

DERIVATION AND MODELLING OF HUE
UNIFORMITY

And Development of the IPT Color Space

by

Fritz F. Ebner

B.S. Carnegie Mellon University (1986)

M.S. University of Rochester (1990)

A dissertation submitted in partial fulfillment of the
requirements for the degree of Ph.D.
in the Chester F. Carlson Center for Imaging Science
of the College of Science
Rochester Institute of Technology

July 1998

Signature of the Author

Accepted by

Henry E. Rhody

Coordinator, Ph.D. Degree Program

9/24/98
Date

CHESTER F. CARLSON
CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Fritz F. Ebner
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation requirement for the
Ph.D. degree in Imaging Science

Dr. Mark D. Fairchild, Thesis Advisor

Dr. Roy S. Berns

Dr. Robert Rolleston

Dr. Kathleen Chen

9/24/98

Date

DISSERTATION RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
COLLEGE OF SCIENCE
CHESTER F. CARLSON
CENTER FOR IMAGING SCIENCE

Title of Dissertation: Derivation and modeling of hue uniformity and development of the IPT color space

I, Fritz F. Ebner, hereby grant permission to the Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature: _____

Date: 9/24/98

DERIVATION AND MODELLING OF HUE
UNIFORMITY

And Development of the IPT Color Space

by

Fritz F. Ebner

Submitted to the
Chester F. Carlson
Center for Imaging Science
College of Science
in partial fulfillment of the requirements
for the Ph.D. degree
at the Rochester Institute of Technology

ABSTRACT

Metric color spaces have been determined to be significantly non-uniform in the hue attribute of color appearance. Several independent sources have confirmed the non-uniformity. A data set was obtained during the course of this thesis work that contains the largest sampling of color space to date which can be used to compare models of color appearance. The data set obtained was compared to existing data sets and found to correspond closely. Lookup table methods were employed to test significant differences between data sets. A simple modeling approach was taken based on commonly understood color space models and knowledge of the visual system. Several color spaces can be derived using the simple model, and one was chosen that models hue uniformity very well and has other desirable attributes. This new color space is named IPT. Many visual data sets were plotted in the IPT color space and all show improved performance over industry standard color spaces. The IPT color space has applications in color data representation, gamut mapping, and color appearance modeling.

ACKNOWLEDGEMENTS

I wish to acknowledge all the observers who selflessly (but not without reward) took time from their busy schedules to stare at a CRT screen and make hard decisions. Names are listed in arbitrary order.

Adam Stein	Bob McCarthy	Brad Smith
Alex Vaysman	Eddy Dalal	Judy Stinehour
Bob Karz	Gaurav Sharma	Tom Lynch
Brian Hawkins	Brian Wiley	Chrissy Larson
Chris VanCampen	Dave Birnbaum	Dorothy Jaehn
Dave Odgers	Hong Li	Rob Rolleston
Jackie Holmes	Jorge Cajiga	Juliet Ong
Karen Braun	Kathy Loj	Laura Flick
Laura Caruso	Liesl Wiley	Martha Birnbaum
MaryJo Ebner	Me	Dianne Steely
Nagesh Narendranath	Nancy Goodman	Raja Balasubramanian
Phyllis Giambrone	Genia Sychtycz	Steve Reczek
Rene Rassmussen	Gus Braun	Jutka Bolgar
Ron Macera	Roger Chan	Scott Bennett
Steve Bloomberg	Skip	Steve Linder
Sue Zoltner	Vlad Bravve	Candice Dobra
Mojgan Rabanni	Jim O'neill	Nancy Mazur
Dick Losz	Ron Johnson	Pat Colclough
Marybeth Poupart	Gloria Wimer	Pat Donaldson
Doug Kreckel		

I would like to acknowledge the following people for going well beyond the call of duty (or a free lunch) by making many observations, providing valuable feedback on the experimental design and presentation to the observers, helpful discussions, and general support: Jacqueline Holmes, Karen Braun , Dorothy Jaehn, Lee Cass, and Dave Birnbaum. I would also like to acknowledge the help of Dr. Robert Chapman from University of Rochester's Center for Visual Sciences, for help with the analysis of semantic differential scale data.

DEDICATION

I would like to dedicate this work to my sister Meg with whom I shared the first 23 years of my life. Although she never knew it, she showed me that it is possible to be courageous in the face of incredible adversity, and that you can keep your sense of humor through it all.

Table of Contents

1	INTRODUCTION	1
2	BACKGROUND	3
2.1	LITERATURE SEARCH	4
2.1.1	<i>Color Order Systems and Color Harmony.....</i>	4
2.1.1.1	Munsell	4
2.1.1.2	Ostwald.....	8
2.1.1.3	NCS.....	10
2.1.1.4	Moon and Spencer.....	14
2.1.1.5	Others.....	16
2.1.2	<i>Color Naming and Color Meaning</i>	17
2.1.2.1	Basic Color Names	18
2.1.2.2	Color Meanings	20
2.1.3	<i>Preferred Reproduction of Pictorial Images.....</i>	23
2.2	PRELIMINARY EXPERIMENTS	30
2.2.1	<i>Experiment P1 (not completed). Semantic differential scaling of colors in the context of business graphics.....</i>	39
2.2.2	<i>Experiment P2. Gamut mapping through the use of color matching (Gamut mapping from below)</i> 46	
2.2.2.1	Perceptually closest color matches (sub-experiment 1).....	46
2.2.2.1.1	Abstract: to P2 (sub-experiment 1)	46
2.2.2.1.2	Introduction to P2 (sub-experiment 1).....	47
2.2.2.1.3	Experimental.....	50
2.2.2.1.4	Conclusions to P2 (sub-experiment 1).....	74
2.2.2.2	Preserving meaning in color matches (sub-experiment 2).....	76
2.2.2.3	<i>Conclusions leading to body of thesis work.....</i>	80
3	APPROACH AND RESULTS	81
3.1	EXPERIMENT E1. FINDING SURFACES OF CONSTANT HUE IN COLOR SPACE.....	86
3.1.1	<i>Abstract</i>	86
3.1.2	<i>Introduction</i>	86
3.1.3	<i>Experimental</i>	88
3.1.3.1	Script.....	89
3.1.3.2	User interface.....	90
3.1.3.3	Viewing conditions.....	92
3.1.3.4	Calibration	92
3.1.3.5	Color selections	92
3.1.3.6	Observers	93
3.1.4	<i>Results and Discussion</i>	95
3.1.4.1	Data set calculation.....	95
3.1.5	<i>Conclusion</i>	110
3.2	DEVELOPMENT OF CONSTANT HUE COLOR SPACES.....	111
3.2.1	<i>Neural Nets</i>	111
3.2.1.1	General overview.....	111
3.2.1.2	Modeling the CIELAB transformation	113
3.2.1.3	Modeling the constant hue data set	114
3.2.2	<i>Color space visualization and parameter tuning</i>	116
3.2.2.1	Model Descriptions	117
3.2.2.2	Visualization tool development.....	122
3.2.2.2.1	Visualization of color spaces.....	123
3.2.2.2.2	Controls for parameter tuning	128
3.2.2.2.3	Quantitative metrics for hue uniformity and psychophysical data set visualization.....	138
3.2.2.3	Properties of the chosen color space (IPT)	142
3.2.2.3.1	Engineering Considerations	142

3.2.2.3.2	Model and Coefficients	146
3.2.2.3.3	Color Matching Functions.....	147
3.2.2.3.4	Constant Hue data sets	149
3.2.2.3.5	Munsell Value 5 data.....	152
3.2.2.3.6	Neutral Lightness Response.....	153
3.2.2.3.7	Chromatic Lightness Response.....	154
3.2.2.3.8	MacAdam (PGN) Discrimination Ellipses and Spectral Locus	159
3.2.2.3.9	Suprathreshold Color Difference Ellipses	160
3.2.2.3.10	OSA Color System medium gray ($L = 0$) constant lightness plane	161
3.3	EXPERIMENTS E2. VERIFICATION EXPERIMENTS	162
3.3.1	<i>Comparing CIELAB to Hung & Berns and Ebner data sets</i>	162
3.3.1.1	Introduction	162
3.3.1.2	Experimental.....	163
3.3.1.3	Results.....	165
3.3.2	<i>Comparing IPT color space to Hung & Berns and Ebner data sets</i>	166
4	CONCLUSIONS	170
5	FUTURE WORK.....	172
6	APPENDICES.....	173
6.1	APPENDIX A: DATA PLOTS FROM EXPERIMENT P1.....	173
6.2	APPENDIX B: DATA FROM EXPERIMENT P2.....	175
6.3	APPENDIX C: THREE PARAMETER FILES CORRESPONDING TO FIGURE 57	177
7	REFERENCES	179

List of Figures

Figure 1. Progression of thesis research direction.....	3
Figure 2. Regions of confusion in color space.....	15
Figure 3. MacAdam's tone compression curves for tone reproduction with limited dynamic range.....	28
Figure 4. Taxonomy of business graphics image content.....	34
Figure 5. Different colors for the face. Face colors that are unnatural look "wrong"	35
Figure 6. Different colors for cars. There is no "natural" color for a car.	35
Figure 7. Different colors for figuratively representative images.	36
Figure 8. Figuratively representative images of automobiles with different colors.....	36
Figure 9. Reduced version of task and semantic differential scales for experiment P1 (abstract sub-class)	41
Figure 10. Example images from P1 experiment. Abstract and Data representative images... ..	41
Figure 11. Variance accounted for by factors from PCA	43
Figure 12. User interface for experiment P2 (sub-experiment 1).....	51
Figure 13. Images presented to the observers for experiment P2 (sub-experiment 1)	52
Figure 14. Example hue leaf of CRT and Printer gamut shapes at $h_{ab}=303^{\circ}$	56
Figure 15. Histogram of ΔE^*_{ab} for exact match exercise for entire observer group including outliers.....	58
Figure 16. Median match values for 25 colors in a^*b^* plane, and in $L^*-C^*_{ab}$ planes.....	62
Figure 17. Example gamut mapping vector with component vectors.	63
Figure 18. 2D weighted vector gamut mapping. Open arrow heads indicate observer results. Filled arrow heads indicate model predictions. Hue leaves shown are for the hue angle of the observer match.....	65
Figure 19. 3D weighted vector gamut mapping. Open arrow heads indicate observer results. Filled arrow heads indicate model predictions. Hue leaves shown are for the hue angle of the predicted match.	66
Figure 20. Hue angle matches on left, and "corrected" hue angle matches on right (using Hung's Constant Lightness data).	68
Figure 21. Comparison of Observer results with Katoh and Ito 1:2:2 vectors. Open arrow heads indicate observer results. Filled arrow heads indicate Katoh and Ito results. Hue leaves shown are for the hue angle of the predicted match.	70
Figure 22. Comparison of Observer results with best 2d enhanced Katoh optimization parameters. Open arrow heads indicate observer results. Filled arrow heads indicate model results. Hue leaves shown are for the hue angle of the observer match.....	72
Figure 23. Comparison of Observer results with best 3d enhanced Katoh model optimization parameters. Open arrow heads indicate observer results. Filled arrow heads indicate model results. Hue leaves shown are for the hue angle of the model prediction.	73
Figure 24. Images used in experiment P2, sub-experiment 2.....	77
Figure 25. Histograms of colors where the null hypothesis was rejected between sub-contexts	78
Figure 26. Mean mapping vectors for each of the 5 sub contexts.	79
Figure 27. Mean mapping vectors of Text sub-context (open arrow heads) and the mean of all other sub-contexts pooled (filled arrow heads).	79
Figure 28. Munsell renotation colors of Value 5 plotted on CIELAB a^*-b^* plane.....	81

Figure 29. Munsell renotation colors with extrapolation boundary and CRT gamut projection onto CIELAB a^*-b^* plane.....	83
Figure 30. Hung and Berns loci of constant perceived hue. Left plot is CL data, right plot is VL data.....	84
Figure 31. User interface for experiment.....	91
Figure 32. Hue leaf sampling at 0 degrees. The large square is the reference color.	94
Figure 33. Projection of test colors onto a^*-b^* plane. Large squares indicate reference colors.	94
Figure 34. Histograms of hue matches for 90 observations at reference hue angle of 0 degrees. Histograms are plotted in the respective locations in the $L^*-C_{ab}^*$ plane of the reference colors.....	95
Figure 35. Average and standard deviation of 95% confidence limits of mean hue matches as a function of reference hue angle. Units of y axis are in degrees (CIELAB Δh_{ab}).....	96
Figure 36. Histogram over the entire data set (306 colors X 30 observation means = 9180) of weights calculated from inverse absolute difference.....	97
Figure 37. Histogram over the entire data set (306 colors X 30 observation means = 9180) of weights calculated from inverse variance.....	98
Figure 38. Weighted mean hue matches in CIELAB space with + - 95% confidence limits.	100
Figure 39. Projection of constant hue surfaces onto CIELAB a^*-b^* plane.....	101
Figure 40. Loci of constant hue in CIELAB a^*-b^* at different lightness levels. Dots represent reference colors.....	101
Figure 41. Weighted mean hue matches plotted in CIECAM97s color appearance space. The x-axis is labeled by CIELAB reference hue angle.	102
Figure 42. Projection of constant hue surfaces onto CIECAM97s $C \cos(h)$ - $C \sin(h)$ plane. Hue leaves are labeled by the CIELAB reference hue angles for comparison.....	103
Figure 43. Loci of constant hue in CIECAM97s $C \cos(h)$ - $C \sin(h)$ at different lightness levels. Dots represent reference colors.....	104
Figure 44. Quantification of average hue non-uniformity for CIELAB and CIECAM97s....	105
Figure 45. Quantification of maximum hue non-uniformity for CIELAB and CIECAM97s.	106
Figure 46. Constant hue surfaces shown in a linear opponent space ($Y, X-Y, Y-Z$). Numbers represent CIELAB reference hue angles.	107
Figure 47. Constant hue surfaces shown plotted in CIELUV color space. Numbers represent CIELAB reference hue angles.	108
Figure 48. Hung and Berns data plotted on top of experimental results in CIELAB space. Left graph compares CL (constant lightness) data. Right graph compares VL (variable lightness) data. Hung and Berns data are shown in bold dotted lines.....	109
Figure 49. 3 node, 1 hidden layer neural net	112
Figure 50. Tanh function is expansive in the negative range, linear near the origin, and compressive in the positive range.	112
Figure 51. Examples of nonlinear functions. Left plot is function f_1 , with parameters $\sigma_1 = 0.45$, $\sigma_2 = 0.6$. Right plot is f_2 with parameters $\sigma_1 = 1$, $\sigma_2 = 1$. Note that at the origin, f_1 has a slope of ∞	120
Figures 52. Visualization tool user interface showing the three slices of color dimension, constant lightness, constant hue, and constant chroma.....	125
Figure 53. Close up of the controls and constant lightness display.....	126
Figure 54. Close-up of the constant hue display.....	127
Figure 55. Two dimensional "slider bars" that controlled the pretransformation matrix.....	131

Figure 56. Examples of CIE 1931 color matching functions transformed through pretransformation matrix. Left image is through an identity matrix. The right image is through Hunt-Pointer-Estevez matrix.....	132
Figure 57. Three examples of CMFs along with their influence on the shape of the color space. Color space images show Hung and Berns CL data set and Munsell V5.....	133
Figure 58. Example hue lines from Guth type model. The first image is rotated at an angle of 0 degrees, the second is at 45 degrees. Note the shape difference of the constant hue loci.....	135
Figure 59. Nonlinear controls for visualization tool. Six parameters correspond to the positive and negative parts of each of the three channels. Functions take the form of hyperbolic, power function, or chroma compression.....	136
Figure 60. A line showing the path of a gradient sweep from gray to the blue primary (in CIE xyY chromaticity space).....	143
Figure 61. Path of a gradient sweep from gray to blue primary in CIELAB a* b* space. Notice that the mid chroma colors are more purple than the constant hue locus. Dotted lines are Hung and Berns constant hue loci.....	144
Figure 62. Path of a gradient sweep from gray to blue primary in IPT P-T space. Notice that the mid chroma colors are more purple than the constant hue locus. Dotted lines are Hung and Berns constant hue loci.....	145
Figure 63. Color matching functions of the IPT color space compared to the Hunt-Pointer-Estevez primaries. Note the larger response on the left tail of the M (green) response function.....	149
Figure 64. Constant perceived data sets plotted in IPT (left) and CIELAB (right). Hung and Berns ⁵³ Constant Lightness (CL) data set is in dotted lines. IPT is plotted with P and T scaled by 100.....	150
Figure 65. Quantification of hue nonuniformity for CIELAB, CIECAM97s, and IPT.....	151
Figure 66. Maximum deviation quantification used in the visualization tool.....	152
Figure 67. Munsell Value 5 data plotted in IPT (left) and CIELAB (right)	153
Figure 68. IPT lightness correlate as a function of CIELAB L*.....	154
Figure 69. Chromatic lightness response of CIELAB. x axis is L* of chromatic stimulus, y axis is L* of neutral stimulus that matched in lightness (mean observer judgement). RMS error between observed lightness and L* of chromatic color is 7.12.....	155
Figure 70. CIELAB matching L* response as a function of predicted $L_{new}^* = L^* + 0.098 C_{ab}^*$. RMS error between observed lightness and predicted is 5.97.....	155
Figure 71. Chromatic lightness response of IPT. x axis is L* of chromatic stimulus, y axis is L* of neutral stimulus that matched in lightness (mean observer judgement).). RMS error between observed lightness and L* of chromatic color is 6.73.....	157
Figure 72. Plot of final lightness predictor for IPT color space for chromatic colors. RMS error between observed lightness and predicted is 3.5.....	158
Figure 73. MacAdam ellipses plotted in IPT (left) and CIELAB (right) along with the spectral locus.....	159
Figure 74. RIT-DuPont visual color-difference ellipses plotted in IPT (left) and CIELAB (right).....	160
Figure 75. OSA color system medium gray constant lightness plane plotted in IPT (left) and CIELAB (right).....	161
Figure 76. User interface for verification experiment.....	163
Figure 77. Scale values for judged uniformity between CIELAB and the 2 constant hue data sets. H&B are Hung and Berns data, E&F are Ebner and Fairchild's data.....	165

