Human Chromatic Contrast Sensitivity and Mean Colour

R. Victor Klassen and Nancy Goodman
J.C. Wilson Center for Research and Technology, Xerox Corporation
Webster, New York

Abstract

We report results of an experiment measuring contrast sensitivity as a function of spatial frequency, location in colour space and direction of variation. Stimuli comprised sinusoidal gratings on one side and a uniform field of the same mean colour on the other. The experiment used forced choice to measure thresholds for many observers and numerous spatial frequency-mean colour-direction of variation combinations. Earlier data covered the spatial frequency range from 0.44 to 20 cycles per degree (cpd) and suffered from quantization effects at peak sensitivities.[1] That data showed no meaningful dependence on any other independent variables, such as C*, as would have been expected from common colour difference metrics. The current experiment corrects the quantization effects, seeks more explicitly chroma or hue dependence and extends the low spatial frequency range to 0.19 cpd. We have found only a weak dependence on hue (in addition to the dependence on spatial frequency).

Introduction

We are interested in the limits of human perception, particularly as they apply to print quality. We seek the equivalent of the "CIE standard" observer for more complex images than pairs of flat squares on constant backgrounds. The contrast sensitivity function (CSF)—the contrast at which a sinusoidal grating of a given spatial frequency is visible—provides a partial answer.

Many experiments have measured luminance-based contrast sensitivity; those few that measured chromatic contrast sensitivity generally used a mean colour along the observers' red-green or blue-yellow axis, and varied the colour perpendicular to the axis or about a neutral mean colour.[2,3,4] An exception is Poirson and Wandell's experiment,[5] involving two observers, in which they found that spatial frequency response could be separated from colour sensitivity. Typically, great care was taken to use the observer's opponent channel axis, so that no excitation of the blue-yellow channel occurred when studying the red-green variation, and vice-versa. We varied colour in all directions about each base colour.

The vast majority of published experiments that measured the CSF involved small numbers of observers (typically 1-3). An exception is the ModelFest[6] data[†] to which 16 observers contributed. Because we used volunteer observers, we collected a small amount of data from each of many, rather than a larger amount from a few observers. Over 100 observers collectively provided over 100,000 samples contributing to over 1000 population threshold estimates.

Prior work primarily falls into two broad categories: colour difference and colour matching experiments, and contrast sensitivity experiments. Colour difference and colour matching experiments motivated the standard colour difference equations ΔE_{CMC},[7] CIE ΔE_{94}[8] and CIE ΔE_{2000}.[9] All three colour difference metrics are based on the L*a*b* colour space, which is intended as a good compromise between computability and visual uniformity. These metrics are specific to solid patches. Colour difference metrics provide guidance for color matching, but give limited insight into perception of colour variation.

Campbell and Robson[10] performed early luminance-based contrast sensitivity experiments, to test linearity in the human visual system. Savoy and McCann[11] found that for fewer than five cycles the sensitivity is driven by cycle count as well as by spatial frequency. Mullen[2] measured red-green and yellow-blue sensitivity, finding low-pass, rather than band-pass behavior. Guth[12] hypothesized that chrominance sensitivity is band-pass at high base chroma, and there is little data to support or refute that suggestion. Contrast sensitivity has been measured in red-green and blue-yellow channels; yet colour difference metrics are based on variation in hue and chroma (weighted differently), rather than the opponent axes. This experiment attempts to determine which are the appropriate primaries for spatial variation and whether the chromatic response is band-pass or low-pass.

Goodman[13,14] published thresholds to variations in printed colour caused by mass variations in primary separations and a few mixed colours. Since lightness dominates the visual response in the spatial frequencies studied (0.14-14 cpd), the response to cyan and magenta mass variation was similar to that in black. The broad peak

[†] www.neurometrics.com/projects/Modelfest/resultsModelfest.htm

in sensitivity at 0.7-5 cpd agrees with other published results. At the peak sensitivity, the population median threshold to variation was $\Delta L^*\sim0.3$ peak-to-peak. For yellow, the peak sensitivity was at a spatial frequency of 0.7 cpd with threshold $\Delta C^*\sim4$ peak-to-peak.

