82±0.06
30	Edge		0.5	0.5		12.8	Edge30.tif	1.96±0.05
31	Line	0.5 min wide	0.5				Line31.tif	0.94±0.07
32	Dipole (+1, 0, -1)	1.5 min.	0.5				Dipole32.tif	0.66±0.04
33	5 Collinear Gabors	8 c/d, in phase	ea. 0.07	ea. 0.07	1	0.25	GaborString33.tif	1.40±0.06
34	5 Collinear Gabors	8 c/d, antiphase	ea. 0.07	ea. 0.07	1	0.25	GaborString34.tif	1.36±0.06
35	Binary noise	1 min pixels	0.5	0.5		12.8	Noise35.tif	1.31±0.06
36	Oriented Gabor	4 c/d, diag.	0.14	0.14	1	1.00	Orientation36.tif	1.58±0.05
37	Oriented Gabor	4 c/d, vert.	0.14	0.14	1	1.00	Orientation37.tif	1.63±0.05
38	Plaid	4 c/d, vert & hor	0.14	0.14	1	1.00	Plaids38.tif	1.42±0.04
39	Plaid	4 c/d, 45 & hor	0.14	0.14		1.00	Plaids39.tif	1.47±0.05
40	Disk	.25 deg dia.					Disk40.tif	1.63±0.06
41	Bessel function	4 c/d	0.5	0.5		12.8	Bessel41.tif	1.53±0.06
42	Checkerboard	4 c/d fund.	0.5	0.5		12.8	Checkerboard42.tif	2.05±0.05
43	Natural scene	San Francisco	0.5	0.5		12.8	NaturalScene43.tif	1.53±0.05
44	Variable noise	1 min pixel	0.5	0.5		12.8		1.12±0.18

Table 1: Stimulus descriptions and psychophysical thresholds

3. RESULTS

3.1 Contrast Sensitivity Function

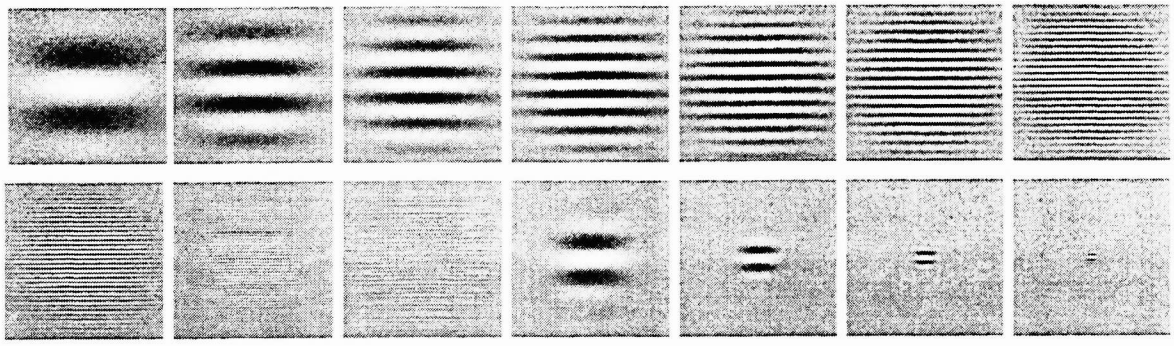


Figure 1: Fourteen Gabor patch stimuli used to characterize the contrast sensitivity function and spatial summation

The first ten stimuli constitute a conventional probe of the contrast sensitivity as a function of spatial frequency (CSF) for a fixed spatial aperture. This provides an important baseline series for comparison with previous data. Although the CSF has been measured extensively over the past half-century, most studies have used extended patterns that stimulated inhomogeneous regions of central retina, and with sharp edges that could mask sensitivity in their vicinity. To focus on a relatively homogeneous zone of the retina, we set the envelope of the stimuli to a full width of 1 deg (at half height). The envelope itself was Gaussian, to minimize the effects of edges on the detectability of the single spatial frequency. Thus, although the low contrast tails of the stimulus extended out beyond a 0.5 deg radius, the stimuli in the entire Modelfest set were essentially restricted to the foveola.

In addition to the foveal location and edge minimization, the stimuli were restricted to the low temporal frequency range by employing a Gaussian temporal envelope with a total duration of 500 msec. This is a sustained temporal presentation paradigm designed to minimize intrusion of transient neural responses over most of the operating range. Based on previous work, we expect the CSF to exhibit a bandpass form under these conditions. The average thresholds for the nine observers (Fig. 2, filled diamonds) indeed exhibit this bandpass form, peaking at about 4 cy/deg. This form is in line with expectations for the foveola, based on previous work on individual observers[14, 15]. The average thresholds for the same observers for Gabor patches with a fixed one octave bandwidth are lower, as expected from limited spatial summation (Fig 2, filled squares). The dashed lines indicate results from individual subjects.

To characterize the data more completely, they are fitted with a subtractive inhibition model. The excitatory component is a simple exponential function of spatial frequency, as well established by Campbell, Kulikowski & Levinson[16] to account for the fall in sensitivity at high spatial frequencies. The inhibitory component is assumed to be a Gaussian function subtracting linearly from the excitatory component as defined in equation one below:

