Human Chromatic Contrast Sensitivity: Exploration of Dependence on Mean Color

R. Victor Klassen and Nancy Goodman
Xerox Research and Technology

Abstract

We report preliminary results of an experiment measuring contrast sensitivity as a function of spatial frequency, location in color space and direction of variation. Observers viewed bipartite fields containing sinusoidal gratings on one side and a uniform field on the other, having the same mean color. The experiment used a forced choice paradigm to measure thresholds for a large number of observers and a large number of mean color/spatial frequency/direction of variation combinations, although no one observer saw every combination. While early data is noisy, we have found contrast sensitivity as a function of spatial frequency, on average, when any of L*, a*, b*, C* or H_{ab} is varied. We have not yet found any meaningful dependency on any independent variable other than spatial frequency, such as C*, as would be expected from such color difference metrics as CIE ΔE_{94} or ΔE_{CMC}.

Introduction

We are interested in the limits of human perception, particularly as they apply to print quality defects. Our interest lies more with the behavior of a population of observers (preferably a population similar to that population making purchase decisions), than with the specifics of any single observer. We seek the equivalent of the “CIE standard” observer, for more complex images than pairs of flat squares on constant backgrounds.

Part of the answer lies in the contrast sensitivity function (CSF), which gives the contrast at which a sinusoidal grating of a given spatial frequency becomes visible. Campbell and Robson[1] found that for spatial frequencies far enough apart (a 3:1 ratio), different spatial frequencies are detected independently. Masking and pedestal experiments[2,3,4] indicate that nearby spatial frequencies contribute to, or interfere with each others’ visibility. The bandwidth for such interference appears to be about one-two octaves, depending on spatial frequency. Moreover, typical printed output exhibits noise containing a multitude of spatial frequencies, not just a few. Thus, the contrast sensitivity function only provides part of the answer.

The vast majority of published experiments that measured the CSF have involved small numbers of observers (typically 1-3). The function was measured and used to study within-observer effects. Little data exists for larger numbers of observers. Many experiments have measured luminance-based contrast sensitivity; fewer have measured sensitivity in the chrominance channels. Those that did measure chromatic contrast sensitivity generally used a mean color along the observers' red-green or blue-yellow axis, and varied the color perpendicular to the axis or about a neutral mean color.[5,6,7] An exception is Poirson and Wandell's experiment,[8] involving two observers, in which they found that spatial frequency response could be separated from color sensitivity. In the typical experiment, great care was taken to ensure that the observer’s opponent channel axis was used, so that no excitation of the blue-yellow channel occurred when studying the red-green variation.

Our pool of observers consists of colleagues from the local technical community. Because they are volunteers, we collected a small amount of data from each of many observers, rather than a larger amount from a small number of observers. Some of the data, therefore, contains more noise than we would like. We could only characterize inter- and intra- observer variation for a subset of the conditions. However, we did measurements with over 70 observers, who collectively provided over 100,000 samples contributing to over 700 population threshold estimates. This is fewer than the 3 million color judgements made by 40 observers for the Munsell Book of Color,[9] but it comes much closer to the "standard observer" than a 2-3 observer experiment. Most estimates have error bounds that closely resemble the inter-observer variation found for those few conditions for which there were enough samples to measure inter-observer variation.

Prior work primarily falls into two broad categories: color difference and color matching experiments, and contrast sensitivity experiments. Color difference and color matching experiments led to the development of such standard color difference equations as the ΔE_{CMC} [10] and CIE ΔE_{94} [11] color difference metrics. Both the ΔE_{CMC} and CIE ΔE_{94} color difference metrics are based on the L*a*b* color space, which is intended as a good compromise between computability and visual uniformity. The alternative color difference equations attest to the fact that distances in L*a*b* space corresponding to one just noticeable difference (under the defined viewing conditions and geometry) vary by a factor of 10 over the space. However this is much better than many of the alternatives, and the data on which L*a*b* and the various derived color difference metrics are based was collected using large numbers of observers.

Moreover, the data was collected over much of color space (L*a*b* is based on a fit to the data used for the

Munsell Book of Color, obtained in experiments which involved reflective media, deficient in high chroma colors, however elsewhere the space was well populated). While these metrics have the advantage of large observer populations and good color space coverage they are specific to very simple geometries. The typical color matching experiment is performed with square constant color patches of a defined visual angle and at a fixed separation on a fixed background color of a defined size. They provide guidance for color matching, but give limited insight into where to set a machine specification for color variation.