Figure 78. Scale values for judged uniformity between IPT and the 2 constant hue data sets. H&B are Hung and Berns data, E&F are Ebner and Fairchild's data.....	167
Figure 79. Mean scale values for colors and descriptive terms for abstract representative image type.....	173
Figure 80. Mean scale values for colors and descriptive terms for data representative image type.....	174

List of Tables

Table 1. Bipolar scales arranged according to their factor loadings	44
Table 2. Model parameters for weighted vector gamut mapping model.....	65
Table 3. Summary of optimization results for Katoh and enhanced Katoh models.....	71
Table 4. Transformation error from hue correction LUTs.....	116
Table 5. Luminance and chromaticities of viewing setup.....	175
Table 6. Colors chosen for gamut mapping color match.....	176

List of Equations

Equation 1. Slope calculation for weighted component vector gamut mapping.	63
Equation 2. Weighted color difference equation.....	69
Equation 3. Enhanced weighted color difference model.....	71
Equation 4. Matrix form of forward computation of 3 node hidden layer neural network ..	113
Equation 5. General form of CIELAB type color model.....	118
Equation 6. Matrix to rotate second and third row elements for the opponent response.....	118
Equation 7. Functional forms for nonlinear step.....	119
Equation 8. General form of CIELUV type color model.....	121
Equation 9. Inverse CIELUV type equation	122
Equation 10. Forward IPT model and coefficients.....	146
Equation 11. IPT inverse model and coefficients	147
Equation 12. Final lightness predictor for chromatic colors from Fairchild and Pirrotta ⁶⁸ . RMS error between observed lightness and $L^{**}(L^*, C_{ab}^*, h)$ is 4.2.	156
Equation 13. Final lightness predictor for IPT color space for chromatic colors from Fairchild and Pirrotta ⁶⁸ . RMS error between observed lightness and predicted is 3.5.....	158

1 Introduction

The progression of this thesis work that led to the eventual characterization of hue uniformity and development of uniform hue color spaces was not linear. In fact, it was more evolutionary, in the sense that the fittest paths to completion survived, whereas the paths that offered less opportunity were abandoned. This dissertation is intended to capture all of the work that was conducted from the completion of the thesis proposal to the completion of the work that was deemed necessary and sufficient in order to graduate.

The background section covers the study preceding the research proposal, as well as two experiments that were performed subsequent to completion of the proposal. These subjects are related to the study of two things. Firstly, the study of color meaning in different contexts as defined by image content, using semantic differential scaling. Secondly, the study of observer color matches to derive gamut mapping models.

The approach section, as well as following sections concentrate on the path that has led to the culmination of the research efforts, which is the characterization and modeling of hue uniformity of color appearance spaces. A large psychophysical experiment was conducted to find surfaces of constant hue in CIELAB color space. Fifteen equally spaced hue angles in CIELAB color space were sampled. Thirty observers performed a hue matching task three times each over 306 colors covering the full gamut of a CRT display. A computer program was created to model color spaces using variations of several simple models. Data from two constant perceptual hue experiments were used to define metrics of uniformity for optimizing

the metric space. Several other visual data sets were used to provide feedback on uniformity and accuracy of color appearance phenomena. A new color appearance space is proposed that has acceptable hue uniformity, models several other visual attributes as well as or better than commonly accepted and industry standard color spaces, and is more closely related to our understanding of the human visual system.

Figure 1 shows the progression of the work performed throughout the research period. The first two experiments, labeled P1, and P2 (P stands for proposal, or preliminary), were part of the initial proposal that was signed off by the advisory committee. P1 did not yield results that were deemed significant enough to follow, thus the path was abandoned. P2 was published in a *Color Research and Application* article¹, and part of the results from this experiment showed that the non-uniformity in hue in CIELAB color space were so bad, the color name was changed when using the color space to perform gamut mapping. The available data sets for constant perceived hue were found to be insufficient, so the subsequent experiments (E1 and E2) were designed around hue uniformity. E1 was presented at EI98 in San Jose in January 1998.² Study of hue uniformity has been significant enough to warrant the completion of the degree based on this work.

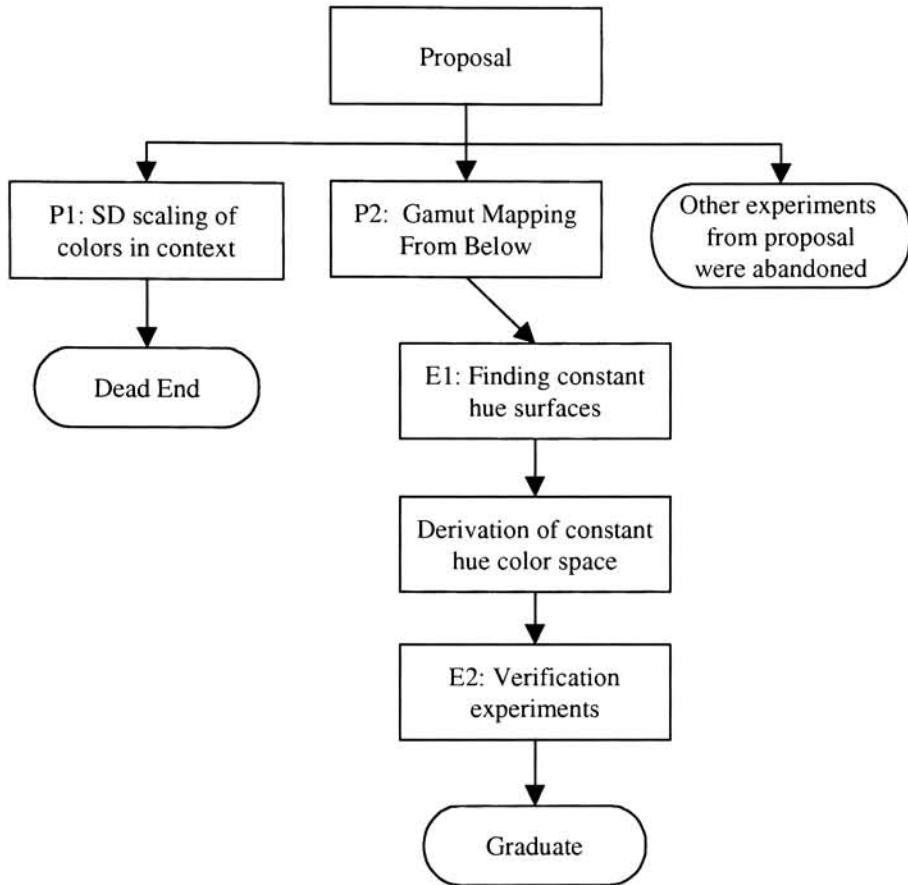


Figure 1. Progression of thesis research direction.

2 Background

The background section is comprised of two parts. First, a literature search is discussed on the subjects of color order, color meaning, color harmony, and color preference. Secondly, experiments that led up to the final thesis subject are described.

2.1 Literature Search

2.1.1 *Color Order Systems and Color Harmony*

A color order system is a way to describe the relationship between colors. Each color order system attempts to create a space within which colors can be specified and compared in an intuitively clear way. The fact that there are so many such systems implies that there may not be a most appropriate way to specify colors for every situation and within every context. From a color harmony point of view, there are a few important color order systems that respective theories are based upon. Color order systems of Munsell, Ostwald, Hard and Sivik, and Moon and Spencer are thus described along with the basic foundations of their respective theories of color harmony. Relevant experiments will also be discussed.

2.1.1.1 *Munsell*

Albert H. Munsell^{3,4} was born in Boston, Massachusetts in 1858. He attended college at the Massachusetts Normal Art School. Following this, he attended the Julien Academy in Paris on a graduate fellowship. He won second prize in the Beaux Arts competition for his painting “The Ascension of Elijah”. Between 1898 and 1905, he developed his ideas on color order, notation, and arrangement.

Although many others before Munsell have suggested ways to arrange colors, he was the first to create an atlas of physical samples that is perceptually uniform. The Munsell color system is arranged as a three dimensional solid with three dimensions of Value(Lightness), Chroma, and Hue. The hue circle is arranged so that there is an equal perceptual distance between each

major hue category. Munsell's major, or simple hue categories were red, yellow, green, blue, and purple. Purple is included here because Munsell thought that there was a larger perceptual distance between red and blue than there was between red-yellow, yellow-green, and green-blue. Intermediate hues, or compound hues, are defined between each of the major hue categories. These ten hue divisions make up the hue circle. The hue circle is then divided ten fold for each simple hue. The value 5 is assigned to each simple hue and 10 to each compound hue. The notation for a given color is defined by a triplet of hue, chroma, and value. The value axis goes from 0-black, to 10-white. The chroma scale goes from 0-neutral, to an open ended high chroma number, although larger values have been assigned by some. On the physical atlas of color, some pigments are naturally stronger than others. Therefore, the strongest pigment of blue-green only goes to chroma of 5 whereas the strongest pigment of red goes to 10. Furthermore, the maximum chroma may not happen at value 5. For example yellow reaches maximum chroma at a value of 7, whereas purple-blue reaches maximum chroma at value 4.

The Munsell notation for a particular color is <hue designation> <value>/<chroma>. For example, a pink color might have a Munsell notation of 5R 7/3. Neutral colors are denoted N.

Munsell's system does not directly correspond to any theory of color vision, as do other systems. Munsell was interested solely in the perception of color and did not attempt to explain how or why.

Munsell, felt that there was a need for balance in every aspect of one's life. This included art and color. He attempted, through heuristics and design tenets, to specify rules for creating harmony and balance in color. Balance, as defined by Munsell, is visual comfort. Balance means that a color scheme is neither too light or too dark, too weak or too strong, too hot or too cold. Balance is described by several general laws:

- The simplest balance of two colors is a balance about neutral gray. This means that two colors that sit on a line drawn through N5 and are equidistant from N5 will be balanced. These colors are considered to be complementary.
- Area of color combinations is important. A strong red will balance a larger field of blue green of small chroma. “The stronger the color we wish to employ, the smaller must be its area, while the larger the area, the grayer or weaker the chroma.”² Area is related to value and chroma by the rule that the product of value and chroma should be inversely proportional to the respective areas of the two colors in question. For example, 5R 7/6 will balance 5R 3/3 in a proportion of 9 parts former to 42 parts latter.
- Colors of the same hue can be made to balance if they are created with different values and chromas. It seems always to be best if colors are connected by a straight line whose center point falls on or around a value of 5 (mid-lightness).
- A small amount of discord is not always undesirable since it makes the scene more interesting.

There are 9 chief features of balance as described by Faber Birren in “Munsell, A grammar of Color”⁵. These are:

1. Gray colors harmonize best when neatly and evenly spaced. Its best to center around N5.
2. Harmonies with only one hue look best when they are arranged neatly and evenly around a center point of chroma /5.
3. Opposite colors of medium chroma /5 can be combined in equal area.
4. Opposite colors of equal value but of different chroma should follow the above mentioned area law.
5. Opposite colors of the same chroma but different value should fall on a straight line through, and equidistant from, N5.
6. Opposite colors of different value and different chroma should follow the area law.
7. Munsell believed, as did others, that color combinations most liked by people comprise either colors that are closely related or colors that are in contrast. “Neighboring colors should be neatly stepped as to value and should find sequence at middle value 5. They should be of the same chroma for good balance. One color could have strong chroma and the other could have weak chroma”. The area rule should apply to these colors.
8. There is harmony in diminishing sequences. These are sequences that start with high value and chroma and each successive color is reduced in chroma and value, and possibly rotated in principle hue.
9. Munsell liked the path of an ellipse through color space from one color to its compliment.

All of Munsell's notions about harmony and balance centered around his color system and simple geometric relations between colors in this space. These relations were drawn from experience as an artist and designer, and as such should be viewed to have more validity than rules that have been generated by workers such as Moon and Spencer¹³ who did not have first hand experience with creating art. Munsell defined complementary colors as those opposite on the hue circle, and characterized them as most strongly contrasting. He also defined them as two colors that made neutral gray when spun on a Maxwell disk. He goes on to say that an admixture of any two colors of equal proportions makes a color that on the Munsell diagram lies at the center of a line connecting the colors. This confusion between perception and stimuli is interesting because it is the only place in his writings that he infers psychological relationships from indirect observation of different stimuli.

2.1.1.2 *Ostwald*

Wilhelm Ostwald^{6,7} was around 60 years of age before he became interested in color. Before this, he had worked in the fields of mathematics, chemistry, and physics. Ostwald's system is comprised of planes of constant dominant wavelength, each of which is defined by a black point, a white point and a "fullcolor". A fullcolor is defined by Ostwald to be an idealized spectral "block dye". There are enough fullcolor descriptions to span the hue circle, thus enabling a 3 dimensional solid to be formed. Ostwald created his system before the CIE 1931 standard observer was adopted, and before Stevens started work on his power law (1953)⁸. Thus he relied on Fechner's law to create equal spacing of colors. Ostwald created his physical atlas from the use of disk colorimetry by mixing proportions of black, white and a

physical analog of a “fullcolor”. All colors that sit in a given triangle are comprised of a mixture of black, white and a given fullcolor. White and black content are denoted by lower case letters of the alphabet, excluding the letter j. Ostwald defined the equation $C+W+B=1$, so only two coordinates need be known to define the position in the triangle. The triangle is set up such that colors of equal white content fall on a line that is normal to a line that intersects the white point, is normal to the equal white content line, and is parallel to the line connecting black and fullcolor. The analogous rule applies for colors of equal black content. Colors of equal fullcolor content form vertical lines.

Colorimetrically, the Ostwald system neither spans the entire visual space (as defined by the CIE xy chromaticity diagram), nor is it perceptually uniform or internally consistent. However, Ostwald’s ideas about color harmony are quite interesting. Because Ostwald defined all colors as mixtures of black, white, and a pure color, his system fits well with how graphical artists deal with, and think about, colors i.e. in tints, tones, and shades. Faber Birren, in his analysis of the Ostwald system, said “When contemplated in the light of average human appreciation, colors are seen not so much in terms of lightness or darkness, but rather in terms of whiteness, blackness, grayness. A lavender (white-containing purple) and a sage green (black-containing yellow) may have equal lightness. Yet they are hardly to be called harmonious.”³

2.1.1.3 NCS

The Natural Color System (NCS) was designed by Hard and Sivik⁹ and is a Swedish standard for color notation. The NCS is intended to be a system that can be used without a physical atlas for reference, although one was created for it for illustrative purposes. NCS provides a way to describe color without reference to a particular viewing situation, thus allows the ability to describe the perception without regard to a particular stimulus. The NCS space is arranged in accordance to Hering's color opponency theory, where yellow and blue are opposed, as are red and green. NCS notation is similar to Ostwald's notation in that a given color lies in a triangle whose vertices are white black and a maximum chroma color. Colors are described by having a resemblance to black, white, and to up to 2 of the 4 primary hues. Within the NCS triangle, whiteness and blackness are described in the same manner as the Ostwald system, and the equation $s+w+c=100$ is conspicuously similar, where the letter s denotes. The notation for the NCS system is a triplet, where the first number is blackness, the second number is chromaticness, defined as the distance from the achromatic axis, and the third element is an alphanumeric that denotes hue angle. The hue circle is broken into quadrants of yellow, red, blue, green, and the hue notation describes the angle from one primary to the next. They make the point over and over again that "subjective phenomena can only be meaningfully measured using subjective methods...Thus, the NCS, as a system for denoting colors as and when they appear to man, is general and not based on the availability of the NCS Color Atlas."

Hard and Sivik have criticized many previous workers in color harmony for creating long lists of dogmatic rules that are mostly opinion and not based on any underlying general rules.¹⁰ They give as an example the definition of complementary colors. There are at least five different, somewhat contradictory, definitions of complementary colors that are founded entirely on what they call “stimulus tricks” instead of perception (for instance that two colors mix to gray on a Maxwell disk makes them complementary). Many of the rules of color harmony of Goethe, Munsell, Moon and Spencer, and Chevreul have been shown to be too rigid and not terribly useful in real world situations.