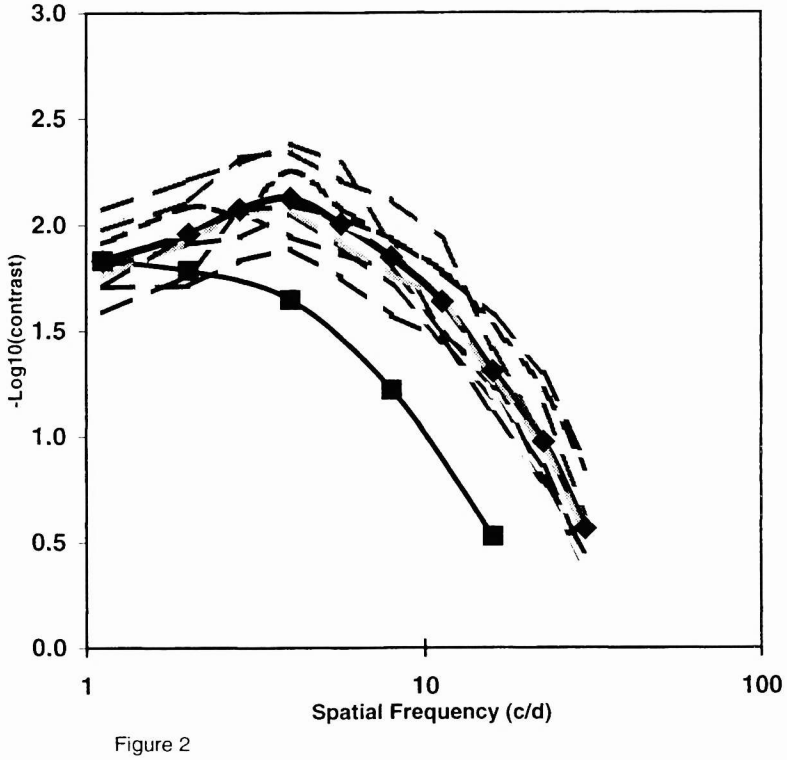


Figure 2

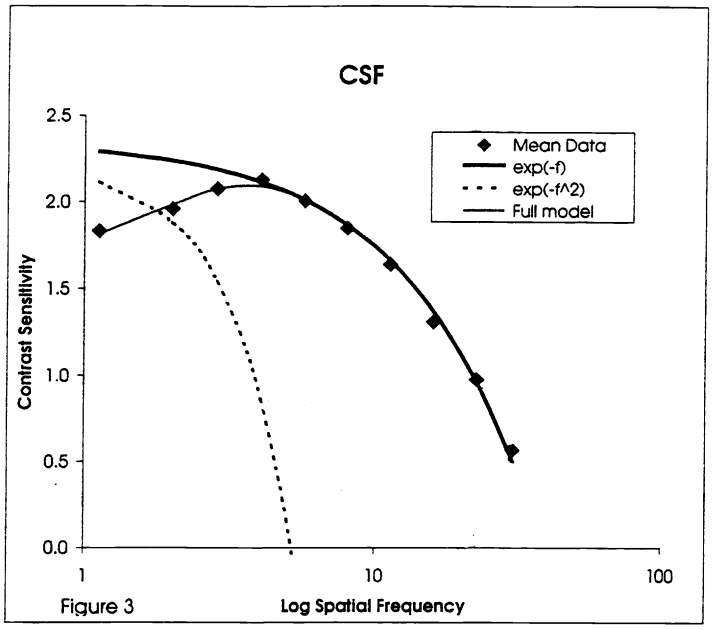


Figure 3

$$CSF = A\left(e^{-f/\omega} - ke^{-(f/\sigma)^2}\right) \quad (1)$$

where the best-fitting constants are A=216, k=0.71, ω=7.22 c/deg and σ=2.2 c/deg.

The curves in Fig. 3 plot the excitatory component (thick curve), the inhibitory component alone (dashed curve) and their subtractive combination (thin curve), which provides a close fit to the data.

If the CSF is mediated by a single-channel process incorporating subtractive inhibition, the components will have to have characteristics very close to those depicted.

The error in the data is represented by the height of the symbols, in terms of ±2 s.e.m. This is not the raw error over observers but the residual error after normalizing the curves for individual observers to the overall mean sensitivity. The normalized error averaged over spatial frequency was 0.065 log units. The normalized value represents the error in the shape of the CSF rather than in its overall sensitivity. The exponential-minus-Gaussian model of the CSF provides a fit within 2 s.e.m. to the data at all spatial frequencies.

In his recent report, Watson[13] showed that the CSF, considered as a spatial filter, could be described by a parabola in a log sensitivity-log frequency space. Since he was fitting particular models to the entire data set rather than just to the fixed size Gabors, we cannot directly compare his fit to that depicted in Figure 3.

3.2 Spatial (area) summation.

A number of stimuli in our battery were selected to investigate the properties of spatial summation in the principal (horizontal) orientation. Stimuli #4, 12, 15, 18, 19, 20, and 21 are all 4 c/deg Gabor patterns with different aspect ratios. Pictures of the 8 stimuli are shown in figure 4. The area of each stimulus (relative to the area of the smallest patch - #12) is given in column 7 of table 1. As shown in figure 4, the larger the stimulus in either horizontal or vertical extend (other things being equal) the lower the threshold. Summation was similar in either direction. As expected, sensitivity for stimulus 12, the smallest, was the lowest of the group, whereas, sensitivity for stimulus 4, the largest, was the highest. To quantify the relationship between stimulus size and sensitivity we examined threshold as a function of stimulus area.