Contrast sensitivity experiments, initially testing luminance sensitivity, were performed by Campbell and Robson[1] as a test of linearity. Savoy and McCann[12] found that for fewer than five cycles the sensitivity is driven by cycle count as well as by spatial frequency. Mullen[5] measured red-green and yellow-blue sensitivity, finding low-pass, rather than band-pass behavior. Van Meeteren[13] found that the sensitivity depends on the adaptation luminance, but for luminances in the range that interests us (typical office), the shape of the curve is pretty much constant. Contrast sensitivity has been measured in red-green and blue-yellow channels; yet both the ΔE_{CMC} and CIE ΔE_{94} color difference metrics show a dependence on chroma and ΔE_{CMC} depends also on hue. Both are based on variation in hue and chroma (weighted differently), rather than the opponent axes. However, Poirson and Wandell's pattern-color separability is based on an opponent color transformation of cone-contrast space, favoring the use of a* and b*. We are testing the hypothesis that there is a dependence on C* and/or H_{ab}, and whether the more meaningful coordinates are hue and chroma or a* and b*.

In a precursor to the current experiments, Goodman[14,15] did a number of print-based experiments. These focussed on aspects most directly applicable to images printed with process color. She published perceptibility thresholds to variations in color as caused by mass variations in each primary separation and a few mixed colors. The resultant color variation was pure lightness in black, nearly pure chroma in yellow, and about equal amounts of lightness and chroma in cyan and magenta. Since lightness dominates the visual response in the spatial frequencies studied (0.14 – 14 cycles/degree or 0.02 – 2.0 cycles/mm, the response to cyan and magenta mass variation was similar to that in black. The observed broad peak in sensitivity at 0.7-5 cycles/degree or 0.1-0.7 cycles/mm, agrees with other published results. At the peak sensitivity, the population median threshold to variation was as low as $\Delta L^*\sim 0.15$ mean-to-peak. For yellow, the peak sensitivity was at a spatial frequency of only 0.7 cycles/degree or 0.1 cycles/mm with an amplitude of $\Delta C^*\sim 2$.

The remainder of this paper discusses the conditions under which the data was gathered, including the display and observation conditions and the experiment protocol, and then describes how the data was analysed and the results obtained so far in this ongoing project. Besides spatial frequency dependence, we found relatively few trends.

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. For reasons detailed below, there were more than 700 conditions. Using a forced choice double staircase paradigm takes anywhere from 30 seconds to five minutes to obtain one estimate of threshold for one condition, depending on the level of confidence required (and the patience one may expect of the observer). This is roughly 350 hours, which was deemed impractical at best. Thus, we sought ways of reducing the amount of data needed, as well as the data collection time.

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 8.2 vertically by 10.5 degrees horizontally of a selected base color with a sinusoidal grating in either the left or right half of the field. The grating contrast varied according to a Gaussian horizontally, and fell to zero linearly in the top and bottom 10%. No visible artifacts resulted from the slope discontinuity of the amplitude. The grating width at half contrast was approximately 2.5 degrees. The grating had mean color equal to the base color and varied in one of L*, a*, b*, C* or H_{ab}. Except for edge effects, gratings were constant in the horizontal direction and varied sinusoidally in the vertical direction. This choice of direction was designed to reduce the effects of monitor non-linearity[16] and to achieve maximum cycle count for a given spatial frequency.

In order to change the grating contrast rapidly (defined as ΔE_{ab} between the minimum and maximum values in the grating) a constant image was computed and loaded into the frame buffer. The actual colors were varied by manipulating the color lookup table interposed between the frame buffer and the digital to analog converters in the video path.

The threshold contrast varies widely over the range of colors 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 grating on the left and the grating on the right so that the contrast of each could be set independently. In any given presentation, the contrast of one or the other was zero.

The monitor was calibrated according to the method described previously.[17] 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 1.75 m; the gratings had a height of 25cm resulting in an angular subtense of 8.2 degrees. According to Rovamo et al,[18] although decreased image size decreases sensitivity, for images the size shown here this will have no effect except possibly at the lowest 1 or 2 spatial frequencies used. In follow-on work, these spatial frequencies will be repeated at a closer viewing distance and corresponding larger angular subtense (and cycle count). For all spatial frequencies used, at least 4 periods were visible (3.5 at maximum amplitude). Illumination was approximately 125 cd/m^2 (normal office lighting cool white fluorescent).

