The development of an image/threshold database for designing and testing human vision models.

Thom Carney[ab*], Stanley A. Klein[b], Christopher W. Tyler[c], Amnon D. Silverstein[d], Brent Beutter[e], Dennis Levi[f], Andrew B. Watson[e], Adam J. Reeves[g], Anthony M. Norcia[c], Chien-Chung Chen[c], Walter Makous[h], and Miguel P. Eckstein[i]

[a] Neurometrics Institute, 2400 Bancroft Way, Berkeley, CA 94704
[b] School of Optometry, University of California at Berkeley, Berkeley, CA 94720
[c] Smith-Kettlewell Eye Research Institute, San Francisco, CA 94115
[d] Hewlett-Packard Inc., Palo Alto, CA
[e] NASA Ames Research Center, Moffett Field, CA 94035
[f] College of Optometry, University of Houston, Houston, TX, 77204
[g] Department of Psychology, Northeastern University, Boston, MA 02115
[h] Center for Visual Science, University of Rochester, Rochester, NY, 14627
[i] Department of Medical Physics & Imaging, Cedar Sinai Medical Center, Los Angeles, CA

ABSTRACT

Models that predict human performance on narrow classes of visual stimuli abound in the vision science literature. However, the vision and the applied imaging communities need robust general-purpose, rather than narrow, computational human visual system (HVS) models to evaluate image fidelity and quality and ultimately improve imaging algorithms. Psychophysical measures of image quality are too costly and time consuming to gather to evaluate the impact each algorithm modification might have on image quality.

Several general-purpose early HVS models currently exist but direct comparisons of the models on the same data sets are rarely made, making it difficult to evaluate their utility. Moreover, researchers designing a new model are confronted with the decision of what data set to use to set model parameters. To address these issues about 60 researchers interested in vision modeling have formed a group tentatively called Modelfest. One of the group's goals is to develop a public database of test images with threshold data from multiple laboratories for designing and testing HVS models. The data set will be available on the WEB for all modelers to use in HVS model development. The group may also provide a threshold database with stimulus specifications derived from the existing vision literature for model design and testing. Although the space of possible stimuli is enormous, this first year's data collection effort is limited to detection thresholds for static gray-scale 2D images. In future years, the database may be extended to include discrimination (masking) as well as detection thresholds for dynamic, color and gray scale image sequences. The purpose of this first report is to invite the Vision Science and Electronic Imaging community to participate in this effort and inform them of the developing data set, which will be available to all interested researchers.

This first paper presents the display specifications, psychophysical methods and stimulus definitions for the Year One effort. The threshold data will be collected by each of the authors over the next few months and presented on the WEB along with the stimuli.

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

1. INTRODUCTION

Digital information technology has progressed rapidly from simple ASCII based text communications to complex interactive multimedia presentations with voice, music, text and visual imagery. The information content has grown at a rate that far exceeds the bandwidth of the hardware infrastructure, from the central digital transmission pipes to the desktop computer interface. Visual images are the most demanding component in terms of system bandwidth. Lossless compression technologies have helped reduce the load, but where perfect reproduction is not required, lossy compression methods are necessary to limit the overall bandwidth while maintaining satisfactory image quality. Satisfactory image quality is application specific. Medical image scans typically require very high fidelity while talking head conference calls can get by with much lower image quality and still be considered satisfactory. In either case, the goal is to reduce the bandwidth as much as possible while maintaining the requisite image quality. Successful lossy image compression requires an accurate model of human

Part of the IS&T/SPIE Conference on Human Vision and Electronic Imaging IV • San Jose, California • January 1999
SPIE Vol. 3644 • 0277-786X/99/$10.00

perception, from the limits of visual sensitivity to the cognitive significance of image artifacts.

Over the past 35 years the vision science community has made significant progress in understanding the early stages of visual processing. Visual psychophysics and physiological studies have revealed a multi-stage parallel processing structure of the early visual pathways. The results have led to the development of quantitative HVS models that can predict performance on a variety of acuity tasks. In the past few years HVS models have become increasingly general purpose. Early models were restricted to simple targets presented in carefully controlled conditions[1]. Recent models report success in predicting visual performance with complex static visual stimuli[2,3,4]. Although most models exhibit similarities, such as banks of Gabor filters, they have distinct differences in how they combine filter responses and account for visual masking. Are the model differences significant? Under what conditions does one model perform better than another? These questions are very hard to answer since model performance is often based on comparisons with a limited psychophysical data set. Models are rarely compared using the same stimulus set[5,6]. Our goal is to provide a readily available stimulus database designed to test many different aspects of HVS models. The database will include psychophysical thresholds obtained from several laboratories using the same stimulus set. The stimulus set will grow each year, beginning with a set of achromatic detection targets and expand to include complex video sequences in future years. The database will be useful for building new models as well as comparing the performance of existing models. Once a large readily accessible database of stimuli with psychophysical thresholds exist, the developers of future general purpose HVS models will be compelled to provide performance data using the database images before their model will be taken seriously. It will become easier to determine which model innovations actually improve model performance. While this may appear competitive, it actually provides a method for modelers to learn from each other's approach and to facilitate the development of high quality HVS models that can be used to improve image compression technologies. The computational models will also be valuable as an accurate means of assessing image quality without lengthy psychophysical studies.