Method

The experiment was designed to run completely under computer control, with a compromise between numbers of samples taken, numbers of observers and numbers of conditions. Given the inter-and intra-observer variance observed during pilot runs, it was clear that for good inter-observer statistics we would need at least 10-20 observers per condition. This was impractical for the number of conditions used. Instead, we used 3-5 observers per condition and relied on smooth behavior among neighboring conditions. The results (see below) show that this did not introduce additional error.

For reasons detailed below, there were more than 300 new conditions (in this phase). As in the first phase of this experiment,[1] we used two-alternative forced choice, without feedback. The grating appeared randomly on the left or the right. The contrast was taken (again randomly) from one of two sequences, one ascending and one descending. The entire range of contrasts was divided into 50^{ths}; we used the even half of these for one sequence and the odd half for the other.

Data Display

Images were computed on a SUN Microsystems ULTRA 2 computer and displayed on a calibrated SUN model GDM20E20 monitor at 9500K using factory settings for brightness and contrast. Images consisted of a solid rectangle subtending 20.4^o vertically by 23.3^o horizontally of a selected base colour with a sinusoidal grating in either the left or right half of the field. Grating contrast varied according to a Gaussian horizontally, fell to zero linearly in the top and bottom 10%, and was otherwise constant. No visible artifacts resulted from the slope discontinuity of the amplitude. The grating width at half contrast was approximately 5.5^o. The grating had mean colour equal to the base colour and varied in one of L*, a*, b*, C* or H_{ab}.

The threshold contrast varies widely over the range of colours and spatial frequencies used. In order to allow both high and low contrasts, an image at a multiple of the estimated threshold contrast was computed and stored in the frame buffer each time the condition was changed. This took several seconds, while changing the contrast required only the time of one video refresh. Different lookup table entries were used for the two gratings so that the contrast of each could be set independently.

Grating images were initially computed to 16-bit precision and then converted to lookup table entries using a variant on Floyd-Steinberg error diffusion.[15] A pixel was computed by selecting the nearest entry and subtracting the error incurred from the next pixel in sequence. Any given grating image contains only colours along a line in colour space between two extreme points. The lookup table entries were chosen to be the six 8-bit-quantized points closest to that line, with three just beyond each of the extreme points.

The monitor was calibrated according to the method described previously.[16] We used the same monitor and found only modest drift over the period of nearly a year that elapsed between the two experiments.

Conditions

Observers viewed the monitor from a distance of 0.78 m; the gratings had a height of 25cm resulting in an angular subtense of 18.2^o. According to Rovamo et al.,[17] decreased image size, below a critical area decreases sensitivity. From Savoy and McCann's work,[11] we concluded that with five cycles or more, this effect would be negligible. In phase one, images similar to the higher spatial frequency phase two stimuli were shown at a viewing distance of 1.75m and corresponding smaller angular subtense (and cycle count).[1] For all spatial frequencies used, at least 4 periods were visible (3.5 at maximum amplitude). We anticipated a small reduction in sensitivity at the lowest cycle counts, which should show up as a higher sensitivity for the spatial frequencies that were repeated in this phase, as there were more periods. As will be discussed later, the effect was more than anticipated. Illumination was approximately 125 cd/m^2 (normal office lighting cool white fluorescent).

Gratings were displayed at 0.197, 0.290, 0.426, 0.626, 0.918, 4.26 and 9.18 cpd, with most of the conditions for this phase at the lower four spatial frequencies. For each base colour, as long as monitor gamut limitations permitted, we measured variation in all five directions, except where one or more dimension was meaningless or redundant with another. That is, for colours near neutral, we did not vary hue or chroma; for colours on or near the a* (b*) axis we varied only one of hue and b* (a*), and only one of chroma and a* (b*).

