Colorimetric Tolerances of Various Digital Image Displays

Jason E. Gibson, Mark D. Fairchild, and Steven L. Wright [*]
Munsell Color Science Laboratory Rochester, New York
[*] ***IBM T.J. Watson Research Center Yorktown Heights, New York***

Abstract

While color difference formulas such as CIE94 work reasonably well for simple uniform color patches in direct edge contact, their application to pictorial images has only recently been investigated. In this study, perceptible color difference tolerances were developed for pictorial images displayed on a CRT display, two LCD displays, and on continuous tone hardcopy output.

Three transforms were used to represent common global image processing steps. Reductions in contrast were simulated using a sigmoidal compression in L^*. Reductions in system gain were modeled using a multiplicative reduction in C^*_{ab}. Overall color casts were modeled using an additive offset in h_{ab}.

For sigmoidal L^* compression, the prints had lower sensitivity than the other three displays. For multiplicative reduction in C^*_{ab}, the prints had higher sensitivity than the others. For rotations in h_{ab}, the prints had the highest sensitivity, followed by one of the LCDs, then the remaining displays, which had similar performance to each other.

Thresholds are also expressed colorimetricly using pixel by pixel comparisons with and without S-CIELAB pre-filtering using three color difference formulas. In general, the overall thresholds ranged from 1 to 3 ΔE for each of the three perceptual attributes. In this experiment, S-CIELAB pre-filtering did not have a significant effect.

Introduction

Much of the early development of color difference formulas was dominated by "traditional" color industries such as textiles, paint and plastics, which were primarily interested in large uniform areas of color in direct edge contact. However, with the advent of computer imaging systems and displays capable of rendering high-quality full-color pictorial images, the need for formulas and methods to deal with this added complexity are now being investigated.

This research project continues the work of Stokes[1] and Uroz,[2] by further examining color tolerances in pictorial images with the aid an IBM prototype 200ppi TFTLCD.[3]

To examine these tolerances, a series of four psychophysical experiments were conducted. The first experiment was performed using a Sony GDM-F500 CRT** to allow comparisons with the previous work of Stokes.[1] The experiment was then repeated on a SGI 1600SW LCD**, which represented currently available flat-panel display technology. The third experiment used an IBM 200ppi TFTLCD Roentgen prototype display3 to see if its improved resolution had any impact on its color tolerance thresholds. The fourth experiment was conducted using hard-copy output from a Fujix Pictrography 3000 printer** to provided a high-end anchor for comparing results.

Experimental Methods

Display Characterization

Procedures for the colorimetric characterization of a CRT are well established.[4] With the advent of affordable, high quality LCD flat-panel displays, procedures for characterizing such devices are now being developed as well.[5,6]

Characterization of the Fujix Pictrography was carried out using an interpolated 3D LUT approach. To create the RGB to L*a*b* 3D LUT, a 10x10x10 sampling of RGB space was printed and measured. To avoid interpolating from an irregular source space back to RGB, a 60x60x60 regularly spaced inverse 3D LUT was created by uniformly sampling the bounding box enclosing the measured output gamut. Performance of this method was deemed adequate.

Representative characterization performance of the four displays used in this experiment is shown in Table 1 below. The relatively large errors for the prototype IBM display are not yet fully explained, but are thought to be due in part to a lack of additivity and or it's poor black state. Being a prototype unit, the IBM display contained several defects that would not be present in a finished product. Every effort was made to minimize or reduce the impact of these defects, but in some cases they were unavoidable.

A brief comparison of each display's specifications is presented in Table 2. Further details of each display, as well as data on other performance factors for the three emissive displays can be found on the MCSL web site at www.cis.rit.edu/mcsl/reports.shtml.[7]

Table 1. ΔE^*_{94} Colorimetric Errors for Independent Data

Display	Quartile 25%	50%	75%	Average	Maximum
Sony CRT	0.23	0.33	0.41	0.36	1.10
SGI LCD	0.83	1.05	1.23	1.01	1.81
IBM LCD	2.49	3.68	4.87	3.73	7.63
Fujix Print	3.18	3.94	4.74	3.97	6.54