Aside from the applied aspects of HVS modeling, this research effort also offers utility for basic vision science research. Applied models used in image compression might use short cut methods to achieve satisfactory results in a time frame that is practical for commercial purposes. Researchers interested in understanding fundamental mechanisms of human vision may use threshold data for building vision models that are consistent with physiological data. In comparing the performance of models, it is important to consider its end application.

This Year One paper describes the data display and collection specifications for the project and the first year's stimulus set. The specifications were based on extended email discussions among Modelfest members regarding advantages and disadvantages of various proposals. The specifications described were voted on and adopted by the Modelfest data collection group. Membership in this data collection group is open to all those willing to collect threshold data once the specifications have been agreed upon. The stimuli and thresholds will be available on the WEB at:

http://neurometrics.com/projects/Modelfest/IndexModelfest.htm

2. DISPLAY AND DATA COLLECITON SPECIFICATIONS

The data collection phase will occur in many laboratories across the country using a variety of display systems. Consequently, the specifications agreed upon are somewhat general so that members can gather data without having to purchase new hardware. We expect these display specifications will be satisfactory for several years of data collection, not just the first year's spatial pattern detection studies.

2.1 Video display specifications and viewing conditions:

In real world applications, the end user viewing conditions will vary somewhat from our specifications, but not significantly in most instances. In general, the Year One threshold data will be collected using standard computer video monitors. Each data collector will specify the specific hardware devices used, which will become part of the database. The following list specifies display conditions required for Year One data collection.

2.1.1 Video display specifications:

2.1.1.1 Display mean luminance: 30 ±5 cd/m^2. While on the dim side, this level is easily achieved by normal monitors, and it allows for luminance losses when pixel replication methods are used for fine gray scale control.

2.1.1.2 Display frame rate: A minimum of 60 Hz is required, higher frame rates are acceptable and encouraged. Sixty Hz is a standard computer display rate. Flicker may be noticeable at this rate, so higher rates are preferred, but not all the display systems available to the data collectors are capable of higher rates.

2.1.1.3 Display pixel size: The image size will be 256 pixels square. The region surrounding the image will be set to the display mean luminance level. The display pixel size is 0.5 min except for cases where pixel replication may be necessary to achieve adequate gray scale resolution, such as for moderately low spatial frequency

stimuli. In the case of pixel replication, the effective pixel size will be 1 min.

2.1.1.4 Display gray scale resolution: The gray scale luminance step size will be 1/4 or less of the final stimulus threshold. Many display systems are limited to 8 bit lookup tables, which is a problem for measuring detection thresholds. Methods such as passive resister networks[7], bit-stealing[8] and virtual pixels[9] exist to extend the hardware lookup table limit to 10 or more bits of luminance control. In future years, the data collection efforts will include color stimuli, therefore the chromaticity of each display at mean luminance will be reported.

2.1.1.5 Stimulus pixel replication: Two or more display pixels can be used to create a virtual pixel that can increase the effective bit depth of the image for improved gray scale resolution and limiting adjacent pixel interaction effects[10]. Stimuli that are essentially one dimensional, such as Gabor patches, will be oriented horizontally when possible to avoid adjacent pixel interaction effects. Other stimuli, such as two-dimensional noise, may require pixel replication or a two-dimensional look up table[11,12] to limit adjacent pixel interaction effects.

2.1.1.6 Stimulus temporal waveform: Each stimulus presentation will last 500 msec. Each stimulus will have a Gaussian temporal window with a 125 msec standard deviation. The relatively short duration was to minimize the burden on data collectors. In future years, we may use longer and shorter windows to explore sustained and transient mechanisms.

2.1.2 Viewing Conditions:

2.1.2.1 Binocular viewing: Observers will view the display with both eyes. Some researchers may elect to include monocular conditions to examine binocular summation issues and to study amblyopia.

2.1.2.2 Natural pupils: Artificial pupils are optional, but may be used to examine the role of pupil size.

2.1.2.3 Fixation: One or two types of fixation target will be used.

1) Prior to the presentation of each stimulus, a one pixel high by two pixels wide fixation pattern will be briefly presented at the center of the stimulus patch. The pattern will be slightly below the background luminance, enough to insure its visibility yet minimize any adaptation effect.

2) A one pixel wide bright square surrounding the central 2-deg wide test patch will remain on during the experiment. This target will be a strong fusion stimulus and will help accommodation.

2.1.3 Data collection and analysis:

When possible, a data collection site will collect data from the same observer over several years so data on detection, masking, motion, color and other aspects will be available from the same subject.

2.1.3.1 Stimulus Presentation: To accommodate the many different laboratories we have agreed on two modes of stimulus presentation, two alternative forced choice or method of constant stimuli. The 2AFC method must be temporal, not spatial, because of screen size limitations.

2.1.3.2 Psychophysical method: Any objective data collection method will be acceptable. Examples include adaptive staircase, Quest, PEST, Psi, and method of constant stimuli. The majority of trials must be located near the final threshold (d' = 1-2) to minimize errors based on assumptions about the slope of the psychometric function.

2.1.3.3 Trials/error bars: Final threshold for a laboratory will be based on a minimum of four runs with stimulus presentation order blocked to avoid order effects. Given the flexibility of the methods adopted (2.1.3.2) we have decided to set a maximum acceptable standard error criteria of 25% of the reported thresholds.. We anticipate that most labs will provide data with average standard errors in the 15% range. The 25% value would be an upper limit for each stimulus.