Gratings were displayed at .44, .94, 2.0, 4.4, 9.4 and 20 cycles/degree (corresponding to .06, .13, .29, .62, 1.3 and 2.9 cycles/mm at normal reading distance of 40 cm). Base colors were as shown in Table 1. For each base color, as long as monitor gamut limitations permitted, we measured variation in all five directions, except for those cases where one or more dimension was meaningless or redundant with another. Specifically, for colors near the neutral axis, we did not vary hue or chroma; for colors on or near the a* (b*) axis we varied only one of hue and b* (a* respectively), and only one of chroma and a* (b*, respectively). L* was varied at all spatial frequencies, color was varied at all but the lowest spatial frequency for all base colors, and also at the lowest spatial frequency for about 30%of base colors.

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

Observer Demographics

Observers were volunteers taken from the pool of technical employees at our site. Over seventy observers participated in the experiment, with ages ranging from 19 to 61 (plus two of the experimenter's children, who together contributed less than 0.4% of the observations). Most observers completed one session (20-30 minutes) covering roughly one to two dozen conditions. A few observers did multiple sessions, such that no observer contributed more than 12% of the data. The gender mix was biased toward male with males contributing 60% of the total observations, but less biased than the workforce. Figure 1 shows the distribution of observations according to the age of the observation population (i.e. an observer is counted in proportion to the number of observations contributed).

Only color normal observers participated, as verified with Ishihara's tests for color deficiency.[19] It should be noted that especially in the male population there is considerable variation in the red-green channel even within the 90% of color-normal observers. It would not be surprising if this created some variability in the observations.

Table 1. Base colors used are randomly distributed through the available gamut

L*	a*	b*	c*	H_{ab}
37.00	0.00	0.00	0.00	0.00
73.94	0.59	0.07	0.59	6.77
85.00	0.59	0.07	0.59	6.77
48.25	60.70	42.27	73.97	34.85
67.07	35.27	26.06	43.85	36.46
76.65	19.00	15.00	24.21	38.29
60.00	25.00	40.00	47.17	57.99
70.00	35.00	65.00	73.82	61.70
70.00	5.00	10.00	11.18	63.44
85.30	3.40	6.90	7.69	63.77
25.00	8.00	24.00	25.30	71.57
30.00	8.00	24.00	25.30	71.57
15.58	0.72	3.25	3.33	77.51
56.00	10.00	60.00	60.83	80.54
80.00	0.00	55.00	55.00	90.00
82.00	-2.83	18.00	18.22	98.94
37.00	-15.00	32.00	35.34	115.11
80.00	-30.00	60.00	67.08	116.57
40.00	-40.00	40.00	56.57	135.00
72.00	-30.77	13.20	33.48	156.78
81.70	-15.99	6.70	17.34	157.27
50.00	-35.00	10.00	36.40	164.05
33.00	-15.00	-6.00	16.16	201.80
82.09	-11.00	-15.56	19.06	234.74
60.00	-15.00	-26.00	30.02	240.02
30.00	-6.00	-20.00	20.88	253.30
40.00	12.00	-36.00	37.95	288.43
53.85	12.69	-27.55	30.33	294.73
70.08	8.48	-17.26	19.23	296.17
30.00	45.00	-32.00	55.22	324.58
47.00	13.00	-9.00	15.81	325.30
79.70	22.17	-3.91	22.51	350.00
68.50	38.82	-5.53	39.21	351.89
49.44	73.67	-7.68	74.07	354.05

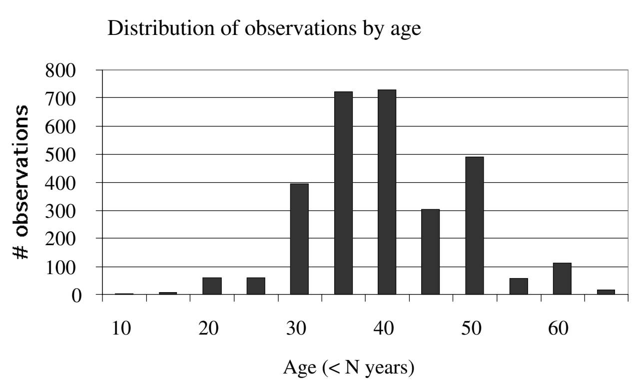

Figure 1: Distribution of observer ages, weighted by the number of observations performed.