Table 2. Summary of Display Specifications

	Sony	SGI	IBM	Fujix
Diagonal	19.8”	17.3”	16.3”	N/A
Resolution	72 ppi	110 ppi	200 ppi	400 dpi
Total Pixels H x V	1280 x 1024	1600 x 1024	2560 x 2048	3800 x 2759
Bits/channel	8	8	6	8
cd/m^2	56	161	153	836
Contrast	427:1	276:1	205:1	30:1

Image Selection and Transforms

Both Stokes[1] and Uroz[2] found little evidence of scene content dependency for systematic uni-dimensional changes in images. Thus, only three images were selected for use in this project to keep the number of stimuli manageable.

The first image, denoted “Girl” is a commonly used test image from a *Kodak Photo CD*™ sample disk. The girl’s bright red and green sweater, multi-colored ribbons and dark black hair provide the observer with varying textures, bright colors, and dark shadows to examine.

The second image, denoted as “Flower”, is a close-up of five purple-blue flowers with yellow-orange pistils on a grassy background. The image includes both deep shadows as well as a fine vein structure in the petals.

The third image, denoted “Currency”, is a close-up of currency of many different countries. It is different from the others in two important features. First, the image has no high chroma features. Secondly, the engravings on the currency dominate the image with high-frequency details. Though this image turned out to be difficult for the observers to judge, it was thought to be the best choice for observing the effects of the varying display resolutions.

Once the images were selected, a set of transforms to be applied was developed. In his thesis, Stokes[1] describes twelve global image color transforms using combinations of four mathematical functions (multiplication, addition, exponents and sigmoids) along with the three CIELAB dimensions of lightness, hue and chroma. Several of these combinations, such as additive offsets in chroma, can be eliminated as having no physical correlate.

Many spatially localized transforms such as low-pass filtering to change sharpness or the addition of random deviations to simulate system noise, as was done by Urzo[2,] could have been applied. However, these transforms would likely be very image dependent[2] and require a significantly larger set of images to test.

To keep the experiment manageable, three transforms were selected as being representative of the most common global color changes during image processing: contrast, gain and overall color cast. Change in image contrast was modeled using a sigmoidal transform in CIELAB lightness (L^*). The sigmoid was simulated using the two part exponential function shown in Eq. 1. Reduction in system gain was simulated using multiplicative transforms in CIELAB chroma (C^*_{ab}) (Eq. 2). Overall color cast was modeled using additive offsets in CIELAB hue (h_{ab}) (Eq. 3).

$$L^*{}' = \begin{cases} \dfrac{\left(2L^*\right)^k}{2} & L^* < 0.5 \\[2ex] \dfrac{\left[\left(2L^* - 1.0\right)^{1/k}\right] + 1.0}{2} & L^* \geq 0.5 \end{cases} \qquad (1)$$

$$C^*{}' = kC^* \qquad (2)$$

$$h_{ab}{}' = k + h_{ab} \qquad (3)$$

To test the symmetry of the thresholds, Stokes applied the chroma and lightness transforms in both the compressive (k<1) and expansive (k>1) directions. For hue, both positive (k>0) and negative (k<0) rotations were applied. His results indicated that the three transforms used in this experiment had symmetrical tolerances. Thus, to avoid gamut boundary issues when manipulating the images, only compressive transforms were used in the current experiment.

With three images, eight levels each of hue rotation and chroma reduction, and nine levels of lightness compression, a total of 75 sets of images per display were used.

Psychophysical Methods- Monitors

For the three monitor experiments, a two alternative forced choice (2AFC) with anchor method was used. During a session, observers were shown a total of 75 image triplets. For each triplet, the known original (anchor) was presented first. Then, using one of three keys, the observer was able to select between the two alternative images or bring up the known original. A brief flash of gray between images was used to preserve adaptation. When the observer identified the different image, a fourth key registered their response and brought up the next triplet. Between triplets, a gray screen was presented for approximately one second to maintain a more uniform state of adaptation.