Hard and Sivik have proposed a model of color combinations in which there are 9 dimensions.¹¹ These are

- Interval, with “subdimensions” of : Distinctness of Border, Kind, and Size
- Chord, with subdimensions of: Complexity, Content, and Type
- Tuning, with subdimensions of : Color similarities, Area relations, and Rhythm.

Interval indicates the contrast between colors. Distinctness of border describes the borderline between color surfaces and how sharp the edge is perceptually. Kind denotes where in space the colors of the pair reside. Size is the distance the colors are from one another. The interval dimension can be thought of as a more detailed description of color difference. Chord describes how individual colors are experienced together. Complexity describes the number of main attributes (the six elementary colors) that are represented in a combination. Content says

which main attributes are in the combination. Type denotes what others may have called complementary colors. It specifies whether colors on the hue circle are opposite one another or in the same quadrant, or in the same half-circle. Tuning refers to the “Harmonics” of the combination. Color similarities describes what aspects of the colors in combination are the same, i.e. which attributes of each color are similar. For example, colors could be all of the same blackness, or whiteness, or chromaticness, etc. Area relations describes the relative sizes of the colors in question. Rhythm has to do with the regularity of a pattern or texture. A repeating pattern with a small period would have lots of rhythm.

This nine dimensional approach to color combinations seems excessive, and the dimensions are hardly orthogonal in concept. These attributes of perception, however, are interesting and should be taken into account when designing experiments dealing with the subject.

Hard and Sivik also define what they term the “color gestalt,” which is the overall experience of color and form when viewed in the real world. The basic elements of their color gestalt are:¹²

1. Color elements: Described by the shape of the contour that separates the color from its surround and the actual color of the shape. There is no mention of smooth gradations from highlight to shadow, but these presumably are part of the same color element.
2. Line Network: The pattern of lines that make up the borders around the colors in a particular area.

3. Overall Form: The pattern that the color elements and line network form. There may or may not be an overall form. This depends on if there is a rhythm in the area in question.
4. Color Character: The color character can change if the color elements change, even if the distinctness of border remains the same. The color character complements the form gestalt to make up the overall gestalt.
5. Balance: This is what may be known in other spheres as color harmony. It takes into account color elements, area balance, rhythm balance, etc.
6. Totality: The influence of context on the rest of the gestalt.

Quantification of some of these aspects of color gestalt may be difficult to obtain.

The study of distinctness of border is of interest to both Hard and Sivik's theory of color combinations and of their gestalt ideas.¹⁰ They show that distinctness of border between adjacent color fields can be thought of as a new way to think about color difference from a phenomenological point of view. Distinctness of border, or GT as they abbreviated it (from Swedish) was shown to be a one dimensional function. The research on the NCS lightness function yields a blackness that is a function of CIE Y as follows:

$$s = 100 - w = 100 - \left[\frac{156Y}{(Y + 56)} \right]$$

which says that blackness (or whiteness) is a hyperbolic function of CIE Y. Although functionally different, the shape corresponds roughly to that of Stevens or Fechner. The

interesting part of this is that a constant difference in GT along the achromatic scale corresponded to a constant blackness (or whiteness) difference. This makes NCS blackness perceptually uniform with both blackness difference and with border distinctness, GT. This relationship also holds with chromatic color pairs where the hue and chromaticness were kept constant and the blackness varied. An equation was derived through multiple linear regression that predicted distinctness of border $GT = f(\Delta s, \Delta c, \Delta \phi)$ with a very high correlation coefficient. Furthermore, it was found that GT was not an additive quantity, i.e. if GT between A and B is the same as GT between B and C (on the achromatic axis), the GT between A and C is not AB + BC. This lack of additivity is not surprising if one considers that center-surround antagonism of the visual system enhances contrast sensitivity at high frequency edges. The assumption is made here also that the borders between colors is sharp. Very different results would have been found if the borders between the stimuli were even slightly blurred.

2.1.1.4 *Moon and Spencer*

Perry Moon and Domina Spencer¹³ took the CIE specification of the Munsell Color Order System and created their “metric color space”, which is basically just a warping of the CIE XYZ space to make it appear more perceptually uniform. They sought to create a mathematical foundation for the rules of color harmony through the use of their new color space. Their two basic assumptions were :

1. Any arrangement of colors that can be sensed as an orderly combination will be pleasing. Thus, in the new ϖ space, simple geometric figures connecting colors will result in a pleasing harmony.

2. The interval (Euclidean distance in the ϖ space) between any two colors is unambiguous.

Moon and Spencer considered that any stimuli that are regarded as confusing will necessarily be disharmonious. From this they created their areas of ambiguity in their color space. For the hue circle, the regions are as in figure 2.

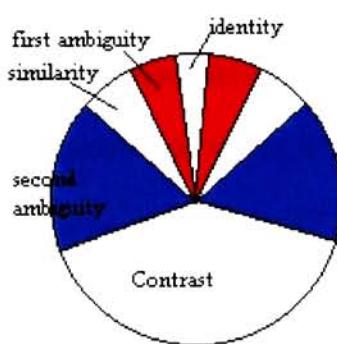


Figure 2. Regions of confusion in color space.

Regions of identity, similarity, and contrast are assumed to be pleasing, whereas the regions of the first and second ambiguities are assumed to be displeasing. The same regions are found in the other two planes in the space and are represented by concentric ellipses in lightness/chroma at constant hue values. Moon and Spencer (also referred to as M and S) go on to classify a large number of harmonies, in one, two and three variables, with corresponding names and denotations in Munsell terms. For harmony in area M and S state

“A pleasing balance among n color patches is obtained when the scalar moments about the adaptation point in ω -space are equal, for all the patches”¹⁴

Pope¹⁵ in his criticism of the Moon and Spencer work points out that they have greatly oversimplified the problem, that classification of harmonies into groups of simple figures looks nice, but needs to be validated by experimentation. He criticizes their application of Birkhoff's aesthetic measure¹⁶ $M=O/C$ (aesthetic measure is order divided by complexity) as a true mathematical equation, since it was intended more as a basic principle for guidance of design. On the ideas of regions of ambiguity Pope has disagreed, but this is dealing with an aesthetic context more generally associated with finer art, and not with graphic design, or page layout. Pope sums up his feelings in the quote “geometric significance in the symbolic representation does not necessarily mean visual significance.”¹³

In Granger's¹⁷ experiment on area balance in color harmony, Moon and Spencer's predictions with regard to area balance showed no predictive value, although Munsell's empirical rule regarding area balance was shown to account for 35% to 53% of the total variance. Strangely, there was a very high inter-observer correlation for area balance, between .67 and .73. This may suggest that there is a predictor for preference of area balance of two color mixes.

2.1.1.5 Others

Nayatani, et al.,¹⁸ in 1969 performed some experiments on two color harmonies. They conducted a paired comparison experiment on 102 two color harmony samples. From this, he derived, through linear regression, an equation to predict harmonies. Then he predicted the

aesthetic measures of 9890 two-color harmonies from the equation. From the predicted aesthetic measures, his data suggested that two color harmonies of same or similar hues tend to be more harmonious than complementary or contrasting hues. Also, colors harmonize better with N-9 (white), than with N-2 (blackish), and that aesthetic measure is affected mostly by value and value differences than other dimensions in color space. The experiment was conducted on pairs of colored patches without reference to particular context.

In summary, it seems that most color harmony models or systems, at best, provide some feeling to designers for a more systematic approach to synthesizing color harmonies. No one system has been shown to be general enough to apply tenets to all situations. To try to design an analysis system that employed one or more of the rules to combinations of color would both be a prohibitively large undertaking, and would not guarantee any useful results. The only way to glean useful information about color harmonies is to study observer's reactions to color combinations, both within the context it is to be used in, and over a wide sampling of the space. Only from results from study of particular context and application can rules be established for that context.

2.1.2 Color Naming and Color Meaning

Object colors are those colors that we experience in normal photopic viewing in the real world. Although viewing conditions can change drastically in the natural environment (and man made buildings), to a great extent colors of objects appear to remain the same, or at least quite similar. This phenomenon, called color constancy, is greater if colors are associated with familiar objects, e.g. people, fruits and vegetables, grass.

The ability to see colors as belonging to objects, in spite of the differences of (to a great extent) the actual physical stimulus reaching the eye allows humans to use color as a tool to help them survive in their environment. The ability to communicate these colors and to assign unambiguous names to them is a useful tool.

2.1.2.1 Basic Color Names

Naming of colors in a given language seems to conform to a fairly strict evolutionary process, whereby the language must pass through successive stages of evolution to get to the next level¹⁹. It has been shown that of the nearly 100 languages surveyed, a large majority of them follow the following rule of evolution (for existence of color names):

$$\left[\frac{white}{black} \right] \leq [red] \leq \left[\frac{green}{yellow} \right] \leq [blue] \leq [brown] \leq \left[\begin{array}{c} purple \\ \hline pink \\ \hline orange \\ \hline gray \end{array} \right]$$

This relationship means that a language most likely will have words for white and black before they have the word for red, and that the language will have a word for red before it has a word for either yellow or green. A language may have, after red, the word for either green or for yellow, but it must have both before it has a word for blue. These rules sound more like arbitrary synthesis rules, but they have been derived from study of a large number of languages in various stages of evolution. In the English language, there are an indefinite number of words or expressions that denote color. They are differentiated from the eleven basic color terms by the following characteristics:

1. The meaning of the color is not predictable from the meaning of its parts. e.g. blue-green.
2. Its significance is not included in that of any other color term. e.g. crimson is a kind of red.
3. Its application cannot be restricted to a narrow class of objects. e.g. blonde may be used to describe perhaps only hair, complexion, and furniture.

The ISCC-NBS Dictionary of Color Names²⁰ uses a similar attitude when assigning color names to volumes in the Munsell color solid. The hue circle is defined in terms of basic color names described above, with the addition of violet and olive. Intermediate hues are described by compound color names, e.g. reddish-orange. A slice of constant hue (value vs. chroma) is denoted using adjectival modifiers such as very pale, strong, deep, and light.

Naming of colors is an outcome of the evolution of a language. As the lexicon grows, more words are assigned to smaller differences in color (up to a limit). As the ISCC-NBS Dictionary of Color Names illustrates, in English, there are a huge number of color names. The meaning associated with, or evoked by a particular color is a less well defined problem, however. Connotation of meaning of a particular color depends on context, experience, culture, time, and other factors. There seems to be one invariant. Black seems to be the universal color of mourning and death. Most people dislike black. Ancient cultures had various and somewhat arbitrary meanings for the colors in their world²¹. To Leonardo da Vinci, yellow stood for earth, green for water, blue for air, red for fire, and black for total darkness. Colors have been used to symbolize royalty. The royal color of the Sung Dynasty

(China) was brown, of the Ming Dynasty was green, and of the Ch'ing Dynasty was yellow. Different colors signify different castes in India. The Hindu bride wears yellow at her wedding, whereas the Chinese bride wears, and is surrounded by, red.

2.1.2.2 Color Meanings

Physiological and emotional responses to colors have been noted and studied by psychologists. For instance, Felix Deutch²¹ found that color brought about reflex actions in the involuntary actions such as pulse rate, breathing, and blood pressure. He found, however, that red may calm one person and excite another. There seems to be no general rule to attribute a response to a given color stimuli. Deutch reasoned that physiological and emotional changes were brought about by subjects' associations, so based on individual experiences, people reacted to color stimuli accordingly.

Color preferences have been studied from a number of different viewpoints. In a survey of 21,060 subjects, H.J. Eysenck reported a mean ranking of preference of basic colors in the following order. Blue, red, green, violet, orange, yellow. However, this ranking was for color in the abstract, without specific context.

Lars Sivik²² has done a number of studies regarding the meaning of color. He used semantic differential scaling, which is a scaling technique that uses a seven step bipolar scale of associations to ascertain subjects' feelings about stimuli. An example of an SD scale is

Warm OOOOOOO Cold

where the subject would fill in the most appropriate circle related to their response. Sivik and coworkers have used the same scales of meaning for many such studies^{23,24,25}. Their list of semantic bipolar variables is: *like-dislike, winter-summer, loud-discreet, unappetizing-appetizing, joyful-serious, hygienic-unhygienic, old fashioned-modern, wet-dry, beautiful-ugly, soothing-exciting, positive-negative, sick-healthy, cultured-uncultured, feminine-masculine, complicated-simple, stimulating-dull, cold-hot, lazy-energetic, friendly-hostile, active-passive, shameless-prudish, old-young, expensive-cheap, tense-relaxed, near-far, secure-anxious*. From a factor analysis (principle component analysis and orthogonalizing) of the resultant data, five major factors were found. These were denoted Excitement, Evaluation, Potency, Temperature, Activity. Since the experiment was performed to be “context free”, the connotations from subjects was general and possibly more arbitrary than would be the case if the stage was set in a more narrow sense. Given this, the inter-observer variance showed to be fairly low. Further studies have been performed that set the context. In “Color connotations of exterior colors,”²⁶ the above experiment was repeated with the difference being that subjects were shown pictures of colors on buildings instead of abstract patches. The list of variables were, for expediency, reduced in number to thirteen: *warm-cold, vulgar-cultured, friendly-hostile, hilarious-serious, pleasant-unpleasant, uncommon-common, soothing-exciting, masculine-feminine, open-enclosed, unclearly-clearly demarcated, strong-weak closure, spacious-cramped*. Two types of buildings were shown, and 67 colors were used that fairly evenly sampled the color space. 168 observers took part in the study. Analysis on the data showed three major factors; emotional (or pleasantness) evaluation, social evaluation, and a spatial factor. The spatial factor is somewhat obvious, as there are four variables that are very closely related to spatial aspects. In addition to factor

analysis, Sivik employed “isosemantic mapping” of the meanings in the NCS perceptual color space. This technique is interesting as it allows visualization of the meanings in planes of the space, although the details as to the mechanics of creating the mappings are somewhat vague. Whereas temperature was separated as a factor (*warm-cold* and *lazy-energetic*) in the context free color study, it was found to be included in the factor called emotional evaluation. For color in the building context the concepts of beauty and warmth are highly correlated, but for color in the absence of context they are independent. This was discussed to be partly a cultural bias, because Sivik states that there are very few “cool” colored buildings in Sweden, although this is not the case in many other parts of the world. The issue that meanings of words are context dependent was also discussed. If the meaning of a semantic variable changes as a function of context, then there is nothing limiting the SD pairs to being opposite in meaning, and the validity of the variable is questionable. This illustrates the need for well chosen variables that are unambiguously understood as opposite within the context of study. For the comparison of individual variables of meaning between color in and out of context, *warm-cold*, *bilious-serious*, *soothing-exciting*, *masculine-feminine* were all found to have fairly high positive correlation. *Vulgar-cultured*, *beautiful-ugly*, and *friendly-hostile* showed fairly low correlation. The factors that were named evaluative in both studies were found to have very low correlation. Thus, this factor was interpreted to have the highest sensitivity to context.

Kobayashi²⁷ has created a “Color Image Scale” that assigns semantic descriptors to perceptual dimensions of color. He uses warm-cool for hue, soft-hard for value, and clear-grayish for chroma. These attributes were derived from a factor analysis. The color space was also

segmented by hue, and by tone, which is a name corresponding to the ISCC-NBS color dictionary tone descriptors of a given hue. Kobayashi states that the three adjectival bipolar scales of warm-cool, soft-hard, and clear-grayish are relatively insensitive to differences of context, personal taste, and other environmental factors, and therefore can be used in a general setting. He goes on to map the space and assign different meanings to different parts of space. This seems to be a gross overgeneralization, and although the adjectives that re-describe the color space may be fairly invariant, there is no reason to believe that this is an exhaustive use of the dimensions of color meaning.

2.1.3 Preferred Reproduction of Pictorial Images

Colorimetry enables color matches to be predicted when original and reproduction are measured under identical viewing conditions. Color appearance models attempt to enable color matches to be predicted under varying viewing conditions. Even if color appearance models could exactly predict matches across differing viewing conditions and media types, they may or may not specify a preferred reproduction. Additionally, if viewing conditions are constrained to be identical, and the original image is free from defect, a colorimetric reproduction may not necessarily be preferred. Evidence suggests that, for several types of image reproduction technologies and modalities, a colorimetric reproduction is not preferred.

In 1948 Buck and Froelich²⁸ studied human complexion under a number of standard light sources. They determined the average complexion of the three major races (Caucasian, Asiatic, Negroid). Since human complexion is a strong memory color and is used by people to

assess the quality of color reproduction, it is a very important volume of color space. They found that "...the preferred source emphasizes the material in brightness and saturation without marked change in hue. Thus, in each case, the preference was for the lamp whose spectral distribution and color most closely matched the material being viewed." This led to the design of the soft white fluorescent lamp, which has a spectral response that is close to average Caucasian complexion. This may explain why a warmer color balance is often preferred when viewing portraits of people. In 1960 Bartleson²⁹ studied the memory colors of familiar objects using color patches. He found that the memory color of flesh was closest to the measured colors than for other familiar objects such as grass, sky, and brick. It also had the least interobserver variance. His overall results indicated that memory colors were generally more saturated than the real object and that the hue shifts of the memory colors were toward the most impressive chromatic attribute of the color in question. This result was also found by Newhall et al.³⁰ in an experiment on a comparison between color matching methods. He found that memory matching, that is successive matching, yielded higher variability, shorter matching times, higher remembered purities and somewhat higher remembered luminances than the test stimulus.