Experiment Protocol

The experiment protocol was largely dictated by the need to obtain data on large numbers of conditions without overly burdening the observer population. The presentation was in the form of two alternative forced choice, without feedback. An audible beep initiated presentation of a stimulus. The observer pressed a key on the right or left side of the keyboard to indicate the side on which the grating appeared. If an observer realized (s)he had pressed the wrong key, (s)he could press 'X', the log file would be annotated to toggle the previous presentation, and the next stimulus would then be re-presented. Allowing the observer to make correction substantially improved the accuracy of threshold estimates (after half an hour one tends to make the occasional error); it also reduced observer frustration.

The grating appeared randomly on the left or the right, with a slight bias in favor of the side on which it had appeared least frequently in the last 4 trials (to compensate for any unexpected bias created by monitor or viewing field inhomogeneity). The contrast was taken (again randomly) from one of two sequences, one ascending and one descending. When an observer had made enough incorrect guesses during a descending sequence or enough correct guesses during the last part of an ascending sequence the sequence was cut short, to avoid wasting observer time. Occasionally (less than 1 in 10) a maximum contrast grating would be displayed as positive reinforcement for the observer.

The entire range of contrasts was divided into 50^{ths}; the even half of these were used for one sequence and the odd half for the other. At small thresholds, quantization resulted in many of these contrasts being the same, but at typical thresholds there were at least 20-30 different contrasts displayed, sometimes as many as 50. When the threshold was very small (<0.5 ΔE_{ab}), there were as few as 5 contrasts displayed. In this case, we had a high level of confidence of the probability of a correct response for each of a small number of contrasts. In the typical case, no observer viewed the same contrast enough times to reliably predict the 75% correct level.

Data Analysis

At each response, the control program recorded whether the observer had answered correctly as well as the L*a*b* of the two extrema colors displayed. These might be different from the L*a*b* min and max values as requested since the requested values were converted to RGB and then quantized to 8 bits. While a grating may have been intended to vary only in a*, the nearest RGB values to the desired max and min may have differed in L* as well. An initial post-process had already toggled those answers where the observer indicated an error.

Subsequent measurement of 45 conditions varying in either L*, a* or b* about 15 base colors showed that this cross-talk was less than 20% in 85% of the cases. However, there were three cases for which b* variation was requested and obtained along with L* variation of ~30% as much. This may have tainted some measured thresholds to b* variation at spatial frequencies for which the threshold to L* is a factor of 3 or more below that for b* variation, but is not expected to have affected the overall results.

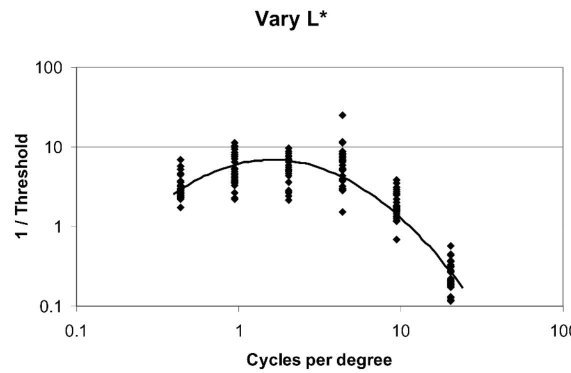

Figure 2: Sensitivity to L variation. The scale is different for this plot than for variations along any of the chrominance directions.*