Three practice trials using obvious changes were presented at the start of each session. This allowed the observer to become aquatinted with the controls, ensured that they understood the directions, and gave them time to adapt to the environment. Visual and verbal feedback was given for each image set during the trials. All observers were verbally reminded to select the different image before beginning the actual experiment.

During the first three experiments, each of the three monitors was masked off to be 11.5" x 7.5" to eliminate differences in aspect ratio. For the Sony CRT display, observers were allowed unrestrained viewing, but were encouraged to maintain a distance of ~20". For the SGI and IBM LCD displays, observers viewed the display through a 0.5"x3.0" slit, 20" from the display to minimize problems associated with viewing angle dependencies.

For the three monitor experiments, the average observer completed the task in 27 minutes with a range of 14–62 minutes. More than thirty observers participated in each of the four experiments. Most observers commented that the task was challenging, but not overly fatiguing.

Psychophysical Methods- Prints

Threshold experiments using prints typically rely on the observer making a 'same/different' judgment versus some standard sample. While experimentally convenient, this method relies on the subjective judgment of the observer. When samples near threshold are presented, an observer may simply reply 'different' in an effort to appear more sensitive.

To provide an objective measure of performance, observers are asked to rank order the Fujix print samples in terms of difference from an unaltered original. This ordering was then compared to the order based on the amount of transformation applied. It is hypothesized that pairs of samples whose mutual difference in the amount of transformation applied is below threshold will be confused more often than pairings whose mutual difference is above threshold. The frequency of confusion for all possible pairings was computed and analyzed using probit analysis.

The average observer spent 47 minutes on this experiment, with a range of 27–100 minutes. Many of the observers commented that this was significantly more difficult than the three monitor experiments.

Parametric Threshold Results

The thresholds presented in this section are referred to as parametric thresholds as they represent the value of the free parameter, k, in Eqs. 1–3 above. Though these comparisons are dependent on the transforms used, they provided a direct comparison of the relative sensitivity of each display tested. The overall average of the three images will be shown for each transform on each display.

The data presented in Tables 3–6 are to be interpreted as follows. The boldface numbers represent the mean threshold for the display in that row for the transform in that column. The plain type numbers to the right of the threshold are the upper and lower 95% fiducial limits. An asterisk indicates a lack of fit at α=0.1. For example, the estimated threshold for chroma on the SGI display was 0.91 with a range of [0.93,0.91].

The upper and lower fiducials are often of different sizes due to the asymmetrical distribution of the pre-defined stimuli levels around the threshold. In most cases it was easier to generate stimuli above threshold than below it.

Table 3. Parameter Thresholds

	Lightness		Chroma		Hue	
Stokes CRT	**0.88**	0.89 0.87	**0.91**	0.92 0.90	**5.9**	6.5 5.2
Sony CRT	**0.87**	0.88 0.85	**0.90**	0.92 0.89	**6.2**	7.1 5.2
SGI LCD	**0.86**	0.88 0.85	**0.91**	0.93 0.90	**3.9***	5.4 1.4
IBM LCD	**0.87**	0.89 0.86	**0.90**	0.93 0.88	**5.9***	8.0 2.8
Fujix Print	**0.78**	0.81 0.73	**0.99***	1.02 0.97	**-0.5***	0.8 -2.7

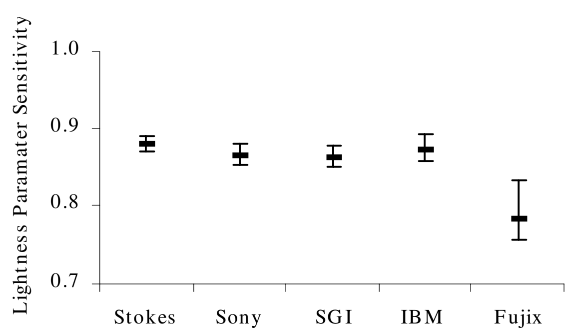

Figure 1. Lightness Parameter Sensitivity

In the lightness dimension, the currency image on the Sony CRT had a slightly lower sensitivity than the other two images. On the SGI LCD, the currency image was found to have a slightly higher sensitivity than the other two images. On the Fujix Pictrography, observers were completely unable to see differences in this image and ranked it randomly. See Fig. 1.