For the first phase, we used colours spread uniformly through colour space. Not having found any dependence on base colour, we chose colours to emphasize the peaks and troughs of the ΔE_{CMC} hue dependence for this phase. Several values of C* were used for each hue. While the hue dependence of CMC has been questioned, at the time it was the best we had, since the ΔE_{2000} data were not available. Base colours were constrained to be within both monitor and xerographic gamuts, allowing for variations of twice threshold in each direction.

Each condition was a combination of spatial frequency, base colour and what was varied. Conditions were selected apparently at random from a list. At least three observers saw each condition. Repeating a small subset of conditions allowed us to quantify inter- and intra-observer variation.

Observer Demographics

Observers were volunteer technical employees at our site. Thirty observers participated in this phase, with ages ranging from 18 to 58. One third of the observers completed one session (20-30 minutes) covering up to 15 conditions. Another third did two or three sessions, completing up to 3-

4 dozen conditions. No observer contributed more than 13% of the data. The gender mix had males contributing 2/3 of the total observations, but was less biased than the workforce.

Only colour normal observers participated, as verified with Ishihara's tests for colour deficiency.[18] Note that especially in the male population there is considerable variation in the red-green channel even within the 90% of colour-normal observers. It would not be surprising if this created some variability in the observations.

Data Analysis

At each response, the control program recorded whether the observer had answered correctly, and the L*a*b* of the two extreme colours displayed. These might be different from the L*a*b* min and max values requested since the requested values were converted to RGB and quantized to 16 bits. The limit of our precision is approximately 0.01 to 0.03 ΔE_{ab}, depending on the location in colour space.

We analyzed the data within session and pooled over all sessions. To pool the data we sorted all responses by condition before processing further. We report results based on the pooled output. The ensemble threshold response was found, along with lower and upper bounds for each threshold using the same methods as in the previous phase.[1] In brief, we found three points. At the threshold estimate the group of observers answered correctly 75% of the time. Between the other two points, the responses appeared to be drawn from a distribution containing 75% correct answers. At the upper and lower bound points, we have 95% confidence that the answers are no longer drawn from such a distribution.

Results

From the data collected, we estimated the percentage error implied by the bounds for each threshold. For a small set of samples, there is sufficient data to estimate inter- and intra-observer variation.

Defining the error as the arithmetic mean between the percent error of the upper bound and the percent error of the lower bound (both taken relative to the estimate), the mode error was around ±25%, and in 95% of all cases, the error was less than ±45%. These error bounds are about typical for contrast sensitivity data. The Modelfest[6] data show a range of variation of about a factor of 3 between lowest-threshold and highest-threshold observer in a group of sixteen observers. This corresponds to approximately ±70% about the geometric mean. Both inter- and intra-observer variation tended to be in the range of 50-75% of the mean—typically larger than the aggregate error bound. When combining responses across locations in colour space, we took the mean across locations, and divided the average errors by the square root of the number of samples averaged, in analogy with the standard error of the mean.

Dependence on Cycle Count

According to Rovamo *et al*,[17] contrast sensitivity depends not only on spatial frequency but also on grating area. Elsewhere,[19] we show how to convert from measured contrast sensitivity to theoretical maximum, grating-size (A) independent, contrast sensitivity, (S) using units of Michelson contrast or ΔE, which we show are nearly equivalent.

$$S_{\max} = S\sqrt{1 + A_c(f)/A}\ , \tag{1}$$

with

$$A_c(f) = A_0(1 + f^2/f_0^2)^{-1},$$
$$A_0 = 269 \text{ and } f_0 = 0.650.$$

We collected data using a monitor with a white point of 56.1 cd/m^2 and a black point of 0.17 cd/m^2. We compared the contrast data obtained by the Modelfest consortium[6] and by Campbell and Robson[10] to our own. We used our white point and an assumed 28.135 mean luminance to convert their contrast data to L*.