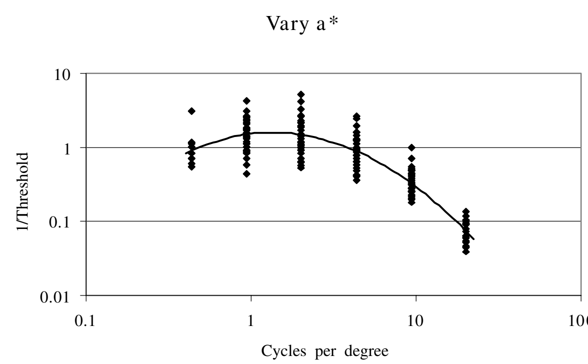

Figure 3: Sensitivity to a variation.*

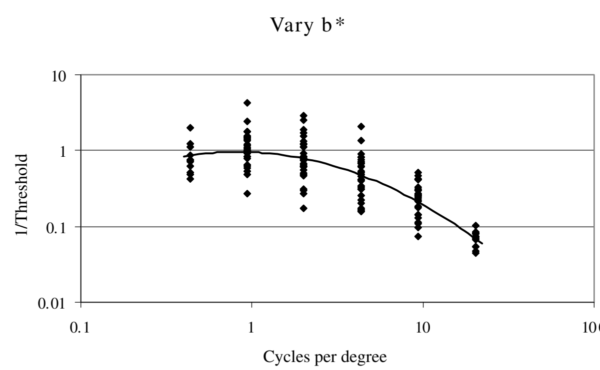

Figure 4: Sensitivity to b variation.*

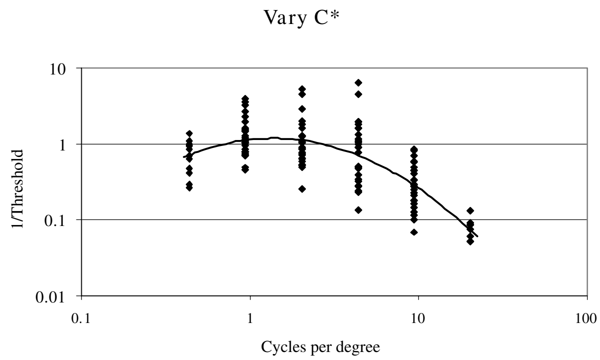

Figure 5: Sensitivity to C variation.*

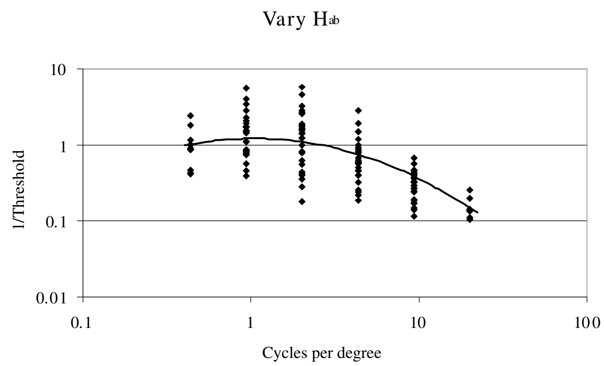

Figure 6: Sensitivity to H_{ab} variation.

The data was analysed within session and pooled over all sessions. Most of what is reported here is based on the pooled output. Pooling was done by sorting all of the responses by condition before any further processing. The ensemble threshold response was found, along with lower and upper bounds (95^{th} percentile confidence interval) for each threshold using the methods described in the Appendix. In brief, the methods used find three points. At the threshold estimate the group of observers answered correctly 75% of the time. Between the other two points, the observers could be said to be drawing at random from a distribution containing 75% correct answers, and 25% incorrect. At the upper and lower bound points, one can say with 95% confidence that the answers are no longer drawn from such a distribution.

Note that by its nature, this method of analysis finds the population ensemble median threshold. Since only a few of the over 70 observers viewed each specific condition (combination of base color, spatial frequency and nature of color variation), the spread of responses to the various conditions gives an indication of the population statistics.

Results

Representative data are shown in Figures 2 through 6. These show the behavior with respect to L*, a*, b*, C*, and H_{ab} variations, respectively. (Note: we show lightness sensitivity as $1/\Delta L$ on a log scale, rather than $-\log(\Delta Y/Y)$; this makes comparison with chrominance easy, but absolute sensitivities are not comparable with contrast sensitivity data). Sensitivity at the highest spatial frequency is to be treated with caution as there were many colors for which we could not increase the contrast high enough to make it visible at that spatial frequency, and therefore they were not included in the ensemble for that spatial frequency. Therefore, the sensitivity at that spatial frequency is somewhat overstated. In addition, for thresholds at peak sensitivity, the precision is insufficient to obtain better than 100% error in the estimate. That is, there is a reliable upper bound, and the threshold is somewhere between that upper bound and zero contrast. For these, we assume a threshold of one half the upper bound. The limit of our precision is approximately 0.1 to 0.3 ΔE_{ab}, depending on the location in color space. At the lowest spatial frequency there were fewer than five cycles (there were 4, with 3.5 at full contrast); hence the sensitivity may be slightly depressed due to low cycle count.