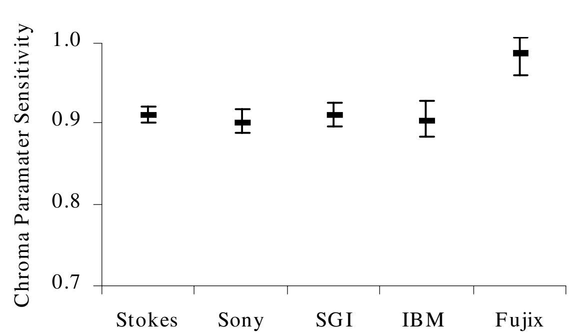

Figure 2. Chroma Parameter Sensitivity

In the chroma dimension, the Girl image on the Sony CRT had slightly higher sensitivity than the other images. On the SGI LCD, the Flower image had slightly lower sensitivity. The lack of fit for the Fujix prints ($p<0.0003$) needs to be considered in interpreting the results. See Fig. 2.

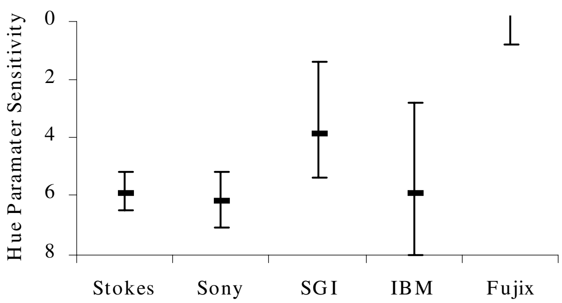

Figure 3. Hue Parameter Sensitivity

On the Sony display, the Girl image was slightly more sensitive in hue than the other two images. On the Pictrography, the Girl image was estimated to have negative sensitivities which pulled the overall results negative as well. Given the lack of fit for the Fujix ($p<0.0001$), these results are suspect. The lack of fit for the SGI ($p=0.0779$) and IBM ($p=0.018$) LCD's must be considered when interpreting these results. Note that the Y-axis in Fig. 3 has been reversed such that the null transform (0) is on top such that sensitivity increases with height along the Y-axis as in Figs 1 &2.

Pixel by Pixel Colorimetric Threshold Results

In this section, the estimated thresholds are expressed in terms of the average pixel-by-pixel color difference between the original image and one transformed by the appropriate threshold amount. For the Fujix prints, the pixel-wise subtraction was performed on the digital source files. Both the printing process and monitor models were assumed to introduce only uniform bias errors. Because of this, only relative comparisons can be made since the actual colorimetry presented to the observer is unknown. Color difference are expressed in terms of ΔE^*_{ab} and ΔE^*_{94} formulas. A candidate formula for ΔE^*_{2000} is also used.[8]

For lightness compression, the three monitors are in close agreement with each other and with the findings of Stokes[1]. The Fujix prints had a slightly higher threshold. The ΔE^*_{ab} and ΔE^*_{94} thresholds are identical since $k_L=1.0$, and the images varied only in L^*. See Fig. 4.

The chroma thresholds for the three monitors are in overall agreement with one another. The choice of color difference formula has a larger impact in chroma than in lightness. The Fujix prints had a very tight tolerance to changes in chroma. See Fig. 5.

Table 4. Delta E's for Lightness Thresholds

	ΔE^*_{ab}			ΔE^*_{94}			ΔE^*_{2000}		
Stokes CRT	**1.94**	2.44	1.14	**1.94**	2.44	1.14	N/A		
Sony CRT	**1.76**	2.05	1.62	**1.76**	2.05	1.62	**1.38**	1.60	1.26
SGI LCD	**1.89**	2.04	1.61	**1.89**	2.04	1.61	**1.48**	1.59	1.26
IBM LCD	**1.75**	1.89	1.46	**1.75**	1.89	1.46	**1.37**	1.48	1.14
Fujix Print	**3.08**	3.55	2.32	**3.08**	3.55	2.32	**2.42**	2.80	1.82