Two threshold sequences in the Modelfest data are comparable to our data. The "fixed size" data sample spatial frequencies from 1.12 to 30, while the "fixed cycle count" data sample spatial frequencies 1.12 to 16. These are identical for the lowest spatial frequency because it is the same stimulus. At higher spatial frequencies the two curves diverge; applying equation 1 (above), the two curves line up remarkably well (as shown in Figure 1).

Campbell and Robson's data is some of the oldest contrast sensitivity data available. Spatial frequencies range from 0.2 to 44 cpd. At lower frequencies they used a 10×10° aperture, while at higher spatial frequencies they used a 2×2° aperture. A significant overlap allows us to compare the two. They used a mean luminance 500 cd/m^2, high enough to cause a slight shift in the peak and a factor of 3 increase in sensitivity in the high spatial frequency range (Van Nes and Bouman[20]). At the low frequency end we expect no luminance-dependent sensitivity change.

The two series diverge significantly at spatial frequencies below 9 cpd. Applying equation 1 brings them closer together, but still leaves them a factor of two apart at 1 cpd.

Phases 1 and 2 differed in two critical respects. First, we changed the viewing distance to accommodate lower spatial frequencies at reasonable cycle counts. The closer display may have had an effect on adaptation, as more of the field of view was the same mean colour. In phase 1 we measured sensitivity from 0.44 to 20.3 cpd, while in the second phase we measured from 0.20 to 4.26 cpd. We offset similar spatial frequencies in the two to make them easy to see when plotted together. Second, we changed the display method to overcome the quantization problem we had seen for many of the phase 1 spatial frequencies near peak sensitivity. This affected significant numbers of samples for which L* was varied at all but the two highest spatial frequencies. Deleting only those data points with quantization problems would leave a biased sample (only the samples with low sensitivity at a given spatial frequency would remain). Therefore only the two highest frequency points from phase 1 along with all the points from phase 2 are shown in Figure 1. The adjustment has no (visible) effect at the high frequency end of the phase 1 data, however it does raise the threshold at the low end.

Figure 1 shows all three groups of (adjusted) data. We expect the Campbell and Robson data to be higher at the high frequency end, and all three groups to converge at low frequencies. Because the Modelfest data stops at 1 cpd it is hard to say whether it would converge, but not unlikely. Our high frequency data matches the Modelfest data well, and the low frequency data matches Campbell and Robson's low frequency data. Oddly, all the phase 2 data matches Campbell and Robson's, whereas we would expect only the low frequency data to match.

We cannot justify applying the same adjustment to chrominance data to account for cycle count. Chrominance channels have different bandwidths, so even assuming a similar mechanism, it may have a different critical area. However, not adjusting creates an obvious discontinuity between phases. The adjustment reduces, but does not eliminate the discontinuity. There were relatively few quantization problems with the chromatic channels, but there were some. This may partly explain the difference; chromatic gratings may have a different grating area dependency, or it may be a difference in adaptation.

L* Variation

Figure 1 shows L* variation sensitivity as a function of spatial frequency (both axes are logarithmic). We found no evidence that sensitivity to L* variation depends on anything other than spatial frequency. The inter-condition (within frequency) variation is no greater than inter-observer variation (within condition). Sensitivity has a broad peak at around 1.5 cpd, with a threshold of ΔL* ~ 0.05 (peak to peak).

Chroma (C*) Variation

Figure 2 shows C* variation as a function of spatial frequency. Two sets of points are shown (phase 1 and phase 2), and a combined series adjusted for dependence on hue (see below) provides the data for the trendline. There may be a peak sensitivity in the neighborhood of 0.3 cpd; it is difficult to discriminate based on the data, whether the behavior is low-pass or band-pass. Regardless, the sensitivity at all measured spatial frequencies appears to be lower than the sensitivity to L* variation. The highest sensitivity is ΔC* ~0.2 for all spatial frequencies below 0.5 cpd, within measurement error.