Data Variance

From the data collected thus far, 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.

With the error defined 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 ±65%. These error bounds about typical for contrast sensitivity data. The Modelfest[20] data, available on the web[*], show a range of variation of about a factor of 3 between lowest-threshold and highest-threshold observer in a group of nine observers. This corresponds to approximately ±70% about the geometric mean. We therefore conclude that the large errors are more related to inter-observer variability than to our experimental technique.

A few (8) conditions were seen more than 10 times, including, in some cases more than once by the same observer. The inter- and intra-observer variability of the threshold estimate in those cases ranged from ±(10 - 15)% to a factor of 4 or more, primarily less than a factor of 3. Again this indicates similar inter-observer variability as seen in the ModelFest data, although the intra-observer variability is somewhat higher (likely due to the relatively small amount of data taken per condition).

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

Non-Trends

There are a number of expected trends that our data cannot confirm. Either confounding factors masked them, they were hidden by noise, the range of our parameters was insufficient or the expectation was unjustified.

First, we did not measure low enough spatial frequencies to test Lee Guth's hypothesis[21] regarding low- or band-pass response in the chroma channels. (Guth hypothesized that the low-pass appearance results from the fact that chroma data is normally collected with zero response at the midpoint of the sinusoid e.g. red-green variation about yellow, whereas luminance data cannot be collected using sinusoidal variation about zero response.) Mullen's data5 (also based on variance about zero response) indicates low-pass, but the flat portion ranges from 0 to about 0.5 cycles/degree, which is our lowest spatial frequency. We plan to extend the experiment to include lower spatial frequency data and to confirm the responses to some spatial frequencies with a greater angular subtense and cycle count.

Second, we would expect that at the highest spatial frequency, varying b* would have a higher threshold than varying a*, based on chromatic aberration arguments. However, the measured difference between the two is not statistically significant. Notably, at the high frequency end we had more conditions in which we could not vary b* sufficiently to reach threshold (due to gamut limitations) than when varying a*. This suggests that the threshold for varying b* is indeed higher, although we were unable to measure it.

Finally, we expected a dependence on C* to appear somewhere in our data. The CIE ΔE_{94} color 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, with a slope of 0.045. This results in a factor of four difference from C* = 0 to C* = 66. ΔE_{CMC} indicates a dependence of C* sensitivity to base C* of similar magnitude, although the detailed dependence is different. We have several colors at each of these extreme values, yet at only one spatial frequency does a line of best fit have an r^2 greater than 0.1: it has an r^2 of 0.217, but a slope of -0.12, which indicates greater sensitivity at higher chroma. In addition, ΔE_{CMC} indicates a dependence of H_{ab} sensitivity to base H_{ab} of a factor of 2, and this was not found either.

We speculate that the dependence on base C* and/or H_{ab} is not observed here due to differences in viewing conditions. These images are spatially varying, as opposed to the pairs of uniform squares used in the experiments to determine ΔE_{94} and ΔE_{CMC}. Perhaps more importantly, the surround was of the mean color 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 thus enhancing the observers' ability to detect color differences.

Trends

As expected, peak sensitivity to variation in L* is at approximately 2.0 cycles/degree, in agreement with our previously published results[15] based on printed mid-tone black, cyan and magenta samples and also in agreement with other reports in the literature. In the current experiment, there were 34 different base colors, and these spanned the entire available color gamut, indicating that base color does not affect sensitivity to lightness variation. Also in agreement with earlier work, sensitivity to variation in L* is significantly (a factor of 6 to 12) greater than sensitivity to variation to any of a*, b*, C* and H_{ab}, at 2.0 cycles/degree.

Surprisingly, sensitivity to b* and to a* variations differ significantly. Sensitivity to a* variation appears to peak at ~1.5 cycles/degree Sensitivity to b* variation may have a peak at ~1.0 cycles/degree, but the current data can not distinguish between this and low-pass behavior. In contrast, Mullen[5] found no peak in red-green or yellow-blue sensitivities, nor did she find any significant difference between them. While there is no significant difference between the measured sensitivities at the highest and lowest spatial frequencies, at three of the intermediate spatial frequencies, 2.0., 4.4, and 9.4 cycles/degree, sensitivity to a* variation is significantly (at the 99% confidence level) larger than to b*. We have no explanation at this time for this difference, but are pursuing it further.