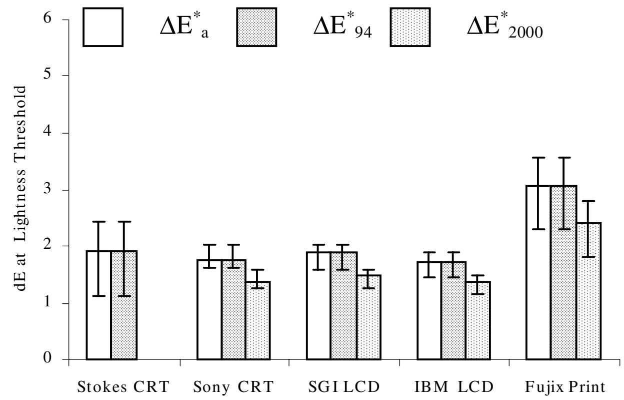

Figure 4. ΔE for Lightness Thresholds

Table 5. Delta E's for Chroma Thresholds

	ΔE^*_{ab}			ΔE^*_{94}			ΔE^*_{2000}		
Stokes CRT	**2.19**	3.41	1.23	**0.94**	1.26	0.48	N/A		
Sony CRT	**1.60**	1.76	1.28	**0.85**	0.93	0.68	**1.00**	1.10	0.79
SGI LCD	**1.44**	1.59	1.12	**0.76**	0.85	0.59	**0.89**	0.99	0.69
IBM LCD	**1.60**	1.92	1.12	**0.85**	1.02	0.59	**1.00**	1.20	0.69
Fujix Print	**0.16***	0.64	0.16	**0.09***	0.34	0.09	**0.10***	0.39	0.10

The Sony CRT used in this experiment had slightly lower threshold than a similar model used by Stokes1 in 1991 indicating that the newer model has improved. The lack of fit for the SGI, IBM and Fujix displays makes interpretation of their results difficult. See Fig. 6.

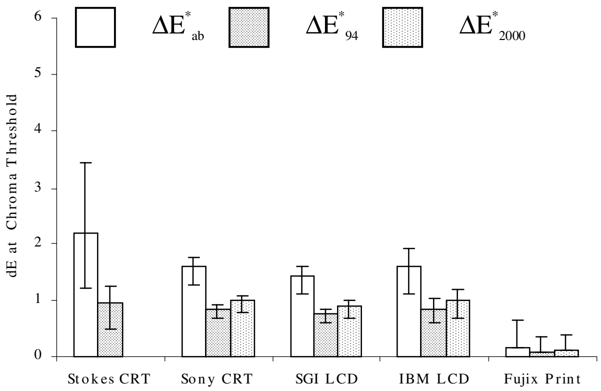

Figure 5. ΔE for Chroma Thresholds

Table 6. Delta E's for Hue Thresholds

	ΔE^*_{ab}		ΔE^*_{94}		ΔE^*_{2000}	
Stokes CRT	**2.52**	3.47 2.01	**1.80**	2.16 1.28	N/A	
Sony CRT	**1.73**	1.98 1.45	**1.31**	1.50 1.10	**2.07**	2.37 1.74
SGI LCD	**1.09***	1.50 0.39	**0.82***	1.14 0.30	**1.30***	1.80 0.47
IBM LCD	**5.93***	7.98 2.81	**1.25***	1.69 0.59	**1.98***	2.68 0.94
Fujix Print	**0.14***	0.76 0.22	**0.11***	0.57 0.17	**0.17***	0.91 0.27