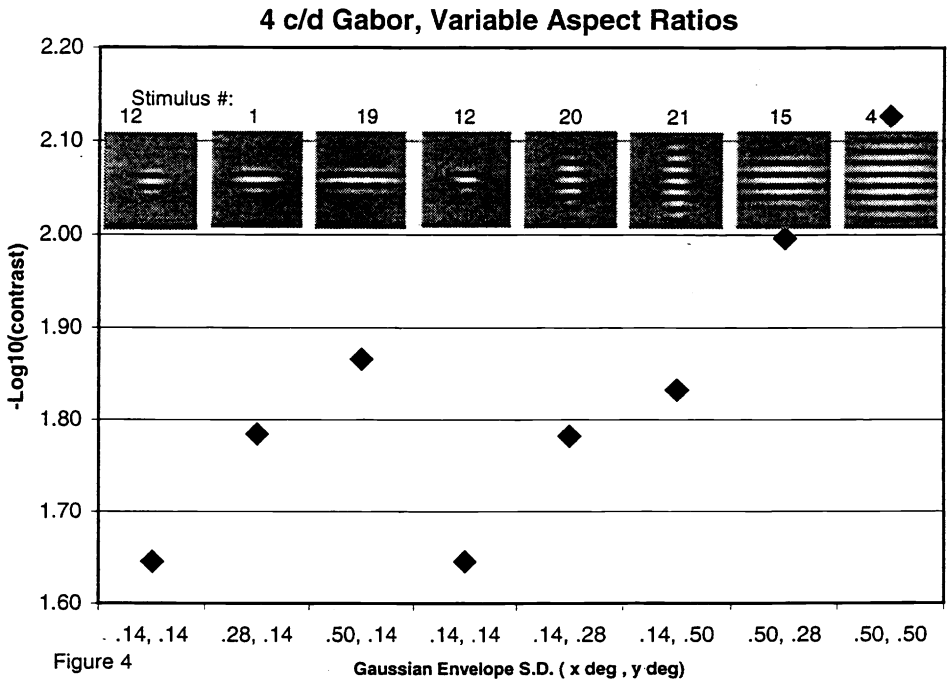
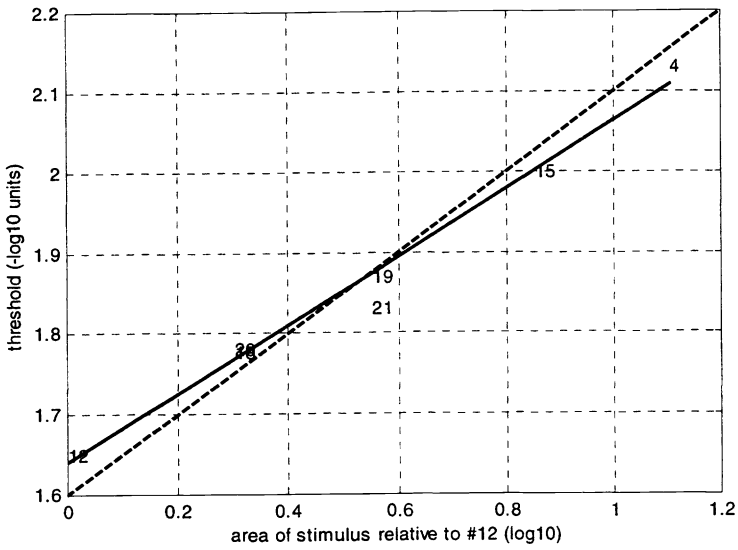


Figure 4

Figure 5 below is a plot of the threshold vs. the log of the stimulus area. The solid and dashed lines are power function fits to the data.



Sensitivity = 1/ contrast = k*Area^(1/p)

where p is the probability summation pooling exponent[17,18].

For the solid line the power, p, is allowed to float and for the dashed line it is fixed at p=2.0. The best fit has an exponent of p = 2.36. Energy summation (p=2) shown by the dashed line does not fit the data as well.

Of particular interest, in light of the findings of Polat and Tyler[19], was the summation along the length versus width dimensions for aspect ratios of 2:1 (stimuli 18 and 20) and 4:1 (stimuli 19 and 21). Thresholds were 16.5 dL for the small patch (stimulus 12), 17.8 dL for double size stimuli #18 and 20) and 18.5 for the average of the 4 times larger stimuli (#19 and 21). Thus the two-fold enlargement reduced thresholds by a factor of 1.35 and the four-fold enlargement reduced thresholds by 1.58. The probability summation exponent for the double size stimuli is 2.3 (i.e. $2^{1/2.3} = 1.35$) and for the 4 times larger stimuli the spatial summation exponent was 3.0 (i.e. $4^{1/3.0} = 1.58$). The slight advantage of length over width summation was not significant ($p<0.07$, one-tailed t-test), but is in the same direction as reported by Polat and Tyler[19]. Polat and Tyler found summation to increase in some of their observers well beyond aspect ratios of 4:1 which were not examined in Modelfest. The greater degrees of elongation used by Polat and Tyler may account for some of the difference in results.

3.3 Subthreshold summation.

The Modelfest battery contains several stimuli that place constraints on how the underlying mechanisms summate. The subthreshold summation task has played an important role in the development of spatial vision models. Campbell and Robson[14] showed that the detection threshold of square wave gratings above about 3 c/deg was based on the visibility of the fundamental. The subthreshold third and higher harmonics did not summate with the fundamental. The implication was that detection thresholds were based on multiple tuned mechanisms that were not able to summate across diverse stimuli. Thomas[20] came to a similar conclusion using localized (non-sinusoidal) stimuli. Graham and Nachmias[21] established a two-component methodology for calculating the amount of subthreshold summation. They measured the detection thresholds of a 3 c/deg grating, a 9 c/deg grating and composite gratings consisting of the two simple gratings. They found that the thresholds for the composite was very close to the independent threshold of either of the simple gratings. There was minimal summation. Only a very tiny amount of probability summation was found. This is not the place to go into the mathematics of probability summation. An important practical advance in the mechanics of doing probability summation was suggested by Frank Quick[17]. He pointed out that if the psychometric function had a Weibull function shape then probability summation would be achieved by summing the various components using a Minkowski summation. Stromeyer and Klein[22] pointed out that the same probability summation formula arises in signal detection theory, given the accelerated transducer function. We will present the derivation in the signal detection framework for two component stimuli.