On preferred reproduction of colors there is a body of evidence that has been reconfirmed by different people over the years. These data stem primarily from photographic reproduction experiments by both Bartleson^{31,32} and by Hunt³³. Primarily, they were interested in the most prevalent colors that existed in natural scenes, specifically grass, sky and skin color.

Bartleson³¹ showed that the preferred color of skin in a reproduction was very close to the memory color of skin, and that it departs significantly from the true color of skin when measured. The mean measured skin color from his subjects was Munsell 10R with a chroma of around 4, while both the preferred and memory colors of skin were about 7YR with a chroma of 3.5. Bartleson³² later showed that the relationship between memory and preferred color did not generalize to grass and sky colors. While preferred skin hue was different than that of the measured values, preferred grass and sky colors were of approximately the same hue. Preferred sky was higher in chroma, while preferred grass was of lower chroma. Bartleson observed that while the preferred grass was of lower chroma, it may have been due to the fact that the available gamut of the process they used could not adequately produce saturated enough greens at the time.

Hunt³³ picked up on this and performed an experiment similar to Bartleson's in 1974 when the process capability was greater for making good greens. He found that the preferred color of skin tended to be yellower and of around the same chroma as did Bartleson. For sky, preferred and measured colors were of similar hue, but the preferred chroma tended to be higher. For grass, preferred color had similar chroma as mean measured, but tended to be yellower. Thus Hunt showed that Bartleson's results for green grass were probably due to process limitations rather than true preference.

Recently, Asada, et al.³⁴ have investigated preferred skin color and compared it to skin color from offset publications. They found that skin colors in publications were very similar to preferred skin color from their own experiments. Plotted in u'v' space, skin colors fall in ellipses that have small radii in hue and larger radii in chroma. As expected, Negro skin color is darker than Asiatic, which is darker than Caucasian, while the hue angle of all skin tones were quite close. Preferred skin color as a function of relative size of faces and color background were studied. There was little difference in the results from differing sizes in the preferred color, but preferred face chroma increased as a function of lightening background color.

In amateur color photography, and more recently in color copying it has been found that reproductions that are “punchier,” i.e. more colorful and possibly lighter and higher in contrast, tend to be preferred over reproductions that attempt to faithfully render color. An attempt to quantify this was done by Fedorovskaya.³⁵ It was found that for four natural scenes as viewed on a CRT monitor, changing chroma, both by scaling and by translation (in CIELUV space) affected preferred image quality. Images where the average chroma was increased by 5 to 10 units were found to be the most preferred. Gamut mapping was done by clipping in chroma. They found a very close relationship between naturalness and quality as well, which is hardly surprising.

Tone reproduction in color printing has been given relatively less importance in the literature than “colorimetric” reproduction. This may stem from the fact that most of the tone research was done within the photographic field in a time when control over processes was determined

solely by exposure and development and chemistry. When the mode of reproduction expanded to allow more arbitrary manipulation as in the case of digital reproduction and the specification of color changed from densities and absorbance to CIE based metrics, it may have been easy to reject or ignore much of the work previously done. Consequently, some workers strive to make colorimetric matches to source data without taking into consideration the source and the output luminance ranges. As Jones³⁶ pointed out in 1944, a proportional reduction of luminance contrasts to fit the dynamic range of the output system creates a far from preferred reproduction. When doing cross platform “device independent color” typically it is suggested that the colorimetry be relative. That is, a CIE Y of 1.0 should correspond to brightest white and a CIE Y of 0.0 should correspond to the darkest black attainable. This, in effect, is exactly what Jones warned was inappropriate; a linear scaling of the available luminance range. Unfortunately, without absolute data about the source luminance range, it is impossible to know exactly how to tune the tone reproduction. However, some assumptions may be made to guess what a typical source luminance range might be for an average outdoor scene and an average indoor scene. Jones states that the average brightness scale, which is the ratio of maximum to minimum object luminance, is 160, or 2.2 log units. If this is used as the luminance range for digitized pictorial images, then we can estimate the amount of luminance compression for a given output process. For instance, the output luminance range of a typical color laser printer is close to 60, or about 1.78 log units. This suggests that a tone curve that is more sigmoidal might give a more preferred reproduction than a linear scaling. Figure 3 shows the curve from MacAdam.³⁷

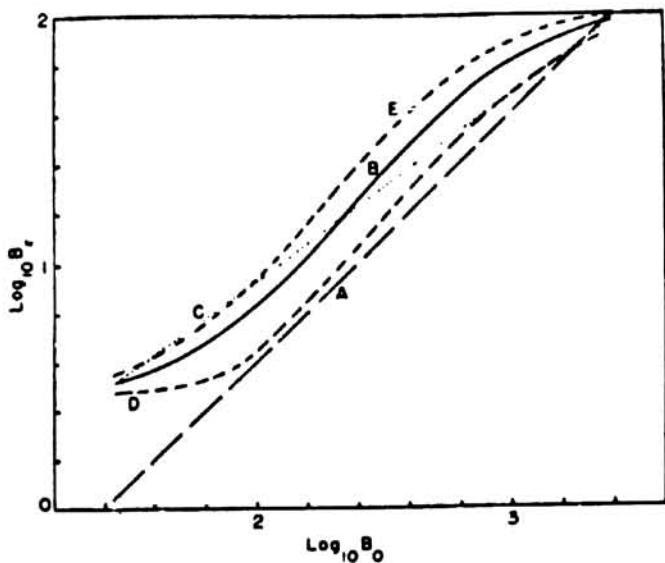


Fig. 9. Tone reproduction curves:

- A, "exact reproduction"
- B, optimum attainable with semimatte paper
- C, proportional reduction of luminance contrasts to fit density range of semi-matte paper
- D, tone reproduction suitable for scene in which highlights are predominant
- E, tone reproduction suitable when only shadow details are important.

Figure 3. MacAdam's tone compression curves for tone reproduction with limited dynamic range.

Roufs³⁸ approached this problem in a somewhat circuitous manner in a study on perceptual image quality. He used gamma functions instead of sigmoidal tone curves. Perceptual quality as a function of gamma was studied. It was found that for different images, the gamma corresponding to the preferred print was found to be different. He then scaled subjective brightness contrast as a function of gamma and then related perceptual quality to subjective contrast. He found that perceptual quality was more closely a function of subjective contrast and so then, less dependent on image content. I think that subjective contrast of the image was a function of where the detail was in the tone scale, as suggested by Jones. The images with the most detail in the shadow regions would find perceptually best reproduction using a gamma function that gave the most dynamic range to that part of the tone scale and where the

slope of the tone scale was similar to that of the luminance gradient of the original scene. As part of the same study, perceived sharpness was scaled as a function of both cutoff frequency and gamma. It was found that within a reasonable range of gamma functions, there is a negligible effect on perceived sharpness. However, when blurred with a gaussian filter, the perceived sharpness is reduced. The study did not include sharpening the images using digital methods such as laplacian filtering, unsharp masking, or error diffusion. Both tone reproduction and sharpness have a significant impact on the preferred reproduction of natural images, and as such are worthy of more consideration.

2.2 Preliminary Experiments

The preliminary experiments were the intended set of experiments from the thesis proposal. There were to be four experiments centered around the meaning of color in the context of business graphics, and the relationship of color meaning to gamut mapping for preferred reproduction. The following text describes the approach and the justification for such an approach.

The problem of gamut mapping has traditionally been approached “from above.” That is, various mapping techniques have been hypothesized and implemented, images were then pumped through the mappings, and finally some type of subjective judgment was made on the various hypotheses to determine which approach worked best under the conditions of the experiment.^{39,40} This approach is then iterated upon until some satisfactory results are obtained. One primary reason that gamut mapping work has been done “from above” is that the images that have been treated have almost exclusively been scanned images. The task of explicitly mapping each pixel of a scanned image whose color is out of gamut is a prohibitively huge task for almost any observer (unless the image is very carefully chosen or manipulated to be to the contrary) and would not lend itself to being able to secure a statistically significant number of subjects. Further, even if the number of colors to be explicitly mapped is small enough to be feasible, the task of color matching individual pixels to closest in gamut colors is likely to be image dependent and may not yield generalized results.

A fundamental drawback of an approach “from above” is that the search space of the problem is essentially infinite for all practical purposes: It is the set of three dimensional mappings that can be made from a source gamut to a destination gamut. In order to bound the search space, “reasonable” assumptions are made for a mapping approach using heuristics based on researchers’ experience. These assumptions may not always be valid for all uses of color reproduction. Furthermore, preference for a given mapping scheme is likely to be highly image dependent, both on color content and on the context to which the image belongs. Color content has been addressed by adaptively mapping colors based on the number of out of gamut points found in a given region of the color space of an image.⁴¹ However, image context has not been taken advantage of as a means to make gamut mapping decisions. This stems partially from the fact that most of the work in gamut mapping has been done on scanned images (typically images from the real world) and it is quite difficult to accurately segment scanned images into recognized objects.

Such may not be the case with computer synthesized graphical images that are commonly used in business and the graphic arts. Typical graphical objects can fall into one of a small number of general categories. One possible taxonomy of computer synthesized graphical objects is given below. It first makes the distinction between foreground and background (also called object and ground), then further categorizes page elements. Background objects can take the form of textures, patterns, or fills. The fills are typically gradient fills and progress from a fairly

light color to white across the page for paper applications (under bright lighting conditions). For applications where the medium is viewed under dark surround (overhead projectors or film projectors) the background is typically quite dark, and the foreground objects are lighter. Other background images are either repeating bitmaps, as in realistic texture representation (marble, wood), or repeating patterns. Backgrounds can be more complex, for example comprised of a scanned image. However this is less common as it may be confusing, thus, detracting from the foreground information.

Foreground images fall into one of at least six categories. The categories heretofore identified are:

1. *Literally representative images*, i.e. images created specifically in reference to the image of a naturally occurring object such as a face, hands, a cow, a computer, a tiger, etc. Images from this category are typified by the “clip art” found in many software packages. Images of this category “look like” their real life counterparts. This is an extremely broad category and spans a similar amount of content as scanned images from the real world. As such, results found from a given image cannot be generalized to the entire category.
2. *Figuratively representative images*, i.e. those images that unambiguously represent either something out of the natural world or a concept (such as “STOP”), but don’t “look like” the image of that thing. One example of this is the stick figures that denote men’s and

ladies' restrooms. This category is quite broad as well, and further categorization would need to be done if the results, as in case 1, are to be generalized.

3. *Externally imposed associations to familiar graphical objects.* These are the images that don't necessarily represent objects from the natural world, yet are so prevalent that they are identified and remembered. An example would be a corporate logo, such as Xerox's digital red X. Another example is a particular corporate color that is in standard use for internal documents.
4. *Data representative images*, or the graphical communication of numbers. Charts and graphs are an example of this type of image.
5. *Simple geometrical objects.* The colors that are attributes of images in this category are the most abstract in terms of concrete association, and are probably more prone to symbolism from personal experience.
6. *Text Objects.* Text is the most prevalent of any object type in most computer communication. Highlighted text directs attention to important parts of document. Colored text may have significantly different connotations than other objects similarly colored.

Taxonomy of computer synthesized graphical imagery

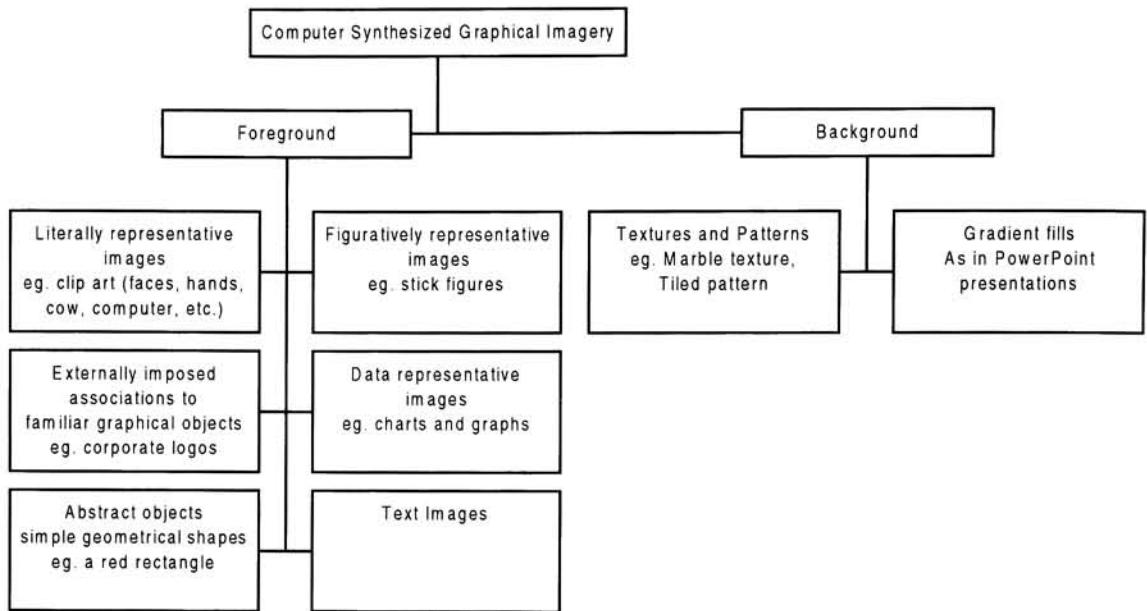


Figure 4. Taxonomy of business graphics image content

For each of the above mentioned categories of images and for images within each category, symbolism and associations of colors attached to objects may be different. For instance, for images that are supposed to look like objects from the natural world (Literally representative images) and for images that are familiar and frequently seen (Externally imposed associations), colors are fairly well defined and color meaning per se may less important than color fidelity (color fidelity in the sense of the color looking like it is “supposed” to). Take, for example, the image of a woman’s face. If the skin is colored blue or green or even red, it looks quite bizarre.

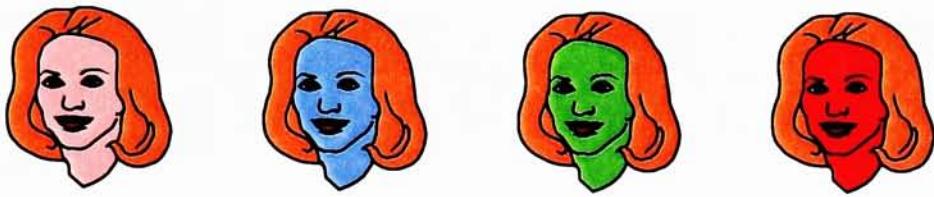


Figure 5. Different colors for the face. Face colors that are unnatural look “wrong”

A study of meaning of the color of these faces that are so clearly unnatural has dubious value. Of course, if the image is of a car then the coloration can take on a much wider range without eliciting such a visceral reaction.



Figure 6. Different colors for cars. There is no “natural” color for a car.

Figuratively representative, data representative, and simple abstract objects have less obvious constraints on their color attributes. Since figuratively representative images clearly stand for some thing or some idea, but don't look enough like what they represent to be literal, the latitude of “acceptable” colors is greater. Take, for example, the following images of a teacher lecturing to two students. The fact that the teacher figure is different colors may have subtle differences in the perceived meaning of the image, but none of the images is any more absurd than another.



Figure 7. Different colors for figuratively representative images.

The subject matter of the figuratively representative object may have significant influence on the associations of the colors contained in them. For example, the same color may have a different connotation when it is associated with a stick figure than when it is associated with a figurative representation of an automobile.



Figure 8. Figuratively representative images of automobiles with different colors.

For this reason, one image selected from the category of figurative representative images will not be representative of the entire group. Because of the need for further category segmentation based on image content, the results from an experiment with a small number of images will not be amenable to generalization. However, it is interesting to investigate the difference in response to the same concept (e.g. car) when presented both as literal and as figurative objects.

Data representative objects can take many forms. However, all charts and graphs exist to graphically communicate numbers and, as such, results from a small sampling of specific images may be able to be generalized to the entire category.

Both text and abstract objects may also be generalized to their category more easily than the first three object types.

The success of the proposed experiments rested on the validity of three hypotheses.

1. The first hypothesis was that a gamut mapping that is created from explicit matching (minimizing perceptual distance) of individual colors will create a useful mapping transform. As will be shown in experiment P2, it was possible to create useful mapping transforms using the matching method. The color matches could be modeled with a number of simple models that are developed in P2. They remain to be verified with psychophysical experiments, but the author is confident that with the proper set of parameters, they will perform as well as most other clipping type transformations.
2. The change in the meaning of colors across contexts (same color in different image types), is statistically significant and meaningful. To test this hypothesis, experiment P1 was conducted. This experiment showed that while there are some significant differences in specific descriptive terms, and the factors found (dimensions of meaning) were very similar to those found in other studies, it was difficult to glean from that useful instances of task descriptions.