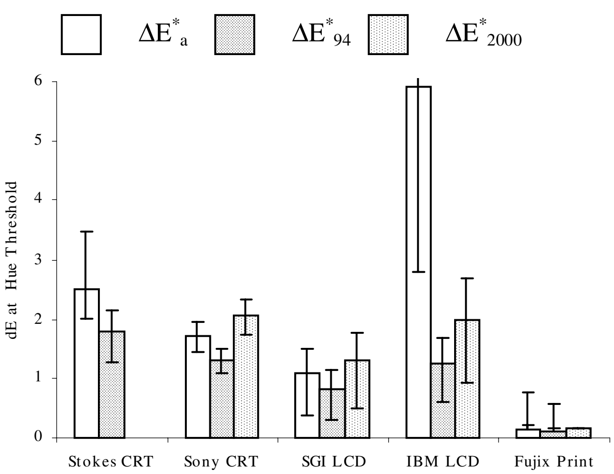

Figure 6. ΔE for Hue Thresholds

Optimizing (l:c) ratios

One of the features of color difference formulas such as CMC, and more recently ΔE^*_{94} and ΔE^*_{2000} is that they include parametric correction factors (l:c in CMC, k_L, k_C, k_H in CIE94 and CIE2000) to account for variation in experimental conditions from the defined reference conditions. In practice, the hue term is typically set to unity, and the lightness and chroma parameters are adjusted to better fit the perception data.

The example below shows the effects of using the k_L and k_C terms in CIE94 to improve the uniformity of the three CIELAB dimensions on the Sony display. Because the data collected varied uni-dimensionally in lightness, chroma or hue, optimization of the l:c ratio was straightforward. The three $k_\bullet$ terms were set equal to the ratio of each dimension's threshold to that of the hue threshold. The k_H term is thus set to unity, and the lightness and chroma thresholds are adjusted to be equal to the hue threshold.

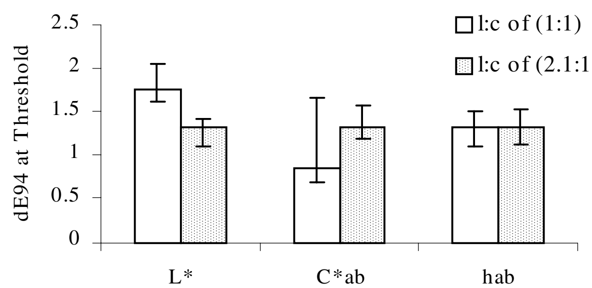

Figure 7. Effects of l:c optimization

In most cases, the data will not or can not be considered to be (as was done in this experiment) uni-dimensional. In these cases, a regression approach can be used as described in Berns.[8] A regression equation is built by re-arranging the terms in CIE94 as shown in Eq. 4. In this way, the $k_\bullet$ terms become the independent variables using the transforms in Eq. 5. The β_0 intercept term is included to allow for a global bias adjustment. If desired, the three $k_\bullet$ terms can each be divided by kH to produce the familiar (l:c) ratio optimization.

$$(\Delta V)_i^2 = \beta_0 + \beta_L\left(\frac{\Delta L^*}{S_L}\right)_i^2 + \beta_C\left(\frac{\Delta C^*_{ab}}{S_L}\right)_i^2 + \beta_H\left(\frac{\Delta H^*_{ab}}{S_H}\right)_i^2 \qquad (4)$$

$$k_L = \sqrt{1/\beta_L} \quad k_C = \sqrt{1/\beta_C} \quad k_H = \sqrt{1/\beta_H} \qquad (5)$$

ΔV is the visual distance measured
i is the value for the i^{th} sample

S-CIELAB Colorimetric Threshold Results

Although the pixel-by-pixel differencing performed has intuitive appeal, it does not account for the potentially complex interactions between neighboring pixels in an image. To address this point, Zhang and Wandell's S-CIELAB[9] filters were evaluated.

S-CIELAB was developed as a spatial extension to CIELAB to be used in measuring color reproduction errors of digital images. It is applied by first separating the input image into an opponent-color space of Lum, R/G and B/Y. Each of the three channels are then convolved with a kernel whose shape is determined by the visual spatial sensitivity to that color dimension and a given viewing angle. The filtered channels are then transformed to a CIE-XYZ representation from which the CIELAB values and color differences are calculated.

One area S-CIELAB has had success is in predicting the differences a continuous-tone image and its halftone representation. Conventional color difference formulas would predict very large differences on a pixel-wise basis. However, at normal viewing distances, the dots within a halftone cell are blurred by the eye and on average match the color of the continuous tone original in that region. The viewing-angle and color dependent filters in S-CIELAB account for this blurring and give more reasonable results.

To determine S-CIELAB's effectiveness in the context of the current experiment, the appropriate filters were applied to each image and the pixel-wise color differences re-computed as described in the previous section.