Signal detection theory says that the visibility of two stimuli processed by independent channels should be given by the Pythagorean summation of the d's associated with each stimulus:

$$d'^{2}_{c} = d'^{2}_{1} + d'^{2}_{2} \quad (2)$$

where d'_c is the d' for the composite stimulus and d'_1 is the d' for the first component. It has been shown that for contrasts near threshold d' is well approximated by an accelerating power function of contrast[23,24]:

$$d' = (c/c_{thresh})^{t} \quad (3)$$

where c_{thresh} is the detection threshold, namely the contrast at which d' is unity, and t is the transducer exponent. In a wide range of experiments we have found that the transducer exponent is between 1.5 and 2.0[24]. Putting Eqs. 2 and 3 together gives an equation for the threshold of the composite pattern to be at threshold, $d'_{tot} = 1$:

$$1 = (c_1/c_{1thresh})^{p} + (c_2/c_{2thresh})^{p} \quad (4)$$

where the pooling exponent, p, is given by: p = 2t. Several of our Modelfest stimuli were composite stimuli in which the contrast of two components were equal. At threshold, the contrast of each of the components of the composite was half of the reported composite threshold (that was how the composite contrast threshold was reported).

Thus, Eq. 4 becomes:

$$1 = (c_{c\ thresh}/2c_{1\ thresh})^p + (c_{c\ thresh}/2c_{2\ thresh})^p \quad (5)$$

Eq. 4 is identical to the formalism of Quick[17] for probability summation based on high threshold assumptions. A value of p=1 corresponds to full linear summation. If p=2 we get Pythagorean summation, which is sometimes called energy summation or ideal observer summation of independent channels. For a very large pooling exponent (p>5) the summation is very close to the maximum of the independent sensitivities. Previous research indicated that the pooling exponent for independent channels was between p=3 and p=4. Future detailed filter modeling based on the Modelfest data will be able to discriminate between pooling exponents for summation across space and summation across multiple mechanisms at one point in visual space.

The Modelfest data had six triplets of stimuli specifically chosen for pinning down the pooling exponent for summation across possibly independent channels. The following table gives the six triplets, with the contrast of the various components at threshold, and the unique pooling exponent that satisfies Eq. 4.

stimuli	stimulus numbers	Thresh 1	Thresh 2	components	pooling exp
2 & 2.8	#2, #3, #22	0.011	0.0085	0.0057	1.36
2 & 4	#2, #4, #23	0.011	0.0074	0.0063	2.17
4 & 5.6	#4, #5, #24	0.0074	0.0098	0.0054	1.55
4 & 8	#4, #6, #25	0.0074	0.0141	0.0074	6.47
horiz & vert	#12, #37, #38	0.0224	0.0234	0.0190	3.73
horiz & diag	#12, #36, #39	0.0224	0.0263	0.0169	1.96

Table 2: Pooling exponents from subthreshold summation stimuli

For example, for the first triple, Eq. 4 becomes: $(0.0057/0.011)^{1.36}+(0.0057/0.0085)^{1.36} = 1$. Note that the contrast in the fifth column ('components') is the contrast of each component, which is half the peak contrast of the total composite pattern that is reported in the data summary in Table 1.

The pooling exponents of 1.36 and 1.55 for summation between components separated by a half octave implies that the mechanism tuning is substantially greater than a half octave. Watson[26] found less subthreshold summation between a 1 c/deg and a 1.41 c/deg grating than what we find here. For components separated by one octave we find substantially less summation with no summation at all between a 4 and 8 c/deg components. This difference is evidence of higher spatial frequencies having narrower tuning.

We also carried out subthreshold summation of across orientation using the one-octave 4 c/deg stimulus patches. As seen in the table the pooling exponent of 3.73 for the summation of horizontal and vertical components is compatible with what would be expected from probability summation. The lower exponent of 1.96 for summation of horizontal and diagonal components indicates that the underlying mechanisms may be broad enough in their orientation tuning to allow some summation across 45 deg.

3.4 Collinear Gabor patches

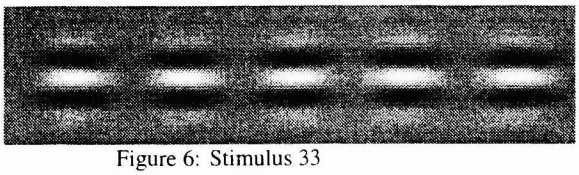
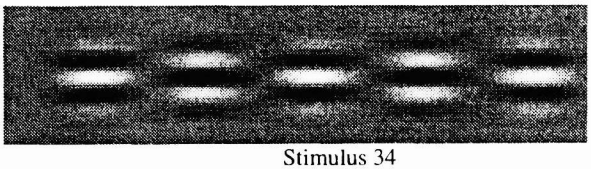


Figure 6: Stimulus 33



Stimulus 34

Stimuli #33 (a chain of five 8 c/deg in-phase Gabor patches) and #34 (a chain of five alternating phase Gabor patches), first used by Chen & Tyler[27], were designed to test the hypothesis that the pattern detection thresholds to images were determined by a band of elongated linear filters in the visual system. Under this hypothesis, compared to a single Gabor patch (stimulus #13), the presentation of extra Gabor patches of the same spatial frequency, orientation, and phase should produce a much stronger response in a linear filter and thus reduce the threshold. The data for a single 8 c/deg patch was 12.2 dL and the threshold of the five in-phase patches was 14.0 dL. The difference corresponds to a factor of 1.51 threshold reduction (threshold (dL) = -10 log10(contrast)).

The nonlinear summation exponent of 2.4 found above for the 4 c/deg Gabor stimuli would imply that there should be a threshold reduction of $5^{1/2.4} = 1.96$. That the actual threshold improvement was less than this implies some different mechanism in operation. One possibility is that there is the same kind of neural summation operative at 8 cy/deg as at 4

cy/deg. The smaller threshold reduction would then imply that this summation had a smaller spatial range at 8 than at 4 cy/deg, so the energy of the outer two patches did not contribute significantly to threshold. Any such neural summation must be phase insensitive because there is no significant difference between the two phase conditions ($t(8)=0.2388$, $p = 0.8173 > 0.05$). The alternative possibility is to postulate that there is no neural summation at 8 cy/deg, and the entire effect is explained by probability summation across local contrasts with an exponent of 4 (i.e., $5^{1/4} = 1.5$). Such probability summation would also account for the observed phase-insensitivity, but it seems implausible to suppose that the neural summation found at 4 cy/deg would have completely evaporated by 8 cy/deg.