2.1.3.4 Threshold reporting: To compare findings across laboratories thresholds will be reported at the 84% correct level for the 2AFC method (or d' = $\sqrt{2}$ when using other psychophysical methods).

2.2. Stimulus Set Specifications:

To ensure that enough laboratories participate in the data collection phase, the number of different stimuli selected was limited to 45. The stimuli are organized into 12 categories, with several stimuli doing double-duty by being listed in multiple categories. The majority of stimuli have a Gaussian envelope with 0.5 deg standard deviation to limit edge effects. Many of the Gabor stimuli have a Gaussian envelope with a one-octave bandwidth that results in a fixed size stimulus in terms of carrier frequency cycles. Unless otherwise specified the Gaussian envelope was circular. The first five categories consist of Gabor patches of different bandwidth, frequency, aspect ratios and orientations. The next three groups consist of combinations of Gabor patches with different orientations, spatial frequencies and phase relationships. The Gabor pattern in particular configurations closely match the receptive fields of neurons in visual cortex[13,14]. The spatial filters common to most early vision models approximate the shape of cortical neurons[15]. The bank of Gabor patches will test the model filter parameters, orientation and frequency bandwidth as well as probability summation

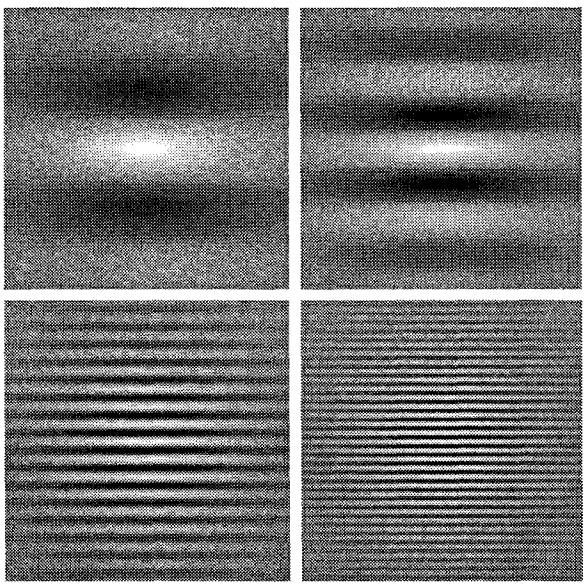
among the mechanisms. The ninth category, Gaussian blobs, examines Ricco's area, very low frequency mechanisms and mechanism integration across orientations. Category ten includes features common to natural scenes: edges, lines and dipoles. These targets along with certain Gabor patterns are also important for estimating mechanism bandwidth. Category eleven and twelve include a variety of complex scenes that combine earlier targets in different ratios at multiple frequencies, orientations and sizes. The categories progress from simple patterns that estimate the spatial sensitivity of the visual system to complex scenes that test how the outputs of the underlying mechanisms are combined.

2.2.1 Gabor patches with fixed size in degrees:

A horizontal Gabor patch, composed of a sinusoidal grating times a Gaussian envelope, is defined as:

$$\cos(2\pi f\, y)\, \exp(-((f\, x/S_x)^2 + (f\, y/S_y)^2)/2) \qquad (1)$$

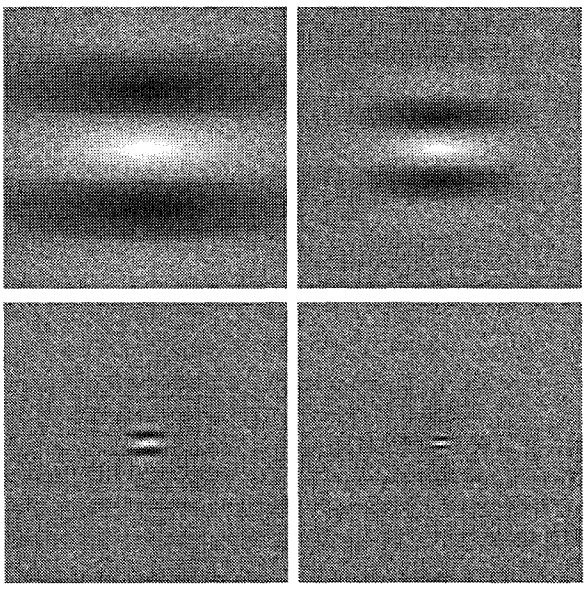
where f is spatial frequency (c/deg) in the y direction and S_x & S_y are the envelope standard deviations in cycles in the x and y directions. In this category, the Gaussian envelope has a constant 0.5 deg standard deviation (S_x/f = S_y/f= 0.5 deg). Data collected using these targets will define the classical contrast sensitivity function for our observers. Carrier spatial frequencies tested are, 1.12, 2.0, 2.83, 4.0, 5.66, 8.0, 11.3, 16.0, 22.6 and 30.0 c/deg. The frequencies are in 1/2 octave steps except for the first and last spatial frequencies. The lowest frequency is chosen to match the one octave bandwidth of the next category of stimuli. The figure below includes the lowest frequency and a few higher frequency Gabor patches. The rectangular background for each patch represents the 128 min square field that all stimuli are constrained to fit within