3. Gamut mapping color matching decisions will be different when either (or both) the type of image presented to the observer changes or the task description changes. Pilot experiments were performed to verify these hypotheses. I concluded from these pilot experiments (both comprise P2) that while the technique can be used to create interesting models for gamut mapping, **there was not a significant difference between color matching means across images tested**. Because of the failings of the P1 pilot study, the differences in task description between the two P2 experiments was the difference between making a closest perceptual match and making a match where the meaning of the image was best preserved (where preservation of meaning was left up to the observer to define). On comparing the two experiments where the only difference was task description (preserving meaning vs. perceptually closest), the comparison was done through looking at differences in the parameter values that went into the gamut mapping model. No statistical study was done of differences, but all model parameters were very close in value.

Although hypotheses 1 and some parts of 2 were shown to be valid, the invalidating of hypothesis 3 forced a change in direction. From the gamut mapping experiments (P2), it was clear that hue uniformity was a significant problem. From these data, and the lack of a significant body of other data in the field, it was decided that hue uniformity be further studied.

2.2.1 Experiment P1 (not completed). Semantic differential scaling of colors in the context of business graphics

An experiment was designed to test if dimensions of meaning could be derived from a semantic differential scaling of colors. Osgood⁴² did pioneering work in the area of meaning, and the derivation of multidimensional spaces to represent it. He assumed that a multidimensional space could be made to represent orthogonal dimensions of meaning, and that one could derive those dimensions from a factor analysis applied to semantic differential scaling of the concept in question. A semantic differential scale is one where polar adjectives (descriptive terms of opposite meaning) sit between a 7 step scale. The task is to judge the concept against where it falls on the scale between the polar adjectives. Osgood repeatedly found that three major dimensions of meaning accounted for a large portion of the variance of the data (between 60% and 70%), irrespective of the subject being measured. These dimensions he called Evaluation (e.g. good-bad), which was almost always accounts for the largest variance (about twice as much as the next 2), Potency (e.g. hard-soft), and Activity (e.g. active-passive). Potency and activity on average accounted for about the same amount of variance. Sivik and others applied Osgood's techniques to the meaning of color (as described earlier on page 20-22), and found additional major factors, the most notable of them being warmth (e.g. warm-cold colors).

The objective of the P1 experiment was to validate that significant dimensions of meaning could be derived from semantic differential scaling as it applied to the meaning of colors in the context of business graphics.

The abstract and data sub-contexts were chosen to be evaluated as a first attempt. (The figurative, literal, and text representative images were going to be added following verification that meaningful factors could be extracted. The experiment was halted before this was achieved.) Fifteen semantic differential scales were used which were copies of the ones Sivik used in reference 22. An example script is shown (reduced in size) in figure 9. For the data representative image, the script read: "Don't judge the color as such, judge it as the coloration of the line chart." There are 15 scales in figure 9. The images used for abstract and data representative images types are shown in figure 10. Each of the image types was printed using 10 colors: Sky Blue (printer Cyan), Red, Purple, Lime Green, Gray, Blue, Violet, Yellow, Green, and Orange. For each image type (abstract and data), for each color (10 of them), 10 people performed the task by filling out the scaling page shown in figure 9. Although this sampling is not sufficient to be significant, it seemed appropriate at the time as a quick test of the method to see how and if useful information could be drawn from the analysis techniques.

Judge each color with respect to the scales shown below.

Interesting Boring

Old Young

Easy Difficult

Dangerous Safe

Warm Cold

Like Dislike

Serious Hilarious

Beautiful Ugly

Soothing Exciting

Positive Negative

Complicated Simple

Real Fake

Passive Active

Expensive Cheap

Success Failure

Figure 9. Reduced version of task and semantic differential scales for experiment P1 (abstract sub-class)

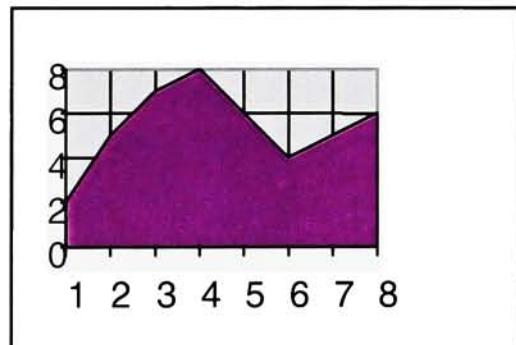
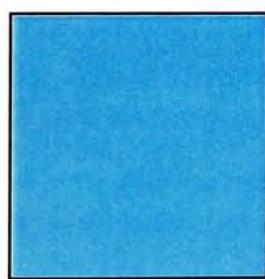


Figure 10. Example images from P1 experiment. Abstract and Data representative images.

Means were calculated for each combination of color and semantic differential scale. Because there are 10 colors and 15 scales per image, there are 150 means that need to be analyzed. Each mean is comprised of 10 samples. For this paltry subset of the experiment, 3000 observations had to take place. Figures showing the mean and standard deviation for each color and scale are quite large, so are shown in appendix A: Mean data plots from experiment P1. As can be seen from the plots, the data are quite noisy as error bars indicate standard deviation of the mean. The only way that useful results would be gleaned from this experiment was if the data indicated that for different image types, there was a systematic shift in meaning that was uniform across all colors. As can be seen from the plots, there are significant differences in the mean values of some scales for some descriptive terms, but they are not uniform across all colors. For instance, the mean value for the descriptive term “interesting”, for the color orange is about -0.4 for the abstract image type whereas it is about -1.6 for the data image type. The significance is judged by the mean of one scale being beyond the error bar (1 standard deviation) of the mean from the other image type¹. Though there are some significant differences between means of colors and descriptive terms across sub-contexts (sub-contexts is referred to as an image belonging to a specific type, e.g. abstract, or data), the trends are not systematic or uniform enough for the results to be used for subsequent experimentation.

¹ This is not strictly correct, as the standard error of the mean should have been used to judge significance.. However, the standard error of the mean would result in a smaller error bar since it is calculated from $1.96 * \text{standard deviation} / \sqrt{\text{number of observers}}$ (10 in this case). Since $1.96/\sqrt{10}$ is less than one, then the standard deviation is larger than the standard error.

To test whether the results from this experiment were similar to those done by Sivik, and by Osgood, a principal components analysis (PCA) was applied to the data set. PCA was done based on Varimax method using the SAS statistical software package. PCA was applied to the entire data set (pooled sub-contexts), and to the abstract and data sub-contexts individually. Because none of this work was continued, only the pooled results will be discussed.

The PCA found three factors that were significant (eigenvalues > 1) which accounted for about 63% of the variance of the data. Figure 11 shows the first 10 factors (of 15) and the amount of variance accounted for. Cumulative and individual variance is shown.

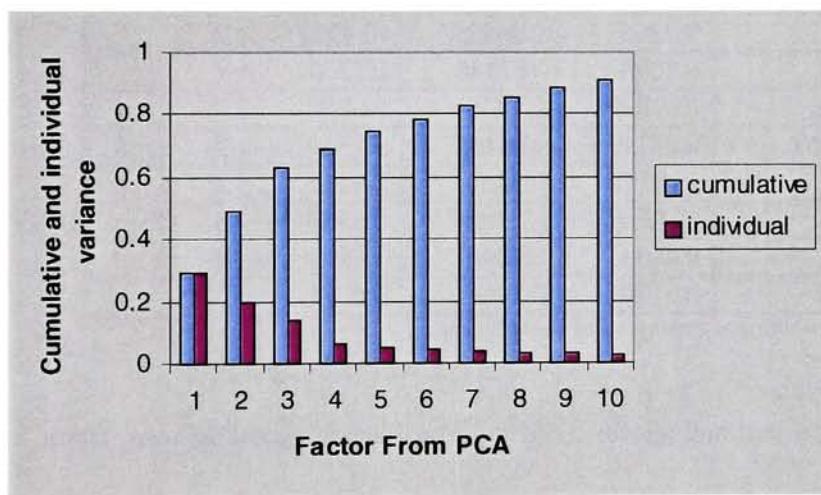


Figure 11. Variance accounted for by factors from PCA

These results correspond roughly to those found by both Osgood⁴³ and by Sivik⁴⁴ although, Sivik's results included 4 factors, and accounted for 85% of the total variance of the system. The three significant factors can be used to construct a three dimensional "semantic space" on

which the descriptive terms used in the experiment can be mapped as vectors in the space. Descriptive terms (bipolar scales) can be grouped by factor loadings, which is used to determine along which dimension a scale is most aligned. Table 1 show the scales arranged according to their factor loadings.

Bipolar scale variable	Factor 1	Factor 2	Factor 3	Primary Factor
Interesting-Boring	0.8256	-0.2699	-0.1856	1
Old-Young	-0.7017	0.2240	0.4499	1
Warm-Cold	0.4692	-0.0044	-0.0631	1
Like-Dislike	0.8254	0.2782	0.0735	1
Beautiful-Ugly	0.8319	0.2129	0.0942	1
Positive-Negative	0.6183	0.3831	-0.1306	1
Success-Failure	0.53665	0.3830	0.1237	1
Easy-Difficult	0.2263	0.5622	-0.4522	2
Dangerous-Safe	0.1098	-0.7946	0.1700	2
Soothing-Exciting	-0.2100	0.7836	0.0900	2
Complicated-Simple	0.2314	-0.6766	0.1889	2
Passive-Active	-0.5585	0.5860	0.1115	2
Serious-Hilarious	-0.1210	-0.0108	0.8380	3
Real-Fake	0.4184	0.2046	0.6589	3
Expensive-Cheap	0.5120	0.1061	0.6268	3.

Table 1. Bipolar scales arranged according to their factor loadings

To apply Osgood's terminology for the three major semantic dimensions, factor 1 would be named the Evaluation factor, factor 2 would be named the Activity factor, and factor 3 would be named the Potency factor. Quite a few of the loadings are not "clean," that is, they don't fall neatly into one factor, rather they project onto more than one dimension of the space. The scale Passive-Active, for instance, is nearly split half and half between the first and second

factors. This analysis does show, however, that the fifteen scales can be reduced to a small number of dimensions that account for a significant portion of the variance of the system.

Although somewhat interesting, the applicability of the experimental results to determining the best task description for gamut mapping was deemed unfeasible, so further analyses and experimentation were halted.

2.2.2 Experiment P2. Gamut mapping through the use of color matching (Gamut mapping from below)

This experiment was started concurrently with experiment P1, and was intended to test the truth of hypotheses 1 and 3. Restated, the objectives of this experiment were:

1. To create a gamut mapping model where the parameters can be fitted to psychophysical data derived from a color matching experiment.
2. To test if there is a significant difference in the mean color matches across sub-contexts (image types).
3. To test if there is a significant difference in the mean color matches across task description (from “find closest perceptual” to “preserve the meaning”).

This experiment was split into 2 sub-experiments. The first experiment concentrated on testing differences between sub-contexts, and the development of a general model to use to create vector directions that make a color mapping similar to the color matching mean matches made by observers. The second experiment was a copy of the first with the task description changed to determine if color matches differed significantly.

2.2.2.1 Perceptually closest color matches (sub-experiment 1)

2.2.2.1.1 Abstract: to P2 (sub-experiment 1)

A colorimetrically characterized computer-controlled CRT display was used to determine closest perceptual color matches of 25 colors when an exact match was not allowed. An artificial but realistic color gamut was created by intersecting the display gamut with a gamut of

a Xerox 4920 color laser printer. Each of 21 observers performed color matches between out of gamut colors and those on the artificial gamut's edge. Each observer made color matches on 4 different images. The images represented some of the categories that business graphic images can fall into. Between the different image types, there were no multidimensional (MANOVA) statistically significant differences at the 10% confidence level in any of the 25 colors tested. The mapping vectors showed that 1) observers don't make simple matches as assumed by most gamut mapping experiments done to date, 2) the influence of image content for simple graphical images tested does not have a large effect when the task is to make closest perceptual color matches, and 3) CIELAB hue angle is not uniform enough, especially in blue and cyan regions, to make adequate gamut mapping transforms. A simple model for clipping type gamut mapping is proposed. Results are compared to predictions of a new gamut mapping technique that minimizes weighted color difference between the target color and the gamut boundary.

2.2.2.1.2 Introduction to P2 (sub-experiment 1)

The gamut mapping problem is well known in color industries. There are also many definitions of gamut mapping. The definition assumed here is: Gamut mapping is a solution to the problem encountered when rendering a color image on a device that cannot accurately reproduce all the colors in the image, after accounting for viewing conditions. Many approaches have been examined, and some have proved more useful than others. Stone, et al.⁴⁵ introduced the concept of gamut mapping, but lumped viewing conditions transforms into the mapping, and used CIEXYZ space to alter color information. Later, Wallace and Stone,⁴⁶

Gentile, et al.,⁴⁷ and others separated the issues of gamut mismatch from adaptation transforms somewhat by performing mappings in more perceptually uniform color spaces, such as CIELAB color space. Wolski, Allebach and Bouman⁴⁸ have created transforms that use different types of mappings in different parts of color space (above and below L* of 50). Hoshino and Berns⁴⁹ provide a good summary of existing techniques for gamut mapping, all assuming that the mapping is done in a perceptually uniform color appearance type space. Montag and Fairchild⁵⁰ have recently completed a large systematic evaluation of existing gamut mapping techniques, including clipping transforms, and compression transforms along lines of constant lightness, constant chroma and constant saturation. Test images were simple spheres and the entire test was performed on a CRT screen.

Gamut mapping transforms have without exception (to the authors' knowledge) been defined functionally, then tested. Many functions have been proposed for mapping a point outside a color gamut onto or into the volume of reproducible colors. Mapping vectors typically follow one of five directions (with 4 of 5 maintaining metric hue angle to be constant): 1) Constant lightness, 2) constant chroma, 3) constant saturation or analog of saturation, 4) centroid, or towards a central point in the gamut (e.g. 50,0,0 in CIELAB), and , 5) minimum distance or minimum weighted distance to the gamut surface. Sometimes a combination of approaches is taken based on location in color space and in certain sequences. A new gamut mapping technique using minimum weighted color difference has been proposed by Katoh and Ito⁵¹, and will be discussed in more detail later. While linear and non-linear compression

techniques have been investigated, the direction of the compression either preserves lightness and hue, or preserves saturation (or it's analog in CIELAB) and hue.

While many of the above transforms work well under certain conditions, the search space for mapping a three dimensional point to another point in three space is essentially infinite. The functions for transforming color data are based on heuristics learned through experience, and from that point are not invalid. However, this does not guarantee that any transform tested is in any way optimal. What is not tested, cannot be found.

With this in mind, an experiment was designed to find the perceptually smallest distance from a point in color space to a closed surface of color (gamut boundary). The focus here was on gamut mapping of business-graphics images, although the results may be generally applied to all image types. Additionally the experiment tested whether image type affected observer's gamut mapping decisions. The task description was the same for all trials:

"Adjust the color of the image on the left hand side of the screen to make it look as close as possible to the image on the right hand of the screen."

Three different image types were presented: Abstract (a colored square), data (a colored chart), and figurative (a cartoon car). An additional image was used (a colored square with a border) to test the dependence of a black border on observer's matches. The experiment

allowed observers to pick colors from a two dimensional image of a simulated gamut edge to match colors that were presented in an image. Twenty five “out of gamut” colors were selected and mapped by 21 different observers. An exact color matching exercise was performed by each observer first to determine subject’s skill at color matching using the interface provided. Using the exact match results, four subjects (of 25) had errors that were too large and were not used in the subsequent experiments.

2.2.2.1.3 *Experimental*

There are two major and presumably separable problems today regarding the reproduction of colors from CRT to hardcopy print, or from any source color gamut/viewing condition to any destination color gamut/viewing condition. These are 1) accounting for differences in viewing condition, i.e. viewing mode (self luminous vs. hard copy), chromatic adaptation, luminance adaptation, etc. and 2) accounting for gamut mismatch. These experiments assume that problem 1) has been solved adequately and concentrate on investigating problem 2).

The following sections explain the details of the experimental design. The user interface tool was designed and implemented. The test images were selected. The viewing conditions were set up and the CRT was calibrated. The simulated gamut edge was created. The task description and selection of 25 colors was made. Twenty one adult observers, 9 females and 12 males, performed the exercise.

2.2.2.1.3.1 USER INTERFACE

Fig. 12 illustrates the user interface for the experiment. The target image was presented at the top right. The image that is manipulated to make a match was presented at the top left. The

gamut edge (set of available colors from which to choose) and a zoomed version of the gamut edge were presented across the bottom.