3.5 Multipoles and mechanism bandwidths

Three of our stimuli were members of the local multipole family: the edge, line and dipole. Each member of this family is the derivative of the preceding member. The multipole family can be used to characterize spatial sensitivity similar to how the CSF characterizes sensitivity in terms of sinusoid thresholds. Just as one can analyze extended patterns in terms of sinusoids (a Fourier analysis) one can analyze local patterns in terms of their moments (a multipole analysis). Klein[28] showed how the ratio of multipole sensitivity to sinusoid sensitivity could be used to characterize the bandwidth of the underlying mechanisms. If one assumes a peak detection model (no probability summation) then:

a) The mechanism that detects the edge is the mechanism that detects the sinusoid at the CSF peak.
b) The mechanism that detects the line is the mechanism that detects the sinusoid at the point where the CSF has a slope of –1 (on log-log coordinates).
c) The mechanism that detects the dipole is the mechanism that detects the sinusoid at the point where the CSF has a slope of –2 (on log-log coordinates).

The formula for the mechanism bandwidth (Eq. 18 of Klein[28]) is:

$$BW = (\pi/2)^{1/2} / (CSF(f)\, M_m\, f^{m+1}) \qquad (6)$$

where M_m is the multiple moment and f has is specified in radians/min. Table 3 gives the values of the various items.

Order m	name	Multipole moment	spatial freq. (c/deg)	spatial freq. (rad/min)	CSF (1/%)	Bandwidth (fractional)	Bandwidth (octaves)
0	edge	2.2 %	3.4	0.35	1.21	0.47	1.6
1	line	5.7 %min	7.4	0.77	0.78	0.36	1.2
2	dipole	10.9 %min^2	14.6	1.53	0.28	0.17	0.6

Table 3: Multipole detection mechanism bandwidths

Consider the line, for example. The line threshold is 5.7 %min. All the multipole thresholds are slightly higher than the values found by others probably because of our relatively low luminance and brief duration. The spatial frequency at which the CSF slope is –1 on log-log coordinates is 7.4 c/deg, using the CSF of Eq. 1. The CSF value at that point is 78 (or 0.78 in units of reciprocal %). The fractional bandwidths in the 7th column are discussed by Klein[28]. They are equal to $(2\pi)^{-1/2}$ times the area under the mechanism tuning curve on a logarithmic frequency axis. The normalization was chosen so that for relatively narrow mechanisms the bandwidth, W, is the ratio of the standard deviation of the mechanism tuning divided by its peak frequency. It is also approximately 0.3 times the number of octaves between the half maximum points. Finally, 1/W is approximately the number of half-cycles in the mechanism's receptive field. The last column of the table, giving the approximate bandwidth in octaves, shows that the mechanism detecting the line has the medium bandwidth that is commonly assumed for the underlying mechanisms. The bandwidth near the peak of the CSF (used for edge detection) is somewhat broader, and the mechanisms at higher spatial frequencies (used for dipole detection) seem to be substantially narrower. The notion that the mechanism tuning gets narrower at high spatial frequencies is not new[29].

Klein (1989) estimated bandwidths for the edge, line and dipole CSF regions, of 0.47, 0.41 and 0.36, which are similar to the present estimates [Note that the bandwidth values in Klein's[28] Table 2 are actually reciprocal bandwidths]. Our estimate of the dipole bandwidth is somewhat lower than expected, possibly because our spatial and temporal uncertainty elevated our dipole thresholds. Improved estimates of mechanism bandwidths will require a full filter model fit to the full data. The multipole thresholds will provide strong constraints on mechanism bandwidths in that modeling.

3.6 Gaussian blobs

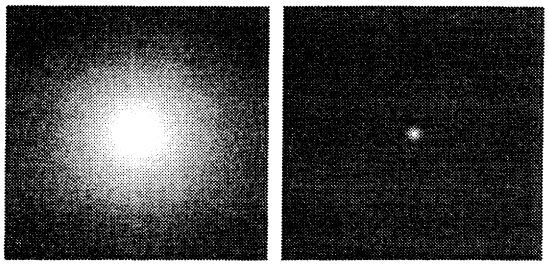


Figure 7: Stimulus 26 and Stimulus 29

Four stimuli, referred to in the first paper on Modelfest[1] as *Gaussian blobs*, have a luminance profile described by the equation:

$$L = \exp(-r^2/2\sigma^2), \qquad (7)$$

where r is any radius and σ is 30.0, 8.43, 2.106, or 1.05 min. The largest and smallest of the four (stimuli 26 and 29, respectively) are shown to the left.

Spatial summation. A reason for including these stimuli was to minimize both spatial frequency and spatial extent within the limits each imposes on the other. Another attractive feature of these stimuli is that they do not selectively stimulate orientation-selective mechanisms. Hence the data may be useful in testing ideas about how such mechanisms interact and in testing models that incorporate mechanisms lacking orientation selectivity.