Guth suggested that the low-pass behavior for chromatic variation should disappear (becoming band-pass) for sufficiently chromatic average colours[12]. Even when we consider only the most chromatic group of colours, we cannot say, on the basis of this data, whether chroma sensitivity is low pass or band-pass.

Hue (H_{ab}) Variation

Figure 3 shows the dependence of hue sensitivity on hue. (As with chroma sensitivity, hue sensitivity shows no dependence on chroma or lightness). The curve shown was computed by fitting the data to a function of the form $1+\sum A_n \cos(H_{ab} n\pi/180+\varphi_n)$. We increased the limit to n until the root-mean-squared (rms) error improved less than 1% when the next term was added. The best-fitting coefficients were

n	1	2
A_n	0.076	0.321
φ_n	2.750	-1.487

The rms error for this set of coefficients is 0.50; the correlation coefficient is r=0.40. We would welcome additional data concerning the presence or absence of a dependence of H_{ab} threshold on hue. A similar fit for a ΔC* dependence on hue had an rms error of 0.74, and r=0.46. Despite the slightly higher correlation coefficient, the dependence was less compelling visually and the rms error is nearly 50% larger.

Figure 4 shows the frequency dependence of sensitivity to hue variation. At peak sensitivity, threshold appears to be approximately ΔH_{ab}=0.3. As with C* variation, it is not possible to say whether the behavior is low-pass or band-pass.

Opponent Channel Variation

Our data for a* variation appears in Figure 5. It is qualitatively very similar to both the hue and the chroma variation data, including the amount of spread in the data. Unlike the hue and chroma variation data, the a* data showed no discernable dependence on hue. Like all the other data, it showed no discernable dependence on chroma. At the maximally sensitive spatial frequencies, threshold is Δa* ~ 0.25.

Our b* variation data appears in Figure 6. The data suggests the possibility of a peak in sensitivity at around 0.43 cpd, a peak which is not nearly so strongly suggested by the other chromatic variation data. The maximum sensitivity for Δb is Δb* ~ 0.2.

Chroma Dependence

There remains no dependence on C* in our data. The CIE ΔE_{94} colour difference metric indicates that for a constant amount of visible change in C*, the amount of actual change in C* increases linearly from the neutral axis outward. This results in a factor of four difference from C* = 0 to C* = 66. ΔE_{CMC} indicates a similar magnitude dependence of C* sensitivity to base C*. We have several colours at each of these extreme values, yet at only one spatial frequency does a line of best fit have $r^2 > 0.1$: it has an r^2 of 0.217, but a slope -0.12, which indicates slightly greater sensitivity at higher chroma.

We believe the discrepancy over dependence on base C* results from differences in viewing conditions. These images vary spatially, unlike the pairs of uniform squares used in experiments to determine ΔE_{94} and ΔE_{CMC}. More importantly, the surround was the mean colour of the test samples, whereas in the other experiments the surround was a constant neutral, having much less contrast with the lower C* test samples. This enhanced the observers' ability to see colour differences. Klassen's work presented at AIC'01 confirms that the C* dependence found in colour difference metrics could be equally well cast as a dependence on difference from surround.[21]

Conclusions

We have reported final results of an experiment to measure ensemble spatial frequency response for a large group of observers for an arbitrary base colour and with colour variation along any of L*, a*, b*, C*, and H_{ab}. We believe that the lack of dependence on base C* is due to differences in viewing conditions between our experiment and typical experiments used to collect data that drives colour difference metric equations, as explained above.

We continue to find that at all but the lowest spatial frequency, L* differences are detected 3 to 10 times as easily as differences in any of the chrominance directions. While the band pass behavior of the luminance detection mechanism was verified, we can not tell whether the chrominance channels are band pass or low-pass. This may be a result of noise, or a lack of sufficient data at very low frequencies. For all conditions studied in this monitor-based experiment and previously published print-based experiments, the results agreed.