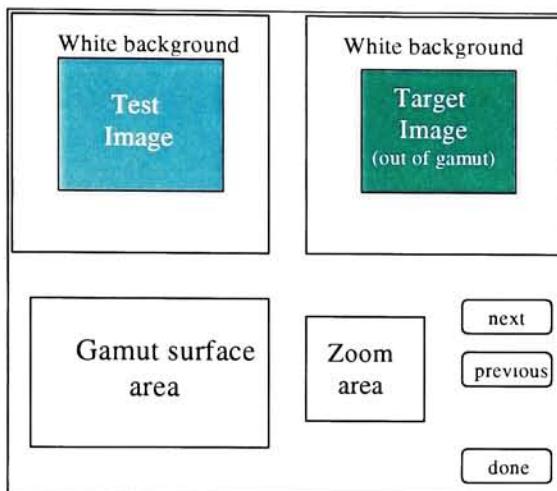


Figure 12. User interface for experiment P2 (sub-experiment 1)

To use the tool the observer picks a color in the “gamut” area. This changes the “zoom” area to make the center color in the “zoom” area the same as the color picked in the “gamut” area. When a point in the “zoom” area is clicked, the color of the parts of the test image that are alterable are changed to have the color selected. The colors can be updated very quickly and change as the user drags the pointer around in the “zoom” area. The “gamut” area is a two dimensional representation of a gamut edge in CIELAB $L^*C^*_{ab}h$ color space. The x axis represents hue and the y axis represents lightness. The color at the point x,y is the $h=x$, $L^*=y$, and C^*_{ab} = the chroma limit at the corresponding L^*,h coordinates. There are 36,000 distinct colors in the gamut area, enough to create the appearance of a continuous gradation of color around the gamut surface.

Any target image can be used in this test. The parts of the image that are updated by the user correspond to an arbitrary length list of polygons described by an initialization file.

Four test images were selected to represent three different image “contexts”. The images shown are simple computer generated graphical type images that represent the abstract, data, and figurative image context groups. Not studied here were representative images from the literal and text contexts due to limitations in the availability of software tools to generate the proper polygon sets to represent the alterable pixels. Fig. 13 shows examples of the three selected image “contexts” that were used in the experiment. Two “abstract” contexts were used, one with no border, and one with a black border.

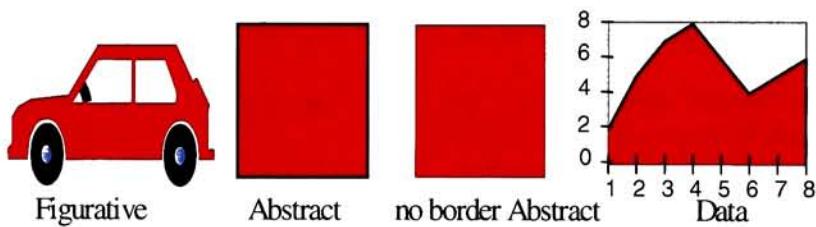


Figure 13. Images presented to the observers for experiment P2 (sub-experiment 1)

2.2.2.1.3.2 VIEWING CONDITIONS

The experiment was performed under simulated D65 lighting. The decision was made to test under bright surround conditions to more closely model real life situations where the results of

this work may be used. The CRT and computer were set up under controlled lighting with a background and surround of gray matte board. An attempt was made to match both the luminance and the chromaticity of the illuminant white and the CRT white point to nullify any chromatic adaptation issues. A mixture of fluorescent D50 and D75 lamps were used because the color rendering index of available simulated D65 source is quite low. A Minolta Chroma Meter CS-100 was used to make all the color measurements. A pressed PTFE puck (Halon) and paper measurements were taken at the same angle as the CRT screen. The xyY values of the surfaces are shown in the first table in appendix B. The ΔE^*_{ab} calculated under D65 for the difference between the measured white of left and right positions on the CRT was 3.5.

2.2.2.1.3.3 CRT CALIBRATION

The CRT calibration was done in two steps. First the white point digital counts were selected, then the system was characterized. A Sun Sparc 5, 18" CRT was used. There is no way to directly alter the gain of the monitor, so the digital count maxima must be altered to affect the white point. The chosen white point was R=255 G=240 B=220, with xyY value of (.305 .321 79.1). The chromaticity of paper white under the selected illumination was (.320 .333 80.3) which is between D65 and D55. The gray background of the matte board at (.307 .321 13.3) is quite close in chromaticity to the CRT white. Based on these measurements it was thought that there would be little confounding from differences in surround luminance or chromaticity.

Once the white point was calibrated, the digital count to luminance relationship was measured for each of the R G and B guns by measuring .2, .3, .4, .5, .6, .8, and 1.0 times the maximum

digital counts separately. A spline fitting routine was used to create smooth curves that mapped digital count to luminance of separate channels. Then XYZ measurements were taken of R, G, and B maximum values, and a 3x3 matrix was used to create the transformation to XYZ from RGB primaries. Flare was measured but not included in the CRT model. The accuracy of this calibration is an average ΔE^*_{ab} of less than 2.

Once the CRT was characterized, forward and reverse models were coded and a 3 dimensional LUT was made. Each of the test images was processed through this LUT to make the neutrals in the image the same chromaticity as the white of the CRT.

2.2.2.1.3.4 GAMUT SELECTION

The reverse CRT model was used to find the gamut of the CRT in CIELAB space. A 16x16x16 sampling of CRT device space was converted to 16x16x16 samples in CIELAB space. The edge of this device cube (16x16x16 – 14x14x14 values) was used to make a piecewise planar surface in CIELAB that represents the gamut edge. A two dimensional gamut boundary description was created by using gamut mapping software that performed chroma clipping with constant hue and lightness. Hue “leaves” were first created by sampling CIELAB $L^*C^*_{ab}h$ space at 100 lightness levels and 360 hue partitions with a chroma that is higher than any device chroma everywhere (128 was used). The gamut mapping software was then used to clip the chroma of each lightness level of each hue leaf. The resulting data can be represented as a 2D image of $L^*C^*_{ab}h$ values, or as a sequence of hue leaves around the hue circle.

A second gamut was created from CIELAB values obtained from sampling output of a color laser printer (Xerox 4920). A combination of the CRT and printer gamut was used to create the simulated destination gamut edge for the experiment. A gamut boundary was created by taking the minimum chroma of the two gamut boundaries at each L^*, h point. This creates an “intersection” gamut. A preliminary gamut edge chosen was an intersection gamut with chroma scaled by 90%. The resulting gamut edge was quite under saturated, though and it was decided to change to an unscaled “intersection” gamut edge. It was felt also, that the intersection gamut is the most reasonable one for simulating real-life conditions.

The intersection gamut boundary was converted to an RGB device image (360x100 pixels) and scaled to fit in the “gamut” window space of the color matching tool. Fig. 14 shows an example of the CRT gamut and the printer gamut at a hue angle of 303° degrees.

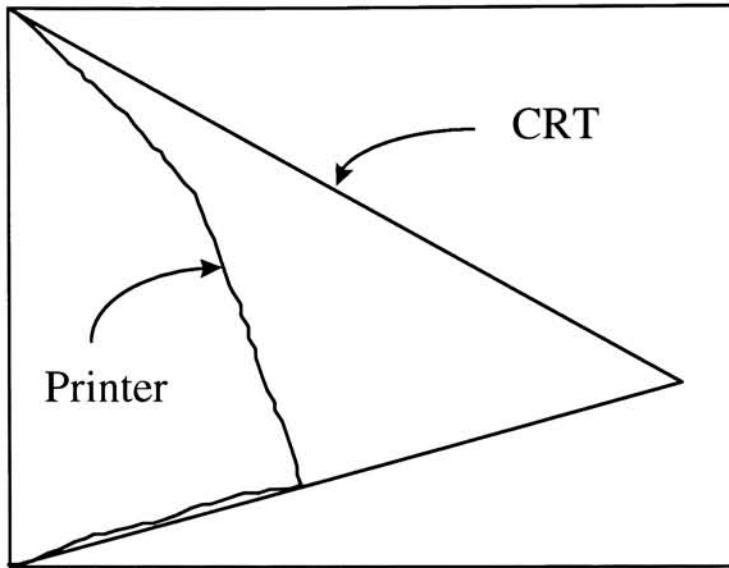


Figure 14. Example hue leaf of CRT and Printer gamut shapes at $h_{ab}=303^0$.

Because the experiment was performed entirely on a CRT display, the only type of gamut mapping that could be tested was where the CRT gamut is larger than the printer gamut. It is well known that many printers can create colors the CRT is incapable of making, especially in cyans and greens of mid lightness, and in bright yellows. Because of the limitations imposed by the display medium, gamut mapping where the printer has higher chroma was not testable.

2.2.2.1.3.5 COLOR SELECTION

Twenty five colors were selected from the surface of the CRT gamut as test colors to match with colors on the reduced “intersection” gamut boundary. They were chosen to span the color space fairly uniformly. Representatives from the 11 most basic color terms were chosen, as well as a concentration about blue since it is an interesting color center (the greatest hue non uniformity in CIELAB space is in the blue region). Out of gamut colors were chosen in parts of the color space where there was a large mismatch in volumes. This means that in

blues and magentas, the target colors were significantly darker than in the greens, cyans, yellows and reds. Because there is no large volume mismatch in light blues and magentas, nor in dark greens, cyans, yellows, or reds, the sampling is sparse in those regions.

The colors selected for the gamut mapping match test are shown in the second table in appendix B. Twenty five colors seem to be close to an upper limit for a matching session without a break. Since most observers performed two tests in a given sitting, 50 colors could have been chosen with a break at the 25th color. The mapping task is quite a bit more frustrating than the exact match exercise though, so it was important not to make the task too overwhelming to the observer.

2.2.2.1.3.6 EXACT MATCH DATA AND RESULTS

The exact match experiment was used as a vehicle to determine observers' individual abilities to 1) match colors and 2) use the interface. The average ΔE^*_{ab} over the entire data set was 4.25. The median ΔE^*_{ab} over the data set was 3.08. The standard deviation was 3.61. The histogram is shown in Fig. 15.

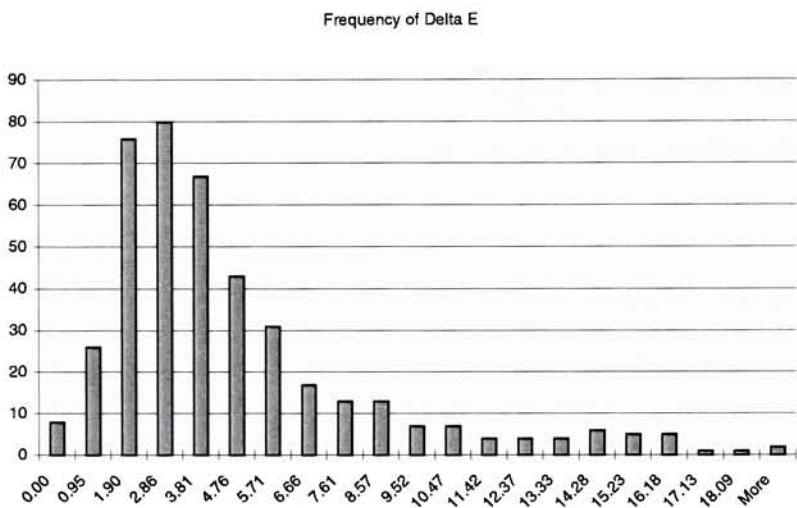


Figure 15. Histogram of ΔE^*_{ab} for exact match exercise for entire observer group including outliers.

The average error is on the same order as the color error between the left and right areas on the CRT screen that presented the target and test color patches. Much larger errors can be seen in low frequency. These may be attributed to carelessness, or to fatigue in observers. Four observers were found to have average errors over 7. These observers were not allowed to continue, as their average errors were too high. The remaining 21 observers had average errors of 7 or less. Most large errors were in very dark colors where influence from flare on the CRT (bright surround conditions) was confounding.

2.2.2.1.3.7 RESULTS AND DISCUSSION

Twenty one observers performed the gamut mapping color match task. Four sets of images were presented to each observer over four different sessions. The image sets were presented in the same order for each observer. Observers typically matched two sets of images in one

sitting. Including the exact match set, an average of three sittings was necessary to complete five sets of matching tasks. The four sets of images were named abstract, data, fig, and noborder. These sets corresponded to an image of a square (with a black border), an image of a data graph, an image of a cartoon car, and an image of a square with no black border. Pictures of these images are shown in Fig. 13. The list of colors was randomized once and used for all sets of images.

Each of the 21 observers performed one set of matches for each of the four image types. The data set contains, for each color, for each observer, for each image type, one color match vector. Each color match vector is composed of three independent numbers and must be treated separately. Given this data set, means can be calculated in a variety of ways. Interesting means include the mean mapping vector for each color and the mean vector for each color for each image type. Additionally, it is interesting to test whether the difference of the means of color vectors for each image type are statistically significant. If there is no statistical difference in the means, they can be pooled to create a better estimate of the overall mean of the matching decisions.

2.2.2.1.3.7.1 STATISTICAL ANALYSIS

For each of the 25 colors, a multivariate ANOVA was performed to test whether mean (three dimensional) mapping vectors from the 4 image types were significantly different from one another. The null hypothesis for this test was

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \text{ where}$$

μ_1 = mean of context 1 (abstract)

μ_2 = mean of context 2 (data)

μ_3 = mean of context 3 (figurative)

μ_4 = mean of context 4 (no border abstract)

H1: at least one mean is different from another.

There are 21 subjects X 4 contexts of data for each color for each of L*, a*, and b* vectors.

Since the mapping vectors lie in three dimensions, it is appropriate to use multivariate techniques to analyze the difference of the means (MANOVA). Three dimensional MANOVA was calculated for each of the 25 colors tested. The SYSTAT statistics software package was used to do all the analysis. Wilks' Lambda, Pillai trace and Hotelling-Lawley trace were all performed to calculate f-statistics. The critical F value for these tests was calculated based on 9/189, 9/240, 9/230 degrees of freedom respectively. The critical value for all tests is equivalent to 9/• degrees of freedom. The critical F value at the 10% level is 1.6315.

The null hypothesis was not rejected at the 10 percent level for any of the 25 colors tested for any of the above techniques, although univariate tests on some colors' independent dimensions did reject the null hypothesis. Based on these data, it was concluded that there is no significant difference between contexts for color mapping decisions. The data from all four sets were pooled and considered to come from one population.

Some comments of observers are relevant though, and deserve comment. Subjects noted that bright cyan and bright green, both device CRT maximum colors, were the most difficult and frustrating colors to match (indeed, it was impossible to make an adequate match). These colors also appeared to be self luminous, or “fluorescent”, and impossible to match with the palette of colors allowed to the user in the limited “gamut edge” set of colors. This is the part of color space where there is the most mismatch between printer and CRT. CRTs can make very light, brilliant greens and cyans, where printer the colors are highly chromatic at much darker colors.

2.2.2.1.3.7.2 MEAN MAPPING VECTORS

Fig. 16 shows the median match vectors in the a^*-b^* plane and as hue leaves in $L^*-C_{ab}^*$ planes. The gamut boundaries shown are for the hue angle that the point mapped to, not the hue leaf associated with the target color.

The general trend seems to be a “constant” (perceptual) hue mapping with lightness and chroma both being traded off. Mapping vectors for the most part point toward the center of the volume, although not exactly at $L^* = 50$, but quite close in many parts of the color space. In the yellow-green quadrant, the vectors all point toward greater than $L^*=50$, while in the yellow red quadrant, the mapping vectors all point toward less than $L^* = 50$. This may be due to the yellow-green hue leaves being convex where there are more high chroma colors near the target lightness. The yellow-red hue leaves are all concave, so more lightness must be sacrificed to get reasonable chroma. The centroid behavior of the mapping vectors is not surprising especially since the task description was to make a closest perceptual match.

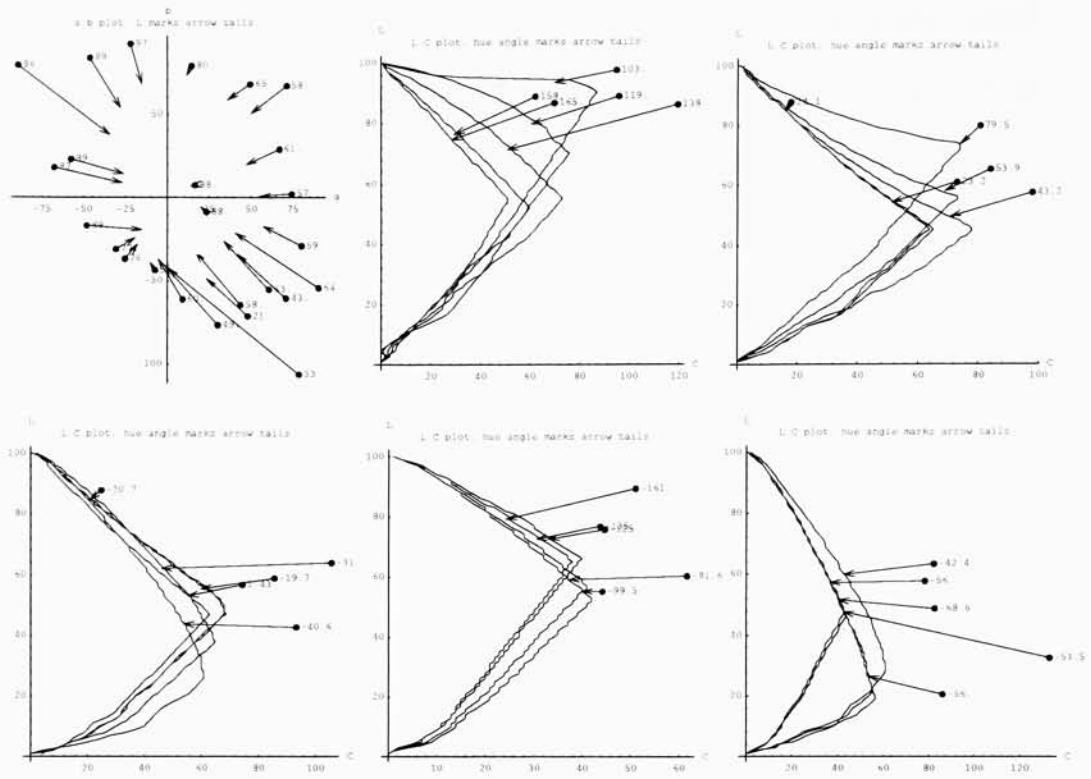


Figure 16. Median match values for 25 colors in a^*b^* plane, and in $L^*-C^*_{ab}$ planes.