These stimuli also allow one to test ideas about the spatial summation of light. Classically, the effects of light falling within certain spatio-temporal limits sum linearly to reach threshold. In the spatial domain, this phenomenon is referred to as *Ricco's law*. The so-called *critical area* within which Ricco's law holds depends on the sensitivity of the experiment and on stimulus conditions, such as luminance, retinal eccentricity, and shape and color of the test stimulus[30,31,32]. In our experiment, Ricco's law did not hold even for the two smallest stimuli, 1.05 and 2.106 min. The mean log sensitivities for these two stimuli were 0.815 and 1.192, respectively, a difference of 0.377. As one stimulus had 4 times the area of the other, full summation of light would have yielded a log difference of 0.602 instead of 0.377. This lack of complete summation is highly reliable statistically. Removal of the effects of differences in overall sensitivities of the different observers, as reflected by their mean thresholds, reduces the standard errors of the sensitivities of the four Gaussian blobs (stimuli 26 to 29) to 0.085, 0.044, 0.023 and 0.034, respectively. A t-score ($t = 5.48$, $p < 0.001$) for the difference between complete summation and the summation observed is highly reliable. This is close to the smallest practical test of neural summation possible with the eye's natural optics using the best clinical correction, for smaller stimuli approach the point-spread function of the eye[33], and any summation observed would be unduly contaminated by optical summation instead of neural summation[34].

However, the sensitivity to the 2.106 Gaussian was nevertheless greater than that to the 1.05 Gaussian, and sensitivity to the 8.4 Gaussian was greater than that to either of the smaller two. This shows some summation of the effects of light over an area exceeding 2.1 min. (All these differences are highly reliable.) The difference in sensitivity to the 8.4 and the 30 min Gaussians, however, was not reliable ($t = 0.55$). So we found no evidence of any summation of the effects of light over an area greater than 8.4 min. These findings are in accord with those of Hillman[35], who reported failure of Ricco's law between 2 and 5 min in the fovea, and those of Davila and Geisler[34], who attribute all summation of light within the fovea to pre-neural or optical factors.

Zero frequency. As the spatial frequencies of the gratings in the Gabor patches (section 3.1) decrease towards zero, the profile of a Gabor patch approaches, as a limit, that of the 30 min Gaussian blob. The question we address here is how well the sensitivity to that 30 min Gaussian blob approaches the Platonic ideal of sensitivity at a spatial frequency of zero cycles per degree. To evaluate that, we plot the contrast sensitivity from section 3.1 against absolute frequency instead of its logarithm, so as to bring zero frequency from negative infinity to the lowest point on the graph, as shown below. Here the sensitivity to the 30 min Gaussian is plotted at zero frequency to see whether it is consistent with the rest of the curve. It does seem approximately in line with the rest of the data. However, extrapolation, to zero, of the equation used to fit the CFS in section 3.1 shows that it has an unintuitive minimum at a frequency of 0.5 cpd. An equation, simplified from that of Yang, Qi, and Makous[36], fits the data nearly as well (not reliably worse) and lacks the infelicity at low spatial frequencies:

$$CSF = a\,(e^{-f/b} - c\,/(d + f^2)) \qquad (8)$$

where f is spatial frequency, $a = 239$, $b = 7.28$, $c = 534.5$, and $d = 2.67$. Evidently, the sensitivity to the 30 min Gaussian blob falls close to the extrapolated function, where one might expect sensitivity to zero spatial frequency to fall. If we ask at what position on the x-axis the point should be placed to optimize the fit with the equation, the value is negative, but placing it there does not improve the fit significantly (statistically or otherwise). As a negative spatial frequency is even harder to interpret than zero spatial frequency, we place the point at zero.

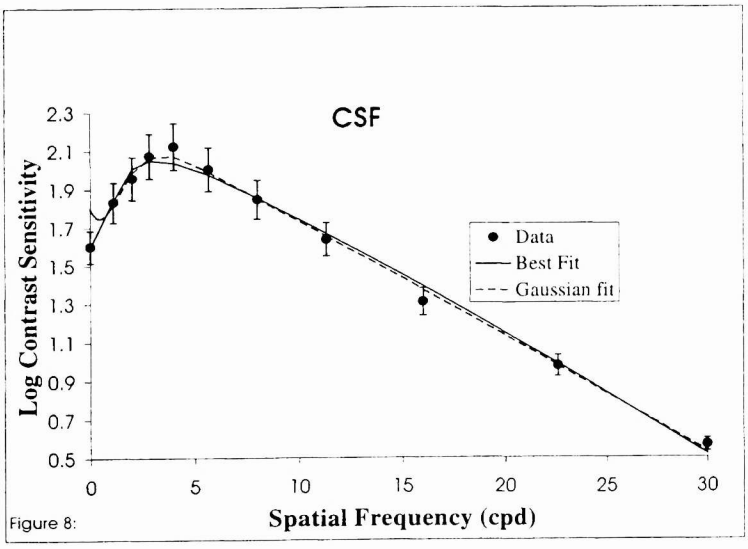


Figure 8:

It may also be noteworthy that the sensitivity to the 30 min Gaussian blob correlates (across observers) with sensitivity to the lowest frequency Gabor patch used but not to the highest spatial frequency Gabor used (ρ = 0.15. This suggests that common mechanisms tend to subserve detection of the lowest frequencies and this Gaussian patch, but that the mechanisms that detect the highest frequency Gabor are independent of those detecting this Gaussian. The case at 4 cpd is different: sensitivity at 4 cpd correlates better with that to the smallest Gaussian blob (ρ = 0.82) than to the largest blob ((ρ = 0.51). This may be because the largest Gaussian blob has practically no energy at 4 cpd, but the amplitude of the smallest blob is about one fourth as great at 4 cpd as it is at its maximum, and the visual system is about 3 times as sensitive at 4 cpd as it is at 0 cpd. So mechanisms sensitive to 4 cpd may contribute substantially to detection of the smallest blob but not to the largest blob.