Sensitivity to L* variation is greatest at around 1.3 cpd, with a threshold of $\Delta L^* \sim 0.05$ peak to peak. Sensi-tivity to chromatic variation in any direction is greatest at or below approximately 0.5 cpd. Minimum threshold is $\Delta E \sim 0.25$ in all chromatic directions in colour space.

References

1. R. V. Klassen and N.B. Goodman, "Human Chromatic Contrast Sensitivity: Exploration of Dependence on Mean Color", *Proc 8th Color Imaging Conference*, Scottsdale, AZ, November 2000.
2. K. T. Mullen, "The contrast sensitivity of human color vision to red-green and blue-yellow chromatic gratings", *J. Physiology*, **359** pp. 381-400, 1985
3. E. Switkes and K. DeValois, "Luminance and chromaticity interactions in spatial vision", *Colour Vision*, pp. 465-470, 1983
4. P. E. King-Smith and D. Carden, "Luminance and opponent color contributions to visual detection and adaptation and to spatial and temporal integration", *Journal of the Optical Society of America*, **66** pp. 709-717, 1976
5. A. P. Poirson and B. Wandell, "Appearance of colored patterns: Pattern-color separability", *J. Optical Soc. of America*, **10** (12),pp. 2458-2470, 12/93
6. T. Carney, C. Tyler, A. B. Watson, W. Makous, B. Beutter, C-C. Chen, A. Norcia, and S. Klein, "Modelfest — first year results and plans for year two" *in Proc. of IS&T/SPIE Human Vision and Electronic Imaging V*, San José **3959,** 2000
7. F.J.J. Clarke, R. McDonald and B. Rigg, "Modification to the JPC79 colour-difference formula", *Journal of the Society of Dyeists and Colourists*, **100**, p. 130, April 1984
8. R. S. Berns, "The Mathematical development of CIE TC 1-29 proposed color difference equation: CIELCH" , *Proceedings of AIC color 93*, **B** (c19-1-4), 1993
9. M. R. Luo, G. Cui and B. Rigg, "The development of the CIEDE2000 colour-difference formula", *Color Research and Applications*, **25**, pp. 340-350, 2001.
10. F. W. Campbell and J. G. Robson, "Application of Fourier Analysis to the Visibility of Gratings", *Journal of Physiology* **197** pp.551-566, 1968
11. R. L. Savoy and J. J. McCann, "Visibility of low spatial frequency sine-wave targets: dependence on the number of cycles". *Journal of the Optical Society of America* **65**(3), pp. 343-350, 1975
12. L. Guth, "The chromatic sensitivity myth", *Proc. CIC 5* 1997.
13. N. Goodman, "Human sensitivity to within page color uniformity", *Proc. SPIE/IS&T Electronic Imaging '97 Conf.: Human Vision and Electronic Imaging* II, 139-146, 1997.
14. N. Goodman, "Perception of Spatial Color Variation Caused by Mass Variations about Single Separations", *proceedings IS&T's NIP14, International Conference on Digital Printing Technologies*, pp. 556-559, 1998.
15. R. W. Floyd and L. Steinberg, "An adaptive algorithm for spatial greyscale", *Proc. SID* **17** pp. 75-77, 1976.
16. R. V. Klassen, K. Braun, R. Buckley, and K. Janamanchi, "Visibility of thin lines on coloured backgrounds", *Proc. 6th Color Imaging Conference*, pp. 27-32, 1998
17. J. Rovamo, O. Luntinen and R. Näsänen, "Modelling the Dependence of Contrast Sensitivity on Grating Area and Spatial Frequency", *Vision Res.*, **33**, pp. 2773-2788, 1993
18. S. Ishihara, "The series of plates designed as a test for color-blindness", Kanehara Shuppan Co., Tokyo, Japan, 1993
19. R. V. Klassen and N. B. Goodman, "Measuring human chromatic contrast sensitivity in an industrial setting", *Proceedings of the CIE Expert Symposium on Temporal and spatial aspects of light and colour perception and measurement,* Veszprém, Hungary, August 2002.
20. F. L. van Nes and M. A. Bouman "Spatial modulation trans-fer in the human eye." *J. Opt Soc Am.* **57**, pp 401-406, 1967
21. R. V. Klassen, "Colour difference metrics and surround effects: preliminary results", *Proc. AIC 2001*, Rochester NY June 2001.