If the task description was to make a preferred “match” or to make a match that preserves the intent or meaning of the color in the context of the specific image shown, this behavior may change significantly.

2.2.2.1.3.7.3 A MODEL OF MAPPING COLOR VECTORS TO THE GAMUT SURFACE:

Every color match vector can be thought of as a combination of component vectors that originate at the target color and end somewhere on the surface of the gamut. A reasonable set of three component vectors is:

- 1) constant L^* vector,

- 2) centroid vector (points toward 50,0,0),
- 3) minimum distance to gamut surface vector in the $L^*-C^*_{ab}$ plane (at fixed hue angle).

Fig. 17 shows an example of a mapping vector with its constituent component vectors. Note that, as is the case in the experimental data set, the match vector is a combination of at least two of the three component vectors and cannot be represented satisfactorily by any one component.

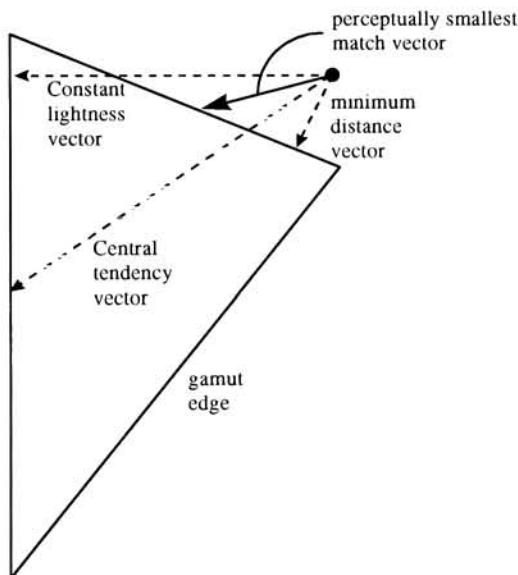


Figure 17. Example gamut mapping vector with component vectors.

The slope P of the line (in the $L^* C^*_{ab}$ plane) from the target color to the gamut boundary is a combination of weighted component vectors. This is shown in equation 1.

$$P = (Lw * \text{lightness.y} + Cw * \text{centroid.y} + Mw * \text{mindist.y}) / (Lw * \text{lightness.x} + Cw * \text{centroid.x} + Mw * \text{mindist.x})$$

Equation 1. Slope calculation for weighted component vector gamut mapping.

Lightness, centroid, mindist are two dimensional unit vectors calculated for each target color. The .x and .y suffix for each unit vector denote x and y axes. The x axis corresponds to chroma and the y axis corresponds to lightness. Lw, Cw, and Mw are scalar multipliers. The mapped color C is a point that sits on the gamut boundary and intersects a line with slope P that passes through the target color.

The three dimensional extension to the model is to find the minimum ΔE^*_{ab} between target color and mapped color P in a range of hue angles about the hue angle of the target color. The parameter Hw determines how many hue leaves will be searched for a minimum ΔE^*_{ab} between the target color and the color P. Hw is an integer that is the extent above and below the target hue the routine will search for a minimum.

A model was computed in both the two and the three dimensional cases. In order to find appropriate parameters, a local optimization was set up that minimized ΔE^*_{ab} between mapped model vectors and observer matches, as a function of the scalar parameters, Lw, Cw, Mw, and Hw (in the 3d case). The “amoeba” routine from Numerical Recipes in C⁵² was used to find minima. The routine implements an N dimensional simplex that crawls toward minima by using reflection, expansion and contraction. Table 2 shows the results for two and three dimensions. The errors calculated in 2 dimensions are only for L* C*_{ab} plane and don’t try to account for target hue angle being different from mapped hue angle. This case would be appropriate if the color mapping space was perfectly uniform in hue angle, and observer’s

choices preserved perceptual hue. Figs. 18 and 19 show the resultant vectors compared to the observer matches. Note that the regions of greatest error are in the blues and in the concave hue leaves of the red-yellow region. The hue leaves shown are for the hue angle of the observer match for the 2d case and for the model prediction in the 3d case.

Minimization	Lw	Cw	Mw	Hw	Max error	Average error
2d ($L^* C_{ab}^*$)	1	1.04	.25	NA	9.8	3.24
3d	1	1.03	.38	+3	33.9	6.84

Table 2. Model parameters for weighted vector gamut mapping model.

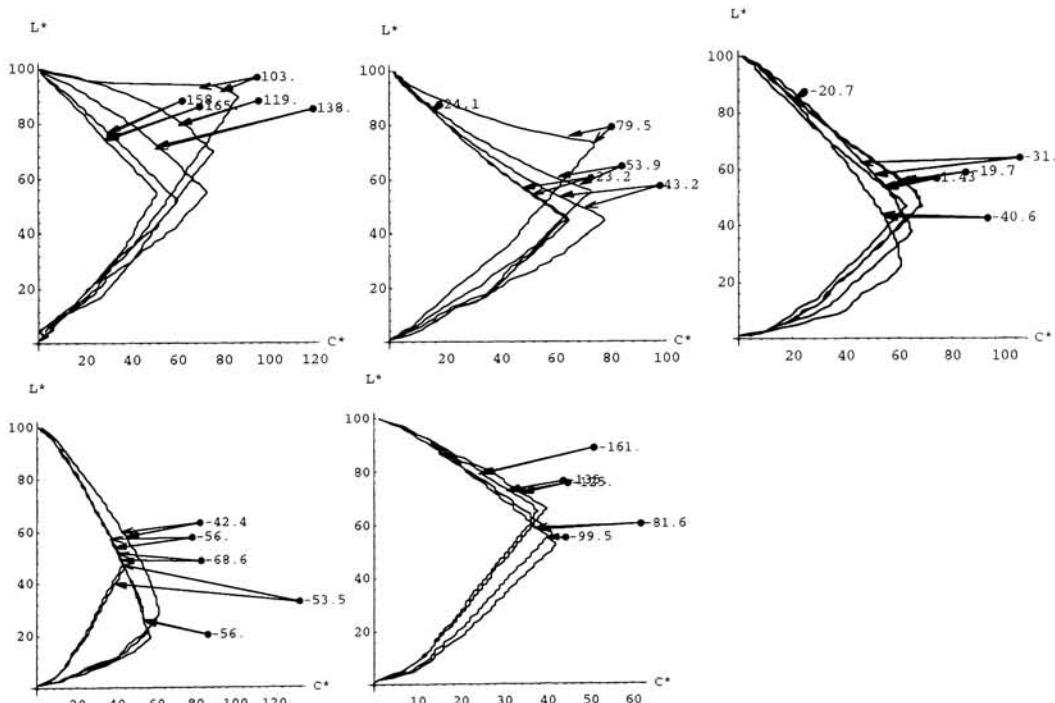


Figure 18. 2D weighted vector gamut mapping. Open arrow heads indicate observer results. Filled arrow heads indicate model predictions. Hue leaves shown are for the hue angle of the observer match.

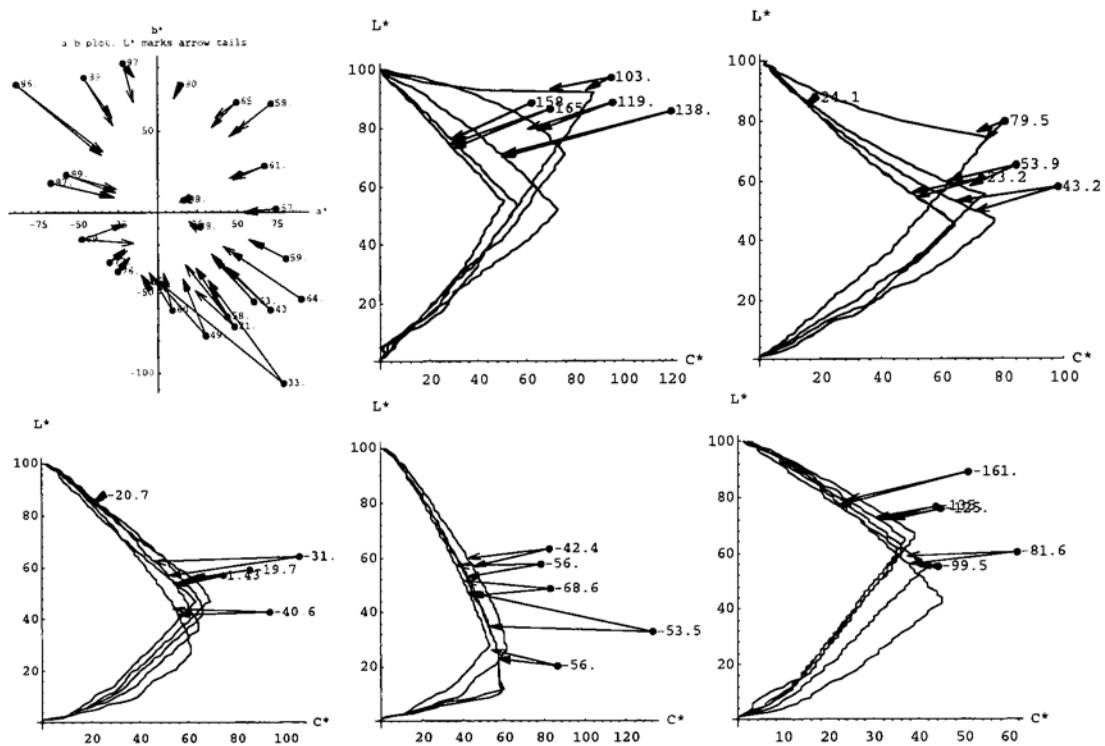


Figure 19. 3D weighted vector gamut mapping. Open arrow heads indicate observer results. Filled arrow heads indicate model predictions. Hue leaves shown are for the hue angle of the predicted match.

2.2.2.1.3.7.4 HUE MAPPINGS

Constant hue loci data from Hung and Berns⁵³ were used to test hue nonuniformity of mapping vectors. If the constant hue loci could be used to correct the hue nonuniformity of the mapping vectors, it may be concluded that observers are performing color matches that preserve the perceived hue of the target color.

The significant areas of hues not mapping toward $a^*=b^*=0$ are in blues and cyan. Using the constant-lightness data set from Hung and Berns, a lookup table was made that mapped

incoming hues that sat on constant hue loci, to the reference hues that were used to find the loci. In between these 12 hue loci, linear interpolation was used to fill the rest of the space. The constant lightness data set created a mapping that significantly corrected hue vectors toward 0 for the blue region, although the cyan region was unchanged and green was made slightly worse (yellower). Fig. 20 shows the original matches, along with the “hue corrected” matches in the a^*-b^* plane. Given that the data used to create this mapping consisted of 36 data points, this is not unimpressive, at least in the blue region of color space. Unfortunately, the cyan region was not helped. This is probably due to the fact that the Hung and Berns data are for constant lightness, whereas the cyan color mapping vector gets significantly darker. Based on observer feedback it seems the mappings are done with a strong emphasis on maintaining perceived hue.

The data here strongly suggest, at the very least in the blue and cyan regions, that CIELAB color space is not uniform enough to do constant metric hue gamut mapping. Additional data must be drawn from constant hue loci experiments to straighten the hue metric before it will be adequate for gamut mapping.

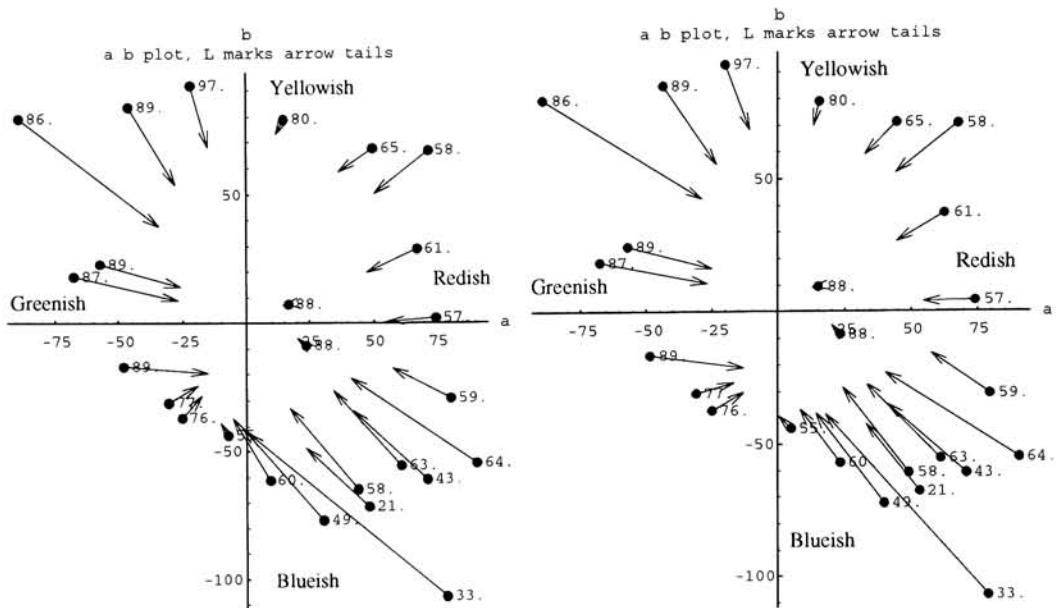


Figure 20. Hue angle matches on left, and "corrected" hue angle matches on right (using Hung's Constant Lightness data).

2.2.2.1.3.7.5 COMPARISON WITH KATOH AND ITO'S MODEL

Recently, Katoh and Ito⁹ have proposed and tested a new gamut mapping model. Since it is similar in some respects to the above model proposed by the author, it deserves inspection. The model defines a gamut mapping direction for each out of gamut color by finding the minimum weighted color difference in CIELAB space. The weighting of the color difference is similar to that of ΔE_{cmc} , where ΔL^* , ΔC^*_{ab} , and ΔH^*_{ab} are weighted by dividing by scalar values. The weighted color difference equation used is shown in equation 2.

$$\Delta E = \sqrt{\left(\frac{\Delta L^*}{Kl}\right)^2 + \left(\frac{\Delta C^*_{ab}}{Kc}\right)^2 + \left(\frac{\Delta H^*_{ab}}{Kh}\right)^2}$$

Equation 2. Weighted color difference equation.

Kl , Kc , and Kh can be altered to change the shape of an ellipsoid like surface of constant error. Mapping vectors are defined by the target color and the point at which the smallest ellipsoid intersects with the gamut volume (with the target color being at the center of the ellipsoid). In a discrete sense (and how it was done in this work), the vector is found by calculating the weighted color difference between the target (out of gamut) color and each point on the gamut boundary. 36000 data points were used to represent the gamut boundary, 100 lightness levels at 360 hue angles. The mapped point is that which has the smallest weighted error. Katoh and Ito tested their model with computer generated images and found that images mapped according to a ratio of $Kl:Kc:Kh$ of 1:2:2 respectively best matched the original images. This may be interpreted as lightness is more important to preserve than chroma or hue in the images used in the experiment.

Fig. 21 shows a comparison of Katoh and Ito's best parameter mappings against the observer results found in this experiment. Katoh and Ito's results are based on the gamut of the printer they used and is likely to be somewhat different from that used in this experiment. There are marked differences in hue in the blue and green regions. The largest differences in chroma also appear in the blue region. This is not surprising as the blue region is most non uniform in

hue. However, the $L^*-C^*_{ab}$ fit is quite close in a large part of the space. The total ΔE^*_{ab} error between the predictions and the observer matches is 318. The average error is 12.72.