As the spatial frequency within a Gaussian envelope decreases, the oriented component decreases and ultimately disappears. This should increase the number of mechanisms stimulated but perhaps decrease the excitation of each individual one. The net effect is problematic. It is satisfying, then, that Watson has shown that the sensitivities to all four Gaussian blobs, and most of the other stimuli in the Modelfest set, are well described by a model based on the stimulus contrast filtered by an emprical CSF, raised to the 2.5 power, and then integrated.[13]

3.7 Miscellaneous Patterns

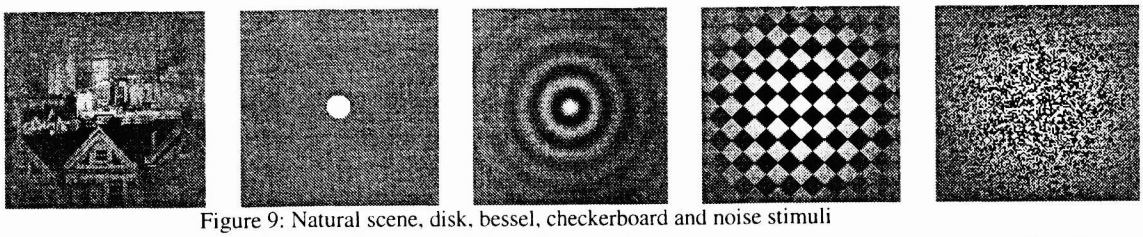


Figure 9: Natural scene, disk, bessel, checkerboard and noise stimuli

By miscellaneous patterns, we refer to the natural image, disk, bessel, checkerboard, and noise (essentially all patterns that were neither Gabors nor multipoles). These patterns were included for various reasons. The first reason was simply so that something other than Gabors and multipoles would be included. The natural image was selected so that at least one "natural" stimulus would be included among the otherwise simple and synthetic collection. The noise stimulus was selected in part because it was the only stimulus whose particular structure could not be preordained by the experimenters. The disk, checkerboard and bessel were included because they have energy at many orientations, and the disk and checkerboard because they have sharp edges. Whatever the initial reasons for their selection, Watson[13] has shown that these stimuli as a class proved to be particularly useful in distinguishing among candidate models. This appears to be primarily because, unlike most of the other stimuli, they are broad-band, that is, their energy is spread over a broad range of spatial frequencies and orientations. Indeed, the noise stimulus, which is the most broad-band of all, proves to be the most effective diagnostic stimulus. Models which did not contain multiple channels tuned for different frequencies and orientations were quite poor at predicting the noise threshold.

Stimuli #35 and #44 were random noise. Each pixel was one min in size, and the noise had a binary rather than a Gaussian distribution. The pattern was then multiplied by our standard Gaussian envelope with a 0.5 deg standard deviation. Stimulus #35 used the same noise throughout, with just the overall contrast changing as part of the threshold seeking staircase. Stimulus #44, on the other hand had a new noise pattern on each trial. It should be noted that only six of the nine observers provided thresholds for #44. As can be seen from the data table, stimulus 44 had the largest SEM of all stimuli. This is partly attributable to the variation inherent in the randomized stimulus, but also partly due to the longer time (for a few laboratories) between trials needed to generate the new stimuli. This could have led to boredom in the subjects and variability in their attention.

The threshold for stimulus #35 of 1.31 corresponds to a 5% contrast threshold. It is interesting to note that the threshold for the line detection of 0.94 corresponds to a 11.4% contrast for a 0.5 min line or 5.7% contrast for a 1 min line. It is noteworthy

that both a 1 min line and a 1 min noise have the same (flat) Fourier spectra and similar contrasts at threshold. This is an extreme example of insensitivity of thresholds to differences in the distribution of the signal over space.

4. CONCLUSIONS

The 'year one' effort of specifying data collection conditions, display characteristics, stimulus specifications and finally collecting the data has been an arduous but rewarding task. The discussions leading up to stimulus selection were often lengthy and sometimes heated, but in the end the final stimuli offered a reasonable balance of the requirements to provide sufficient data to aid in model design and testing without unduly taxing the data collection efforts of the individual laboratories. Where limitations in the stimulus set have been perceived, some groups are gathering additional data which will be posted on our WEB site as soon as they are made available. The research benefits of this exercise are just now being realized with the active modeling efforts of several laboratories that are using this dataset[8,9,10,11,12,13]. As the dataset grows and more modeling results are published the various weakness and strengths of different approaches will become self evident. This is the benefit of using a common dataset for developing and comparing models.

Building on what we have learned in this first data collection phase, future data collection groups will have a much easier time specifying the stimuli and collecting the data. In future years, whenever possible, we will adopt the same psychophysical methods and display specifications. The major task will be to identify the most critical stimuli for designing and testing models for the dimension of the stimulus space under study. A new data collection group has formed to consider of spatio-temporal luminance detection. Potential stimulus sets have been presented at recent meetings of the Modelfest group[4,5]. Interest has also been expressed for establishing a data collection group to consider the critically important area of spatial masking. An accurate model of spatial masking would be invaluable for the image compression community. As the bandwidth demands on the internet are growing at an astonishing rate; improved image compression could have a significant impact on the required bandwidth. Many corporations are working on means of improving image compression technologies: this is an open invitation for those companies to join and participate in the Modelfest activities.

5. ACKNOWLEDGMENTS

We thank the members of the greater Modelfest group who have also contributed to this effort. This research was supported by: Air Force Office of Scientific Research F49620-95; NASA RTOP 548-51-12-4110; NEI RO1-4776; EY-4885; EY-1319.