2.2.2 Gabor patches with fixed size in cycles:

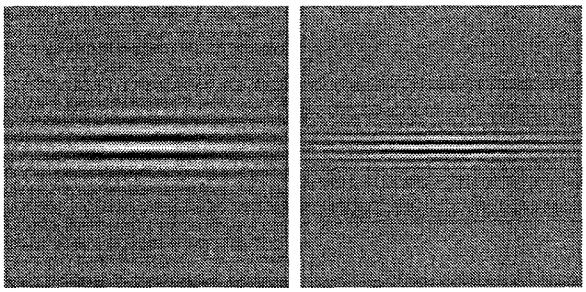
The overall spatial frequency tuning (the CSF) of central vision has been well established for as much as half a century, using sinusoidal grating stimuli of various extents. What has not been carefully evaluated is the tuning for local stimuli of a scale that might be expected to stimulate the most local receptive fields known from visual neurophysiology. One of the closest approaches to this question was by Wilson, McFarlane and Phillips[16], using a stimulus bar with a D6 profile, which is slightly more than one cycle wide at half-height. Their tuning function peaked at 2 c/deg with a maximum sensitivity of 50, substantially lower than for a more extended grating target. However, the bars were elongated in the vertical direction, to different extents at different peak frequencies. So the stimuli were still not as local as possible, and extended over inhomogeneous retina in the perifoveal region. The value for Modelfest is that these local Gabor stimuli are measured at the same spatial frequency as the corresponding full-width Gabors of category 2.2.1. The difference between the categories will thus reveal the extent of summation at each spatial frequency, from essentially a single cycle to the full extent permitted within a reasonable definition of the fovea.

Stimuli in this category are one-octave (full bandwidth at half-height) Gabor patches. The Gaussian envelope, with S_x = S_y = 0.56, results in 1.12 cycles between the ± standard deviation points. The carrier frequencies tested are 1.12, 2.0, 4.0, 8.0, 16.0 and 22.5 c/d. The low frequency target is the same as the low frequency target in section 2.2.1. The figure below shows the stimuli relative to the display window.



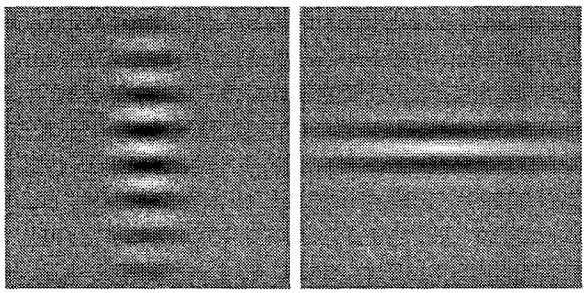
2.2.3 Elongated Gabor patches:

These Gabor patches have a horizontal carrier frequency with Gaussian envelope with standard deviation of 0.5 deg in x. The envelope in y has a half octave bandwidth (S_x = 1.12) corresponding to 2.24 cycles per ± standard deviation points. The carrier frequencies are 4, 8, and 16 c/deg. The purpose of these three stimuli are to provide a comparison to the three multipole stimuli (edge, line, dipole). The ratio of multipole sensitivity to sinusoid sensitivity is directly related to the bandwidth of the underlying mechanisms[17]. The stimuli in this category have the same horizontal extent as the multipole stimuli and they have narrow enough tuning to be close to simulating a full grating.



2.2.4 Gabor patches with different aspect ratios:

This class of Gabor functions, that we call Baguettes, have unequal aspect ratios in two orthogonal directions, so that the envelope is elongated into an elliptical Gaussian, as defined by equation 1.

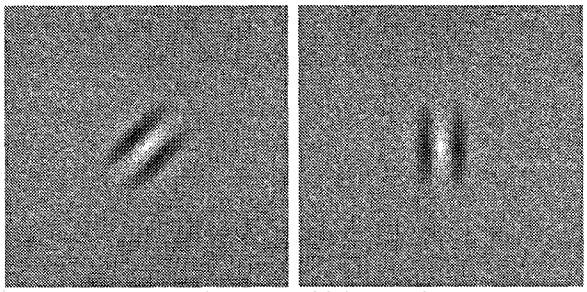


There are four extra stimuli in this category with horizontal and vertical bandwidths of (S_x, S_y) = (1.12, 0.56), (2.0, 0.56), (0.56, 1.12), (0.56, 2.0). The specific case where (S_x, S_y) = (0.56, 2.0) (short stripes across the elongated envelope) has been termed a "tigertail"[18].

The value of baguettes is to allow the evaluation of the two-dimensional structure of the summation units contributing to psychophysical detection mechanisms. Evidence that such detection mechanisms are tuned to both spatial frequency and orientation[19,20] implies that they are selective for particular carrier properties. Given such carrier selectivity, one can then ask what is the extent of the summation field for each carrier selectivity? Is it small (one cycle) or large (many cycles); is it isotropic (circular) or anisotropic (elongated); if elongated, is it aligned with (collinear) or across (orthogonal to) the orientation of the stripes in the carrier? Of course, there may be many such summation fields overlapping for each carrier frequency. In this case, the measured summation curve will reflect the envelope of the set of summation fields at that carrier frequency, but it will characterize the summation extent of the largest fields in each direction, at least. Prior results for such stimuli are controversial. Watson, Barlow & Robson[21] varied aspect ratio around a circularly-symmetric envelope of about 3 cycles diameter at half-height, for a drifting Gabor of 4 c/d, and found that all other aspect ratios were less sensitive than the circularly symmetric one. Polat and Tyler[22] on the other hand, used a core stimulus as small as one cycle diameter at half-height and found that elongation improved sensitivity up to at least 6:1 in the direction collinear with the bars of the carrier. Elongation in the orthogonal direction was less effective, implying pronounced anisotropy in the summation envelopes. Inclusion of this condition in the Modelfest stimulus set will resolve these discrepancies with respect to static stimuli and challenge existing models.