Sensitivity to C* and H_{ab} are similar to a* and b*, respectively. In general there is a trend for chromatic sensitivity to be highest for a*, then H_{ab} and C*, then b*. Sensitivity to L* has by far the greatest dependence on spatial frequency, then a* and C*, then H_{ab} and b*.

Sensitivity to variations in L* is at least six times as high as sensitivity to variation in any other direction for all but the highest spatial frequency we measured. We have not yet analysed this for dependence on C*. As discussed earlier, measured thresholds at the highest spatial frequency were limited by our inability to display higher amplitudes.

All of these results are preliminary. Some apparent trends may disappear as we collect more data and further analyse the data we have, possibly removing some outliers. It is possible that other trends will appear as well, such as a confirmation of whether chromatic sensitivity is characterized by low pass or band pass.

Conclusions

We have reported on *preliminary* results of an experiment to measure ensemble spatial frequency response for a large group of observers for an arbitrary base color and with color variation along any of L*, a*, b*, C*, and H_{ab}. Several anticipated trends have not been observed, but these may be masked by the large amount of noise inherent in such a measurement process. These include a dependence on C* which corresponds to the reduction in ability to detect differences in chroma and hue at larger values of C* as captured in CIE ΔE_{94} or ΔE_{CMC}, and a difference between performance in detecting variation in a* vs. in b* at the highest spatial frequency. We speculate that the lack of dependence on C* is due to differences in viewing conditions, as explained above.

We did find that at all but the highest spatial frequency, L* differences are detected at least four times as easily as differences in any of the chrominance directions. While the band pass behavior of the luminance detection mechanism was verified, two different forms of behavior were observed for chrominance channels: band-pass and 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.

We intend to continue the experiment in order to increase the likelihood that the expected trends will show up if they are in fact real, and to verify or eliminate the unexpected trends we found in the preliminary data.

References

1. Campbell, F.W. and J.G. Robson, "Application of Fourier Analysis to the Visibility of Gratings", Journal of Physiology 197 pp.551-566, 1968
2. Stromeyer, C. and B. Julesz, “Spatial-frequency masking in vision: critical bands and spread of masking” , Journal of the Optical Society of America, 62 (10),pp. 1221-1232, 1972
3. Legge, G.E. and J.M. Foley, “Contrast Masking in human vision” , Journal of the Optical Society of America, 70 (12),pp. 1458-1471, 1980
4. Foley, J.M., “Human luminance pattern vision mechanisms: Masking experiments require a new model” , Journal of the Optical Society of America A, 11 pp. 1710-1719, 1994
5. Mullen, K.T., “The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings” , J. Physiology, 359 pp. 381-400, 1985
6. Switkes, E. and K. DeValois, “Luminance and chromaticity interactions in spatial vision” , Colour Vision, pp. 465-470, 1983
7. King-Smith, P.E. and D. Carden, “Luminance and opponent colour contributions to visual detection and adaptation and to spatial and temporal integration” , Journal of the Optical Society of America, 66 pp. 709-717, 1976
8. Poirson, A.P. and B. Wandell, “Appearance of colored patterns: Pattern-color separability” , J. Optical Soc. of America, 10 (12),pp. 2458-2470, 12/93
9. Newhall, S.M., D. Nickerson and D.B. Judd, “Final report of the O.S.A Subcommittee on the spacing of the Munsell colors” , Journal of the Optical Society of America, 33 (7),pp. 385-418, July, 1943
10. Clarke, F.J.J., 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
11. Berns, R.S., “The Mathematical development of CIE TC 1-29 proposed color difference equation: CIELCH” , Proceedings of AIC colour 93, B (c19-1-4), 1993
12. Savoy, R.L 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
13. van Meeteren, A., and J.J. Vos, "Resolution and contrast sensitivity at low luminances", Vision Research 12 pp. 825-833, 1972
14. Goodman, N. “Human sensitivity to within page color uniformity”, Proc. SPIE/IS&T Electronic Imaging ’97 Conference: Human Vision and Electronic Imaging II, 139-146, 1997.
15. Goodman, N, “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.
16. Brainard, D, "Calibration of a computer-controlled color monitor", Color Research and Applications 14 pp. 23-34, 1989
17. Klassen, R.V, K. Braun, R. Buckley and K. Janamanchi, "Visibility of thin lines on coloured backgrounds", Proc. 6th Color Imagin1g Conference, pp. 27-32, 1998
18. Rovamo, J., Luntinen, O. and Näsänen, R., “Modelling the Dependence of Contrast Sensitivity on Grating Area and Spatial Frequency”, Vision Research, 33, pp. 2773-2788, 1993
19. Ishihara, S., “The series of plates designed as a test for colour-blindness”, Kanehara Shuppan Co., Tokyo, Japan, 1993
20. Carney, T., C. Tyler, A.B. Watson, W. Makous, B. Beutter, C-C. Chen, A. Norcia, S. Klein, "Modelfest — first year results and plans for year two" to appear in Proceedings of IS&T/SPIE Human Vision and Electronic Imaging V, San José 3959, 2000
21. Guth, L. "The chromatic contrast sensitivity myth", Proceedings of the 5th Color Imaging Conference, pp. 23-26, 1997