6. REFERENCES

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. R. Eriksson, B. Andren and K. Brunnstrom, "Modeling the perception of digital images: A performance study," *Proceedings of SPIE, Human Vision and Electronic Imaging III*, ed. B. E. Rogowitz and T.N. Pappas, **3299**, 88-97, 1998.
3. B. Li, G. W. Meyer and R. V. Klassen, "A comparison of two image quality models," *Proceedings of SPIE, Human Vision and Electronic Imaging III*, ed. B. E. Rogowitz and T.N. Pappas, 3299, 98-109, 1998.
4. S. A. Klein, "Modelfest '99 Workshop: Comparing detection models," *Optical Society of America Annual Meeting, , Digest of Technical Papers* pp. SuE., 1999.
5. T. Carney, "Modelfest: Vision Modeling - Progress and future plans", *Investigative Ophthalmology and Visual Science* **40**, insert pp. 9, 1999.
6. T. Carney, "Modelfest Web Site", http://neurometrics.com/projects/Modelfest/IndexModelfest.htm , 1998.
7. A. B. Watson, "ModelFest Web Site", http://vision.arc.nasa.gov/modelfest/ , 1999.
8. L. Walker, S. A. Klein, and T. Carney, "Modeling the Modelfest data: decoupling probability summation," *Optical Society of America Annual Meeting, Digest of Technical Papers*, pp. SuC5., 1999.
9. A. B. Watson, and J. A. Solomon, "ModelFest data: Fit of the Watson-Solomon model," *Investigative Ophthalmology and Visual Science* **40**, S572, 1999.
10. 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* pp. SuC6, 1999.
11. T. Carney, L. Walker, S. A. Klein, "Multi-scale spatial detection model prediction of the Modelfest dataset," *Investigative Ophthalmology and Visual Science* **41**, (submitted) 2000.
12. C. -C. Chen and C. W. Tyler, "Modelfest: imaging the underlying channel structure," "Human Vision and Electronic Imaging IV" Ed: Rogowitz, B. E. Proc. SPIE **3645**, inpress 2000.
13. A. B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models," *Optics Express* **6(1)**, 12-33, 2000 (http://www.opticsexpress.org/oearchive/source/14103.htm).

14. F. W. Campbell and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," Journal *of Physiology (London),* **197**, 551-566, 1968.
15. C. Blakemore and F. W. Campbell, "On the existence in the human visual system of neurones selectively sensitive to the orientation and size of retinal images," Journal *of Physiology (London)* **203**, 237-260, 1969.
16. F. W. Campbell, J.J. Kulikowski and J. Levinson, "The effect of orientation on the visual resolution of gratings," Journal *of Physiology (London),* **187**, 427-436, 1966.
17. R. F. J. Quick, "A vector-magnitude model of contrast detection," *Kybernetic* **16**, 65-67, 1974.
18. C. F. Stromeyer III and S. A. Klein, "Evidence against narrow-band spatial frequency channels: detectablility of frequency modulated gratings," *Vision Research* **15**, 899-910, 1975.
19. U. Polat and C.W. Tyler, "What pattern the eye sees best," *Vision Res.* **39**, 887-?? , 1999.
20. J. P. Thomas, "Model of the function of receptive fields in human vision," *Psychology Review* **77**, 121-134, 1970.
21. N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: a comparison of single channel and multichannel models," *Vision Research* **11**, 251, 259, 1971.
22. C. F. Stromeyer III and S. A. Klein, "Evidence against narrow-band spatial frequency channels: detectablility of frequency modulated gratings," *Vision Research* **15**, 899-910, 1975.
23. J. Nachmias and R. V. Sansbury, "Grating contrast: Discrimination may be better than detection," *Vision Research* **14**, 1039-1042, 1974.
24. C. F. Stromeyer III, and S. A. Klein, "Spatial frequency channels in human vision as asymmetric (edge) mechanisms," *Vision Research* **14**, 1409-1420, 1974.
25. S. A. Klein and C. F. Stromeyer III, "On inhibition between spatial frequency channels: Adaptation to complex gratings," *Vision Research* **20**, 459-466, 1980.
26. A. B. Watson, "Summation of grating patches indicates many types of detector at one retinal location," *Vision Research,* **22**, 17-25, 1982.
27. C. -C. Chen, and C. W. Tyler, "Spatial pattern summation is phase insensitive in the fovea but not in the periphery," *Spatial Vision*, **12**, 267-286, 1999.
28. S. A. Klein, "Visual multipoles and the assessment of visual sensitivity to displayed images," *Proceedings of SPIE, Human Vision, Visual Processing, and Digital Display*, B. Rogowitz and J. Allebach, eds, **1077**, 83-92, 1989.
29. H. R. Wilson, D. K. McFarlane, and G. C. Phillips, " Spatial frequency tuning of orientation selective units estimated by oblique masking," *Vision Research*, **23**, 873-882, 1983.
30. N. R. Bartlett, "Thresholds as dependent on some energy relations and characteristics of the subject," in Vision and visual perception, C. H. Graham, Ed. New York: John Wiley & Sons, pp. 154-84, 1965.
31. E. Baumgardt, "Threshold quantal problems," in Visual psychophysics, vol. VII/4, Handbook of physiology, D. Jameson and L. M. Hurvich, Eds. New York: Springer-Verlag, pp. 29-55, 1972.
32. L. A. Olzak and J. P. Thomas, "Seeing spatial patterns," in Handbook of perception and human performance, vol. 1, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds. New York: Wiley, pp. Chapter 7, 1986.
33. D. R. Williams, D. H. Brainard, M. J. McMahon, and R. Navarro, "Double-pass and interferometric measures of the optical quality of the eye," *Journal of the Optical Society of America A* **11**, pp. 3123-3135, 1994.
34. K. D. Davila and W. S. Geisler, "The relative contributions of pre-neural and neural factors to areal summation in the fovea," *Vision Research.* **31**, 1369-80, pp. 1991.
35. B. M. Hillmann, "Relationship between stimulus size and threshold intensity in the fovea measured at four exposure times," *Journal of the Optical Society of America.* **48**, pp. 422-8, 1958.
36. J. Yang, X. Qi, and W. Makous, "Zero frequency masking and a model of contrast sensitivity," *Vision Research* **35**, pp. 1965-1978, 1995.

---

*Correspondence: Email: thom@neurometrics.com, WWW: http://www.neurometrics.com/modelfest; Telephone: 510-644-3112;