2.2.5 Gabor patches with different orientations:

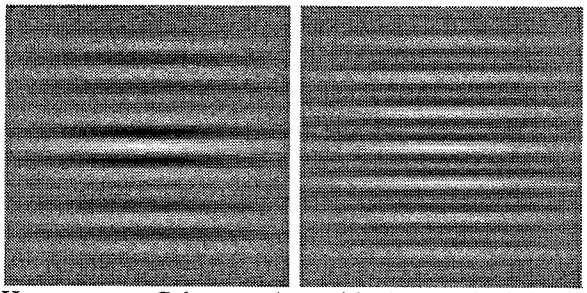
Aside from the standard, horizontal orientation, vertical and oblique 45 deg gratings of 4 c/deg will be tested as well. The so-called oblique effect demonstrates that visual sensitivity is anisotropic[23,24]. An estimate of the magnitude of this anisotropy is required not only for a satisfactory characterization of the system, but it is particularly important for modeling sensitivity to radially symmetrical



stimuli and natural images, which contain components at all orientations, and other stimuli with oblique components, such as the plaids and checks. Evidence that spatial frequency has little effect on the magnitude of the anisotropy[25] suggests that testing at a single spatial frequency is likely to suffice. A test of the anisotropy also requires a vertical stimulus, and these data are particularly important for evaluating sensitivity to other, more complex stimuli containing vertical components, such as plaids.

2.2.6 Gabor patch subthreshold summation:

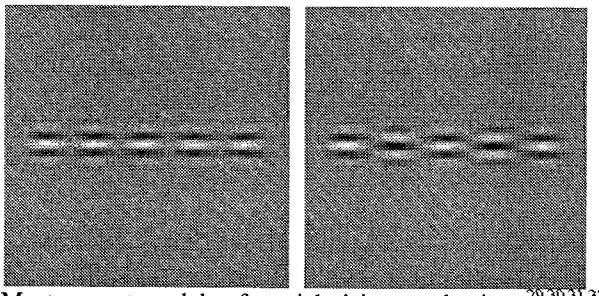
Subthreshold summation between grating or Gabor targets of nearby frequencies has been a traditional means of constraining the frequency bandwidth of the underlying spatial channels[26,27,28].



Here we use Gabor patches with the standard 0.5 deg Gaussian envelope, and spacing of one half and one octave. The experiment is repeated at base spatial frequencies of 2 and 4 c/deg.

2.2.7 Five collinear Gabor patches:

This category consists of two stimuli, both composed of a horizontal string of five one-octave bandwidth Gabor patches with an 8 c/d carrier frequency. Each patch is separated by 5 envelope standard deviations ($S_x = 0.56$ so the center-to-center spacing is 5*0.56/8 deg). The first stimulus is composed of five of the 8 c/deg patches used in category 2.2.2. The second stimulus is the same as the first except successive patches have the carrier grating shifted by 180 deg.



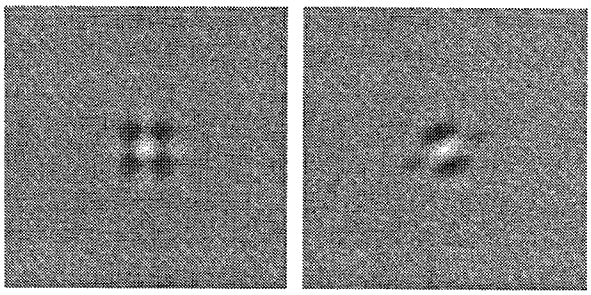
Most current models of spatial vision mechanisms[29,30,31,32] assume that the first stage of visual information processing contains localized linear filters whose spatial sensitivity profile resembles those of simple cells in the striate cortex[33,34]. Each linear filter analyzes a small portion of the whole visual field, which, by analogy to the receptive fields of neurons, is called the perceptive field[35]. Although recent studies show that localized linear filters are inadequate to account for the performance on "complicated" or "higher level" discrimination tasks such as texture discrimination[36,37] and contour integration[38], detecting periodic patterns is a task expected to be governed by the localized linear filters.

Chen & Tyler[39] developed a stimulus to evaluate whether the localized linear filters alone can account for the detection of elongated periodic patterns. They measured the effect of phase coherence on the detectability of a string of Gabors consisting of 1 to 8 vertical Gabor patches aligned vertically and either in phase or 180° out of phase from their immediate neighbors. Summation beyond that attributable to probability summation was observed. In the fovea the phase configuration had no effect on detection threshold. This result cannot be accounted for by linear filters but suggests that a nonlinear transform is required before the pattern detector sums information across space. The Gabor strings therefore provide a challenging test for models of spatial vision.

2.2.8 Plaid patterns:

The principal challenge in generalizing the results of experiments with simple, grating stimuli to more complex stimuli, such as are of practical importance and are encountered in natural images, lies in learning the rules governing the interactions among the components of these images.

The first, basic step in that direction lies in testing the interactions between pairs of gratings with differences in orientation too great to stimulate the same low-level mechanisms. The principles of probability summation alone almost demand that the threshold for the combination differ from that of either alone, but there are, of course, many alternatives to this combination rule, and it is important to get an idea of which are plausible. We propose to start in Year One with plaids in which both gratings are of 4 c/deg; one is horizontal, the other either vertical or oblique 45 deg. Use of these two orientations seems most likely to reveal any anisotropy in the combination rule.