Based on an initial analysis of the results, the S-CIELAB filtering had little effect in the context of this experiment. This was not an entirely unsuspected result. S-CIELAB is designed such that for uniform areas of color, the color difference predicted with be equivalent to that of an unfiltered version. The images used in this experiment were displayed with relatively high sampling rates and at distances such that individual pixels were not visible and were thus only minimally altered by the S-CIELAB filters.

Conclusion

Four experiments have been conducted to determine the visual thresholds for color changes in pictorial images displayed on four displays. Thresholds were expressed in terms of parametric values and colorimetric differences using three CIE color difference formulas.

For sigmoidal L* compression, the Fujix prints had the lowest sensitivity. For multiplicative reduction in C^*_{ab}, the Fujix prints had the highest sensitivity. For rotations in h_{ab}, the Fujix prints had the highest sensitivity. The SGI display was second most sensitive, and the IBM and Sony displays third. Colorimetric tolerance varied between 1 and 3 ΔE depending on the particular dimension and formula selected.

A possible explanation for the lightness results is that the tolerances depend on the dynamic range of the image displays. Since the printed images had significantly less lightness range that the soft displays, this might explain the significantly increased tolerances for the printers. For the chroma and hue dimensions, the results roughly correlate with display luminance. Since the colorfulness of the displays increases with luminance, it is reasonable to expect that observers might become more sensitive to small changes in hue and chroma since these would be perceptually amplified in a colorfulness space. Thus, one might cautiously conclude from these experiments that colorimetric tolerances for displays would decrease as display luminance and dynamic range increases. Further experiments would be required to confirm this conclusion and also to systematically investigate the influence of addressability, if any.

Acknowledgements

The author would like to thank Dr. Mark Fairchild for his advice during this project, and the IBM T.J. Watson Research Center for funding and supporting this project.

**Trademark or registered trademark of Sony Electronics, Inc.; Silicon Graphics, Inc.; or Fuji Photo Film Company, Limited.

References

1. M. Stokes, M.D. Fairchild, and R.S. Berns, Precision requirements for Digital Color Reproduction, ACM Transactions on Graphics V. 11, pg. 406 (1992).
2. Joan Uroz, Colour Difference Perceptibility for Large-Size Printed Images, MSc Thesis, University of Derby, Derby, (1999).
3. K. Schleupen, et al., High-Information-Content Color 16.3" Desktop-AMLCD with 15.7 Million a-Si:H TFTs, *Proc. Int. Display Res. Conf.*, p. 187 (1998).
4. Roy S. Berns, Methods for characterizing CRT Displays, Displays, 16, 173-182 (1996).
5. Mark D. Fairchild, David R. Wyble, *Colorimetric Characterization of the Apple Studio Display (Flat Panel LCD)*, Munsell Color Science Laboratory Technical Report, July 1998.
6. Albert Cazes, Gordon Braudaway, et. al., Color calibration of liquid crystal displays, *Proc. SPIE*, **3636**, pg. 154. (1999).
7. Jason E. Gibson, Mark D. Fairchild, *Colorimetric Characterization of Three Computer Displays (LCD and CRT)*, Munsell Color Science Laboratory Technical Report, (2000).
8. Mark D. Fairchild, Personal Communication (2000).
9. R.S. Berns, Deriving Instrumental Tolerances from pass-Fail and Colorimetric Data, *CR&A*, **21**, pg. 459. (1996).
10. X. Zhang, B. Wandell, A Spatial Extension of CIELAB for Digital Color Image Reproduction, *Proc. SID Symposiums*, **27**, pg. 731. (1996).

Biography

Jason Gibson received his BS degree in Imaging Science from the Rochester Institute of Technology in 1994. He returned to RIT in 1998 to pursue his MS in Color Science which he expects to receive in November 2000. Between degrees, Jason worked for Hewlett-Packard's Ink-Jet Media Division in San Diego. Jason currently works as a color scientist for Hewlett-Packard's Color Pipeline Team in Boise, Idaho. He is a member of IS&T and the ISCC.