Biography

R. Victor Klassen received a B.Sc. in Physics and a Ph.D. in Computer Graphics from the University of Waterloo, then moved to Xerox, where he is now a Principal Scientist. His current interests include limitations of human vision applied to hard copy image quality. A member of ACM SIGGRAPH, IEEE-CS, the Canadian Human-Computer Communications Society, Eurographics, and IS&T, he has served on the program committees of SIGGRAPH, Graphics Interface, and RIDT.

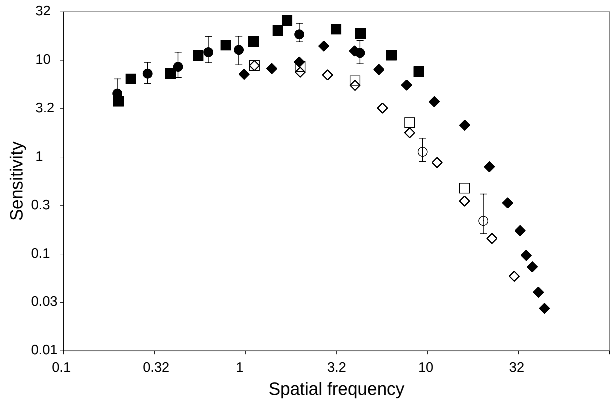

Figure 1. Sensitivity to L variation. In all figures, sensitivity is in units of inverse ΔE: a sensitivity of 10 corresponds to 0.1 threshold ΔE. Filled diamonds: Campbell and Robson 2×2˚ aperture; filled squares: 10×10˚ aperture. Open diamonds: Modelfest fixed size; open squares: fixed cycle count. Open circles: phase 1; filled circles: phase 2.*

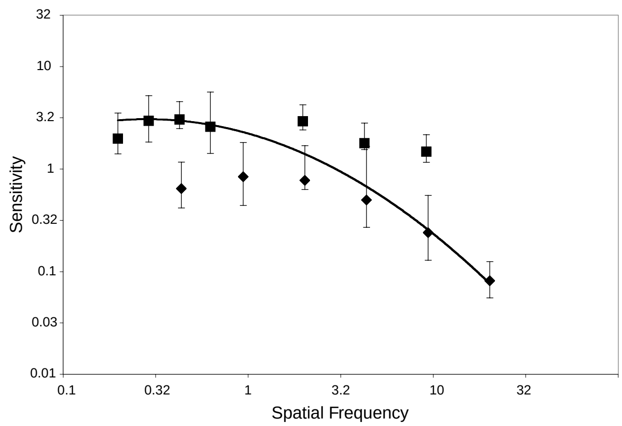

Figure 2. Sensitivity to C variation. Diamonds are phase 1 data; squares are phase 2 data; the curve is a simple quadratic fit (in log-log space) to all data combined.*