Appendix: Estimating Ensemble Thresholds from Combined Forced-Choice Response Data

A data stream was segmented by condition and then each condition was analysed. Within condition, the trials were sorted by the amount of variation in the independent variable. Beginning at the largest variation and proceeding to the smallest, the number correct in a sliding window of n>10 responses was counted, and the first occurrence of k•3n/4 correct was taken as threshold. If a later (lower contrast) occurrence of k>3n/4 correct occurred, it was most likely a random fluctuation and not an indication of a lower threshold.

The use of a sliding window of n observations, not all of which are necessarily at the same contrast (and most of which are at different contrasts for single-session analysis) can be justified on the basis that we expect the threshold as estimated to be within the range of contrasts represented by the window centered at the threshold. There are two limiting cases to the common psychometric function typically fit in such instances (a method we avoid for reasons to be explained later). At one limit, the slope at threshold is infinitely steep. In this case, any window centered at threshold is expected to contain 50% correct answers in one half the window and 100% correct in the other half, for a net result of 75% correct. In the other limit, the function is very shallow, and for a significant distance on either side of threshold is approximately linear. In this case, the average response on one side of the window will exceed 75% by the amount by which the average response on the other side of the window is less than 75%. Finally, so long as the psychometric function is symmetric, the second argument continues to hold. We do not know that the function is symmetric (although typically chosen functions used in data fitting are), but have not seen evidence to the contrary.

For any given window size, it is straightforward to calculate the maximum number that may be correct before one must discard the hypothesis (with any particular confidence level — we chose 95%) that the observer is at the 75% level. At the 75% level we expect an observer to guess correctly with the same frequency that one would draw black balls (with replacement) from an urn containing 75% black balls. Similarly, it is straightforward to calculate the maximum that may be wrong before one must discard the hypothesis. While scanning from high to low contrast, we find the upper bound in the same pass as when looking for the threshold. A second scan from low to high finds the lower bound as the first time the 75% correct hypothesis cannot be discarded. This gives 95% confidence levels on all threshold estimates.

We have shown that including values from adjacent contrasts is unlikely to affect the estimate of threshold, but we have yet to show the same about bounds. Suppose that the upper bound is found at some (relatively) small window size. As the window size increases, but remains centered at the same location, the window will include a larger number of correct answers (from the higher contrast half of the window), but from the other half of the window, it will include increasingly many incorrect answers. In the usual situation, one would expect the bounds to grow further apart as the window size increases. Working against this effect is the fact that for very large windows, one need be only slightly higher than 75% before one may reject the hypothesis. With this in mind we test all window sizes from 11 samples to 75 and take the answer with the tightest bounds.

The reason we chose this rather unconventional analysis scheme is that it is robust to relatively small sample size. With as few samples as 16, one may, if lucky, locate the threshold. Secondly, it allows us to explicitly test for conditions such as one in which the upper bound does not exist (the threshold was never exceeded by enough to confidently determine an upper bound). And, unlike fitting to a psychometric function, we may compute an exact value of the upper and lower bound, based only on two assumptions: that of a binomial distribution, and the validity of windowing data. With a psychometric function, one obtains the slope of the function, and a correlation coefficient giving goodness of fit. These two pieces of information can aid one in estimating the bounds but do not provide a straightforward way to conclude that the threshold is most likely at contrast c, and we know with 95% confidence that it is not at a contrast greater than u or less than l. In fact, few other methods will provide asymmetric bounds.