2.2.9 Gaussian blobs:

Pure Gaussian stimuli serve at least two functions in the Modelfest context:

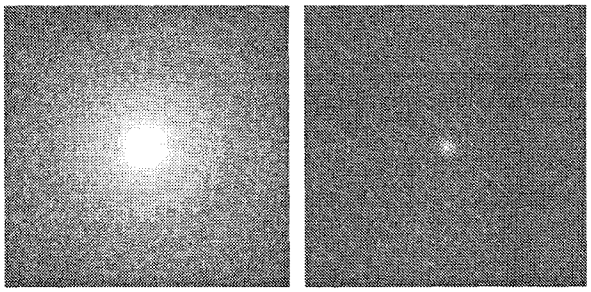
1) The small Gaussian stimuli test for the existence of local detection mechanisms, because they provide stimulation of a local part of the summation field. Detection of such local targets should therefore be inefficient if performed by such one-by-one cycle summation fields. If local detection is found to have full

efficiency, it will imply the presence of summation fields in the form of small circular spots. However, this test is made more difficult by the possibility of probability summation over a set of Gabor filters of different orientations.

2) As non-oriented stimuli, Gaussians constitute a test for the existence of large non-oriented summation fields. Stimuli in this set extend to the largest width compatible with the 2 deg frame of the test field. The large Gaussian targets will place constraints on the sensitivity of the low frequency mechanisms. Some ideal observer models may have trouble with the expected low frequency falloff of mechanism sensitivity.

The most local stimuli used for the study of spatial frequency tuning were those of Tyler, Chan & Liu[40] (1992), who used spots with a circularly symmetric difference-of-Gaussian profile weighted so as to be balanced in mean luminance. They found that sensitivity increased down to a spatial frequency of about 0.5 c/deg. Summation for non-oriented stimuli such as Gaussians may therefore be much more extensive than for oriented Gabor patches, whose sensitivity peaks two octaves or more higher. They therefore test for the summation properties from a small region up to the size of the smallest balanced summation field (such as are used in stimulus set 2.2.2).

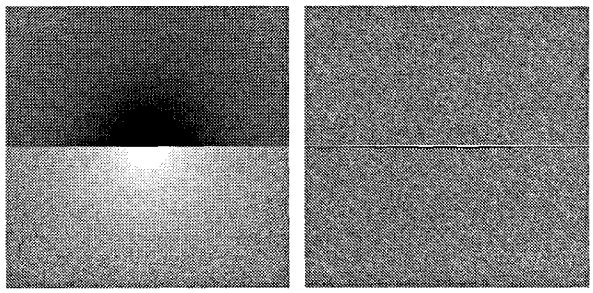
This category includes four circular Gaussian blobs with standard deviations matching the envelopes used in category 2.2.2. The standard deviations are 30.0, 8.43, 2.106, and 1.05 min. The two extremes are shown below.



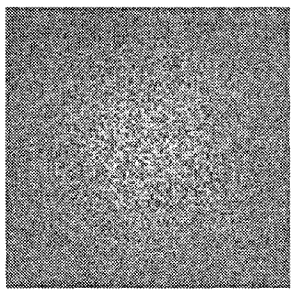
2.2.10 Multipoles - edge, line & dipole:

Multipoles are a family of stimuli where each higher order multipole is the derivative of the previous, in this case going from edge to line to dipole. The line is one pixel wide. Edges and lines are familiar stimuli, the dipole is composed of adjacent opposite polarity lines. The envelope used for this group is the standard circular 0.5 deg Gaussian.

There are several reasons for including the multipoles: 1) They provide the most robust means for measuring the spatial frequency bandwidth of the underlying mechanisms. The ratio of multipole sensitivity to sinusoid sensitivity gives a direct measure of the bandwidth (neglecting probability summation considerations)[17]. 2) The multipoles will test the sampling structure of the models. The model's mechanisms either need to be centered on the multipole or there needs to be enough mechanisms to capture the multipole in an arbitrary position. 3) The edge and dipole stimuli are the only purely antisymmetric stimuli in the Modelfest battery. This will test the relative sensitivity of even and odd symmetric mechanisms.



2.2.11 White noise:



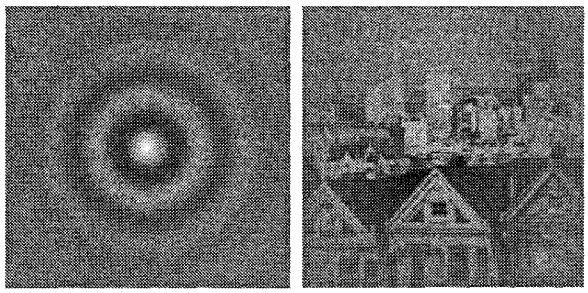
The first stimulus in this category is a static white noise patch with a 0.5 deg Gaussian envelope. The pixel size for this stimulus will be a 1 min effective pixel which is comprised of 4 physical screen pixels. The same noise patch is presented on each trial. The second stimulus in this category is the same except the noise itself changes on each trial. The randomized noise will be generated on each trial based on a random number generator. This is the only stimulus in the battery that has a different pattern from trial to trial. The purpose of the noise stimulus is to test the model's ability to handle complex stimuli with summation across space, spatial frequency and angle. The purpose of the randomized version is to provide a stimulus for testing the effect of stimulus uncertainty. Some of the data collectors plan to use a method of constant stimuli with enough trials so that the slope of the psychometric function can be measured. We anticipate that the fixed noise condition with known phase of the 'hot spot' area will have a shallower psychometric function than the random noise case. Some models may be able to predict this effect.