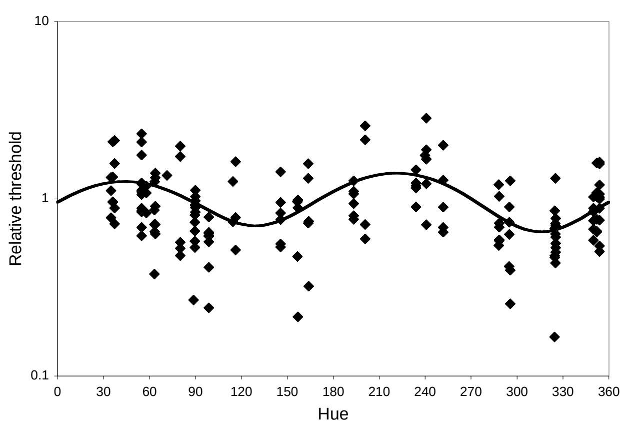

Figure 3. Relative sensitivity to hue variation as a function of hue. Each point is normalized to the average sensitivity at that spatial frequency. There appear to be two peak sensitivities in the red and cyan regions, with roughly a factor of two between max and min sensitivities.

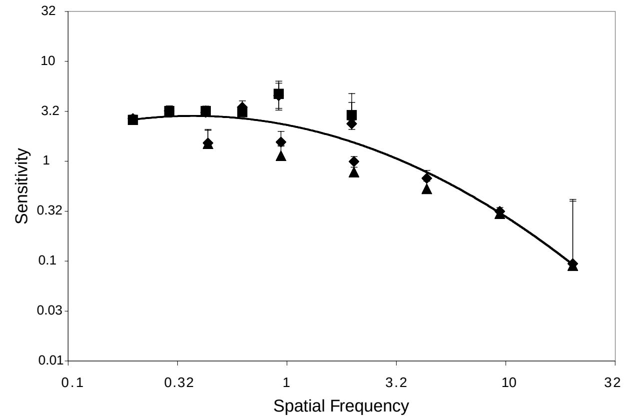

Figure 4. Sensitivity to hue variation. Diamonds: phase 1; squares: phase 2. The curve is fit to the combined data, after adjusting for the hue dependence in Figure 3.

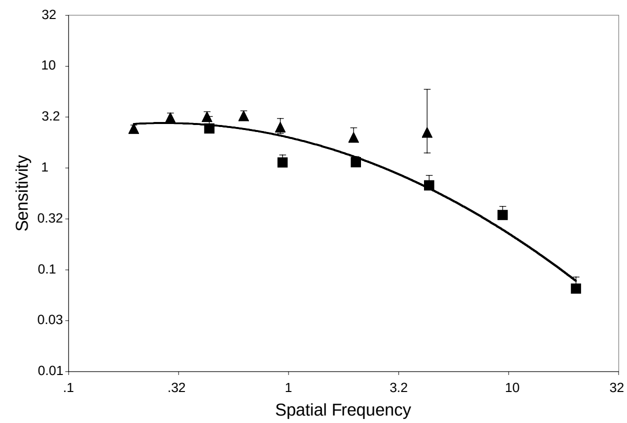

Figure 5. Sensitivity to a variation. Squares: phase 1; triangles: phase 2. The curve is a simple quadratic fit in log-log space to the combined data.*

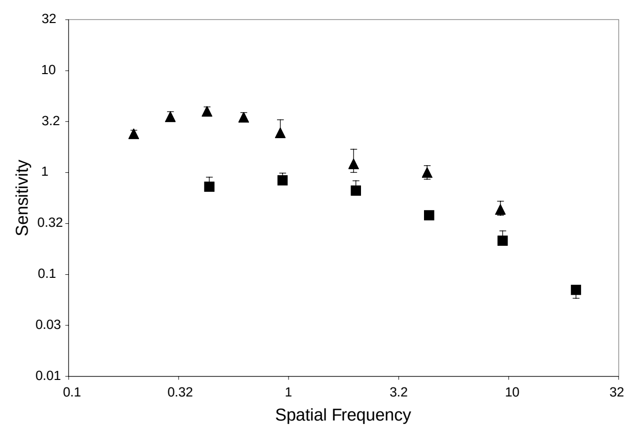

Figure 6. Sensitivity to b variation. Squares correspond to phase 1, triangles to phase 2.*