2.2.12 Miscellaneous patterns

This test battery includes, a Bessel function, a 0.25 deg diameter disk target, a checkerboard, and a natural scene.

Bessel: The 4 c/d Bessel function stimulus, apart from being very different from all our other stimuli, is of interest also because it is a stimulus that is narrow-band in spatial frequency, but broad-band in orientation. It thus tests the degree of summation across many frequency

components at different orientations. The Fourier transform of the Bessel function is a annulus with a Gaussian profile. The maximum of the Gaussain is at 4 c/deg and its standard deviation is $1/\pi$ c/deg.



Disk: The disk is a familiar target used in medical imaging research. The .25 deg disk target also provides another test of summation across orientations or the presence of circular mechanisms.

Checkerboard: Most models operate primarily on the Fourier components of stimuli, but spectrally complex stimuli with unique properties in the spatial domain, such as edges, should not be overlooked. A pattern of checks in a checkerboard is a good stimulus to test for such special effects because its salient properties in the spatial domain are a set of high contrast edges that have no corresponding Fourier components. The width of the checks will be 0.18 deg, so that the fundamental Fourier components have a spatial frequency of 4 c/deg.

Natural image: We have included a natural image (albeit weighted by the standard envelope) in our test battery for several reasons. The first is that we hope ultimately to apply our models to natural images, and it may be wise to test their ability to predict such targets at this time. Second, natural images incorporate the statistics of natural images, while most of our other stimuli do not. A third reason is that, like the static noise image, it is a stimulus for which we have few preconceptions, and whose visibility cannot be predicted without an image-driven model.

3. RESULTS

Table 1 summarizes all the stimulus categories, including number of stimuli in each category and number of stimuli contributing to the total of 45 stimuli. Year One data collection will be completed this April. The results will be available on the WEB and a manuscript of the findings will be submitted soon thereafter.

4. CONCLUSIONS

Computational models of human vision are needed to measure image fidelity for evaluating image compression algorithms. Ultimately, HVS models will be incorporated into the compression algorithms to achieve optimal results. Modeling human vision has advanced to the point that general-purpose models of human vision being developed but direct comparisons are difficult to make. A public image/threshold database is needed to compare HVS models and facilitate model development.

The Modelfest group, which consists of about 60 researchers from industry and academia who are interested in developing accurate models of human vision, have joined forces to develop the needed public image/threshold database. This year's data collection focuses on the detection thresholds of static 2D stimuli. The stimuli were carefully chosen to provide critical data for the creation and testing of HVS models. The stimuli comprise 12 categories ranging from simple Gabor patches to natural scenes.

In future years, the image and threshold database will be expanded to include masking, color, motion and other dimensions of human vision. Achieving consensus on appropriate stimuli across a large number of independent laboratories is a difficult task. However, the results will have significant and lasting impact on vision modeling and its application to image compression, making the effort well worth while. To join the Modelfest group send email to: thom@neurometrics.com.

Stimulus Category	*Stimuli*	*New Stimuli*	*Comments*
Fixed size Gabors (degree)	10	10	1 - 30 c/deg in 1/2 octave steps
Fixed size Gabors (cycles)	6	5	One-octave bandwidth, 1 - 22.5 c/d
Elongated Gabors	3	3	.5 octave by .5 deg envel; 4, 8, & 16 c/d
Gabor Aspect Ratios	7	4	1/1, 2/1, 4/1, 1/2, 1/4, 4/2, & 4/4 ratios
Gabor Orientations	3	2	0, 45 & 90 deg from horizontal
Subthreshold summation	4	4	$f + \sqrt{2}f$ or $f + 2f$, with $f = 2$ or 4 c/d
Gabor - 5 collinear	2	2	Adjacent patch - in or anti phase
Gabor plaids	2	2	2 orientations
Gaussian blobs	4	4	30, 8.43, 2.106, 1.05 min s.d.
Multipoles	3	3	Edge, line & dipole for bandwidths
White noise	2	2	1 min pixel, .5 deg Gaussian envelope
Miscellaneous	4	4	Bessel, natural scene, disk & checkerboard
Total		45	

Table 1 - Summary of Stimulus Categories

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.

REFERENCES

[1] S. A. Klein and D. M. Levi, "Hyperacuity threshold of one second: theoretical prediction and empirical validation," *J. Opt. Soc. A.* 1170-90, 1985.

[2] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," *Digital Images and Human Vision*, ed. A. B. Watson, MIT Press, 163-178, 1993.

[3] S. Daly, The visible differences predictor: an algorithm for the assessment of image fidelity, *Digital Images and Human Vision*, ed. A. B. Watson, 179-206, MIT Press, Mass, 1993.

[4] C. Lambrecht, "A working spatio-temporal model of the human visual system for image restoration and quality assessment applications," *Proceeding of IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, 2291-2294, Atlanta, GA, 1996

[5] R. Eriksson, B. Andren and K. Brunnstrom, "Modelling 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, SPIE, Bellingham, 1998.

[6] 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, SPIE, Bellingham, 1998.

[7] D. G. Pelli and L. Zhang, "Accurate control of contrast on microcomputer displays," *Vision Research,* 31, 1337-1350, 1991.

[8] C. W. Tyler, H. Chan, L. Liu, B. McBride, and L. Kontsevich, "Bit-stealing: how to get 1786 or more grey levels from an 8 bit color monitor," *SPIE Proceedings Human Vision, Visual Processing and Digital Display III,* ed. Rogowitz, B. 1666, 351-364, SPIE, Bellingham, 1992.

[9] S. A. Klein, and T. Carney, "How the number of required gray levels depends on the gamma of the display," *Society of Information Display Digest*, 623-626, 1991.

[10] A. C. Naiman, and W. Makous, "Spatial non-linearities of grayscale CRT pixels," *SPIE Proceedings*, 1666, 1-15, 1991.

[11] S. A. Klein, Q. J. Hu, and T. Carney, "The adjacent pixel nonlinearity: problems and solutions," *Vision Research*, 19, 3167-3182, 1996.

[12] T. Carney and S. A. Klein, "Gray scale adjacent pixel luminance nonlinearity compensation with three color guns," *SPIE Proceedings Color Imaging: Device-Independent Color, color Hard Copy, and Graphic Arts II,* Ed. G. B. Beretta and R. Eschbach, B. 3018, 188-197, SPIE, Bellingham, 1997.

[13] M. A. Webster and R. L. De Valois, "Relationship between spatial-frequency and orientation tuning of striate cortex cells," *J. Opt. Soc. Amer, A* 2, 1124-1132, 1985.

[14] J. D. Daugman, "Two dimensional spectral analysis of cortical receptive field profiles," *Vision Research* 20, 847-856, 1980.

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

[16] 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.

[17] S. A. Klein "Visual multipoles and the assessment of visual sensitivity to displayed images." *Proceedings of SPIE, Human Vision, Visual Processing and Visual Display.* ed. B. E. Rogowitz and T.N. Pappas, 1077, 83-92, SPIE, Bellingham, 1989.

[18] M. J. Morgan and C. W. Tyler "Mechanisms for dynamic stereomotion respond selectively to horizontal velocity components," *Proceedings of the Royal Society B.* 262, 371-376, 1995.

[19] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *J. Physiology* 187, 437-445, 1966.

[20] C. Blakemore and F.W. Campbell "On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images," *Journal of Physiology (Lond)* 203, 237-60, 1969.

[21] A. B. Watson, H. B. Barlow, and J. G. Robson, "What does the eye see best?" *Nature* 302, 419-422, 1983.

[22] U. Polat and C. W. Tyler "What pattern the eye sees best," *Vision Research* (1999 in press).

[23] 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.

[24] M. M. Taylor, "Visual discrimination and orientation." *Journal of the Optical Society of America*, 53, 763-765, 1963.

[25] C. W. Tyler and D. E. Mitchell, "Orientation differences for perception of sinusoidal line stimuli," *Vision Research*, 17, 83-88, 1977.

[26] A. B. Watson, "Summation of grating patches indicates many types of detector at one retinal location," *Vision Research*, 22, 17-25, 1982.

[27] M. B. Sachs, J. Nachmias, and J. G. Robson, "Spatial frequency channels in human vision," *Journal of the Optical Society of America*, 61, 1176-1186, 1971.

[28] N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models," *Vision Research*, 11, 251-259, 1971.

[29] S. A. Klein, and D. M. Levi, "Hyperacuity thresholds 0f 1.0 second: theoretical predictions and empirical validation." *Journal of the Optical Society of America A*, 2, 1170-1190, 1985.

[30] R. J. Watt, and M. J. Morgan, "A theory of the primitive spatial code in human vision." *Vision Research*, 25, 1661-1674, 1985.

[31] A. B. Watson and J. S. Solomon "Model of visual contrast gain control and pattern masking." *Journal of the Optical Society of America A*, 14, 2379-2391, 1997.

[32] C. C. Chen, J. M. Foley, & D. H. Brainard, "Detection chromatic patterns on chromatic pattern pedestals." *IS & T Proceedings: Optics & Imaging in the Information Age*, 19-24, 1997.

[33] D. H. Hubel, and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex." *Journal of Physiology*, 195, 215-243, 1968.

[34] G. C. De Angelis, I. Ohzawa, and R. D. Freeman, "Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development," *Journal of Neurophysiology*, 69, 1091-1117, 1993.

[35] L. Spillmann, A. Ransom-Hogg, and R. Oehler, "A comparison of perceptive and receptive fields in man and monkey," *Human Neurobiology*, 6, 51-62, 1987.

[36] G. Sperling, "Three stages and two systems of visual processing." *Spatial Vision*, 4, 183-207, 1989.

[37] N. Graham, J. Beck, and A. Sutter, "Nonlinear processes in spatial frequency channel models of perceived texture segregation." *Vision Research*, 32, 719-743, 1992.

[38] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: evidence for a local 'association field'," *Vision Research*, 33, 173-193, 1993.

[39] C. C. Chen, and C. W. Tyler, "Spatial pattern summation is phase insensitive in foveal but not in peripheral vision," *Spatial Visi*on, 1999 (in press).

[40] C. W. Tyler, H. Chan and L. Liu,. "Different spatial tunings for ON and OFF pathway stimulation," *Optometry and Physiological Optics,* 12, 233, 1992.

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