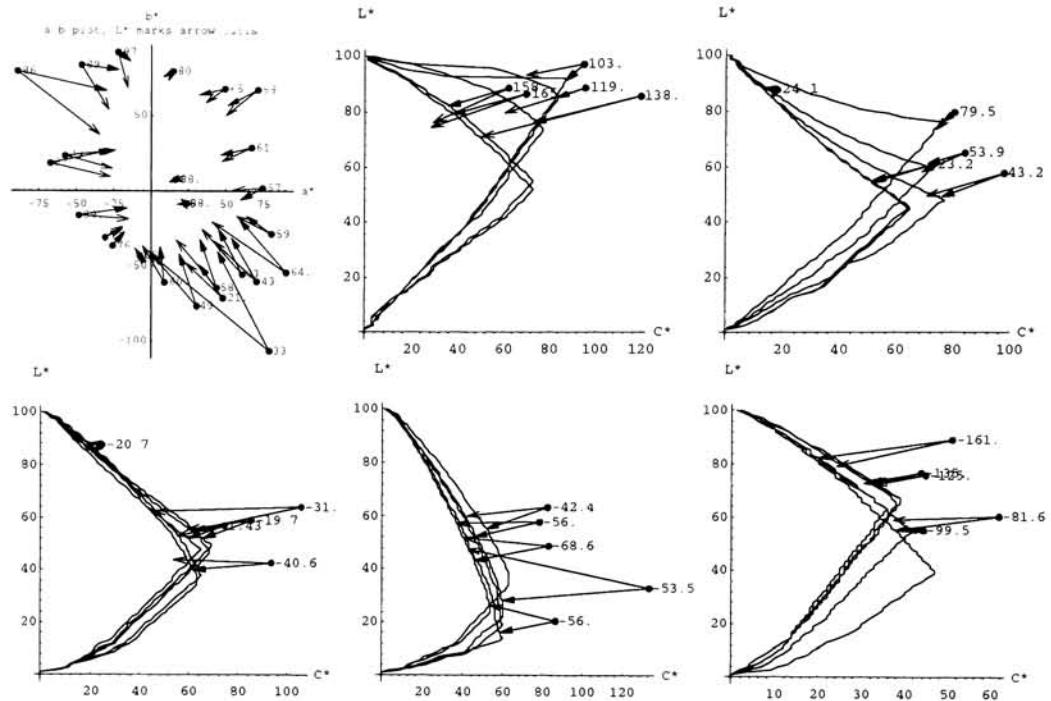


Figure 21. Comparison of Observer results with Katoh and Ito 1:2:2 vectors. Open arrow heads indicate observer results. Filled arrow heads indicate Katoh and Ito results. Hue leaves shown are for the hue angle of the predicted match.

In order to see if other parameters existed that fit the observer data better, the same optimization technique as described above was used. In addition, an enhancement to the model was made that gave better fit to the experimental data. This enhancement allows the ellipsoid to be rotated in the $L^* C^*_{ab}$ plane. The enhanced model has the form in equation 3.

$$A = \Delta C^* \cdot \sin\left(Kt \cdot \text{ArcTan}\left(\frac{L^* - 50}{C^*}\right)\right) + \Delta L^* \cdot \cos\left(Kt \cdot \text{ArcTan}\left(\frac{L^* - 50}{C^*}\right)\right)$$

$$B = \Delta C^* \cdot \cos\left(Kt \cdot \text{ArcTan}\left(\frac{L^* - 50}{C^*}\right)\right) - \Delta L^* \cdot \sin\left(Kt \cdot \text{ArcTan}\left(\frac{L^* - 50}{C^*}\right)\right)$$

$$\Delta E = \sqrt{\left(\frac{A}{Kl}\right)^2 + \left(\frac{B}{Kc}\right)^2 + \left(\frac{\Delta H^*_{ab}}{Kh}\right)^2}$$

Equation 3. Enhanced weighted color difference model.

The L^* and C^*_{ab} in equations A and B are elements of the target color. Kt is the parameter that determines the degree of rotation of the ellipsoid.

Optimized model parameters were found for both Katoh and Ito model and the enhanced model. The models were also optimized in 2D space (L^* C^*_{ab} planes) to remove error due to nonuniform hue in CIELAB space. Table II summarizes findings.

Model	Kl	Kc	Kh	Kt	Max error	Average error
Katoh 3d best parameters	1	2	2	NA	48.46	12.72
3D Katoh	1	2.38	.64	NA	38.24	7.32
3D enhanced	1	3.22	.61	-.41	35.32	6.48
2D Katoh	1	2.66	NA	NA	8.07	3.68
2D enhanced	1	3.33	NA	-.45	7.98	2.64

Table 3. Summary of optimization results for Katoh and enhanced Katoh models.

The results from the optimization show that the best fits occur when the coefficient of chroma is greater than that for lightness, but coefficient for hue is less. This implies that observers are more sensitive to lightness changes than to chroma changes, but they are more sensitive to hue changes than to lightness changes. This result is somewhat contradictory to Katoh and Ito's results that showed less sensitivity to hue shifts. The rotation parameter K_t shows that rotating the ellipsoid can fit the observer match data better, but it complicates the interpretation.

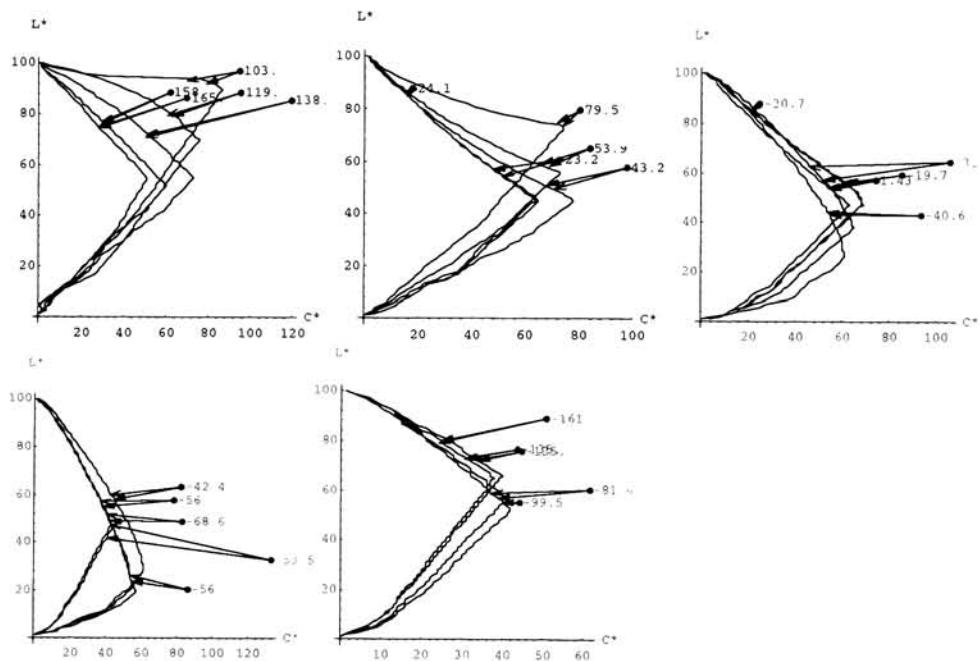


Figure 22. Comparison of Observer results with best 2d enhanced Katoh optimization parameters. Open arrow heads indicate observer results. Filled arrow heads indicate model results. Hue leaves shown are for the hue angle of the observer match.

Figs. 22 and 23 show the best results from the enhanced model. Fig. 22 shows the results of 2D matches. Notice that this fit is very close in all regions of the space. This implies that the

model may work very well when a perceptually uniform hue color space is developed. Whereas this enhanced model, as well as the original model fit the data adequately, there are some problems inherent in a mapping approach that seeks to minimize a color difference. For instance, there is a greater tendency for many to one mappings when target colors lie beyond the highest chroma values in a given region (colors lying beyond the tip of the hue leaf). This will tend to create "flat" spots on an image. A weighted vector mapping has less tendency to create many to one mappings since it is influenced more by static directions in space. Note that the weighted vector approach can fit the observer match data better than the Katoh and Ito model, but performs slightly worse than the enhanced color difference model.

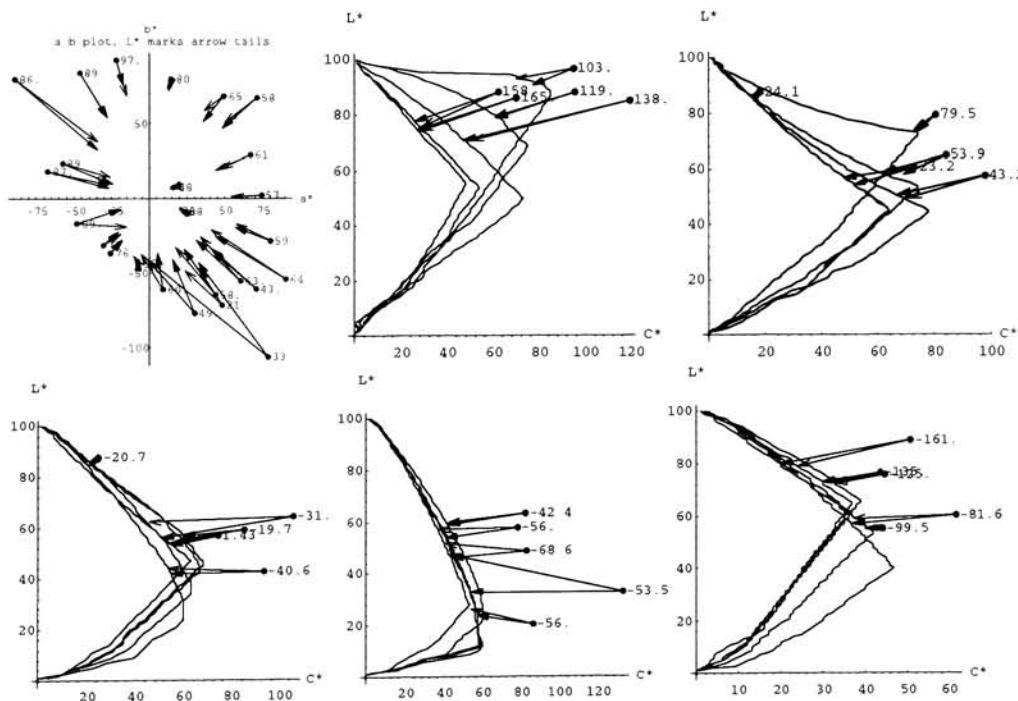


Figure 23. Comparison of Observer results with best 3d enhanced Katoh model optimization parameters. Open arrow heads indicate observer results. Filled arrow heads indicate model results. Hue leaves shown are for the hue angle of the model prediction.

These results show that both types of gamut mapping models can be used to fit the observer data fairly well. The question remains as to the effectiveness of a color match approach in determining gamut mapping vector directions

2.2.2.1.4 Conclusions to P2 (sub-experiment 1)

The major points of summary are:

1. Observers don't make simple matches of the kind that are assumed in-gamut mapping experiments to date. However, three simple models were optimized against the observer match data and found to fit the match vectors acceptably, and may be used to create a clipping or compression type gamut mapping that will create perceptually closest matches to out of gamut colors. Although the experiment was performed using solid color matching for the most part (not complex imagery), this may also apply to pictorial images.
2. This experiment shows that there is no multidimensional, statistically significant difference in any of 25 gamut mapping decisions over the four different images shown. The entire data set has been treated as if it came from the same population.
3. CIELAB, arguably the most popular color space for performing gamut mapping transforms, is not uniform enough in hue angle to make adequate (perceptually constant hue) matches.

The experiment clearly showed that CIELAB is nonuniform in hue. An experiment needed to be designed to characterize the nonuniformity over more of the color space in a simple,

macroscopic manner that can use results to create a transform to a color space whose metric hue angle is perceptually constant hue.

2.2.2.2 Preserving meaning in color matches (sub-experiment 2)

The above work was repeated almost exactly, with the exception of the addition of two new image types (sub-contexts of literally representative, and text images), and the change in the observer task. The script for this experiment was as follows:

“In the last experiment, if you did it, you attempted to find the closest perceptual match between two colored images. The criterion for matching was to find the color that made the images look as close to one another as possible.

In this experiment, you are presented with the same images and colors, but instead of attempting to make a color match between images, you are to adjust the color of the image on the left to try to make the image appear to convey the same meaning as the image on the right.

For instance, you may be making a presentation in Microsoft PowerPoint, and the color you want to use is the one in the image on the right. Your task is to pick a color from the available color palette that will convey the same message, or the most similar message as that conveyed in the image on the right. Please don’t make an adjustment that is too radical to the color of the image on the left, i.e. the basic color “name” or color category shouldn’t change (reds should remain reds, greens should remain greens, etc.).

As an example, I may be presented with a red color as a target color to match. But, the perceptually closest color is a bit less saturated. In order to convey what I might want to convey with a saturated red (for instance, danger or attention), I may want to find a color near red that doesn’t match in hue as much, but is more saturated.

In this experiment, you are presented with the same images and colors, but instead of attempting to make a color match between images, try to find a best substitute for the target color. This may or may not be the same as the closest perceptual color.

One approach to the problem is to first find a closest perceptual match. Once this is done, try to find a color that you would use in place of the target color without changing the color category (or color name)."

Some observers had difficulty understanding the task. I discussed the intention of the task with all observers who did not understand what they were to do. Even when some observers understood what they were to do, performing the task in a way that is different than a perceptually closest match was difficult.

The set of images used in this experiment is shown in figure 24.

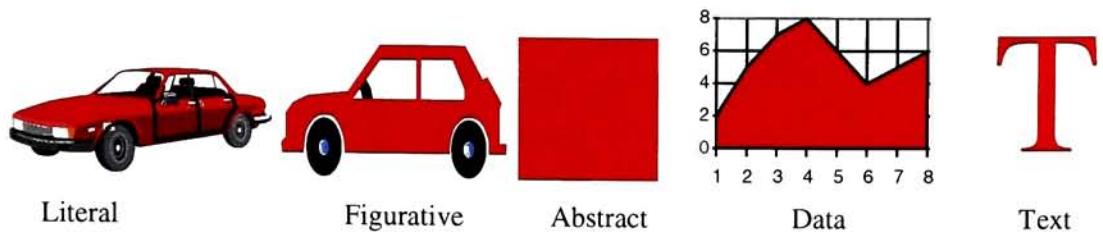


Figure 24. Images used in experiment P2, sub-experiment 2.

Twenty five observers performed the meaning preservation task for each of 5 sub-contexts and over the same 25 colors as in the previous experiment.

The results from statistical analysis showed that for the multivariate ANOVA (MANOVA), the null hypothesis of all means being the same was rejected for 17 of the 25 colors at the 10% level. Further analysis showed that the text sub-context was the most different from the other contexts. When the MANOVA analysis was re-run with the text data set removed, the null hypothesis was rejected for only 5 of the 25 colors. A univariate analysis on differences of data between individual sub-contexts shows that most univariate rejections happen between the text sub-context and others. Figure 25 shows a graphical depiction of the colors rejected across sub-contexts.

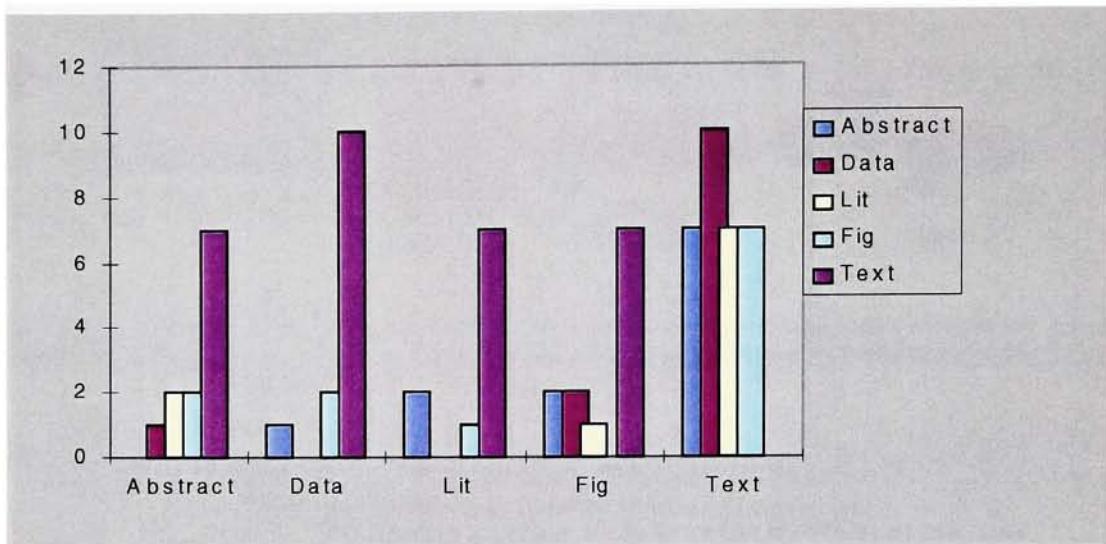


Figure 25. Histograms of colors where the null hypothesis was rejected between sub-contexts

Although the null hypothesis was rejected for several of the colors across sub-contexts, as figure 26 shows, the actual value differences of the mean mappings for each context does not show a readily identifiable systematic trend. However, when plotting the pooled mean of all sub-contexts except text against the text sub-context, several comments can be made. Figure 27 shows mean mapping vectors for text sub-context (open arrow head) and the mean of the rest of the subcontexts pooled (filled arrow heads). For many colors, the text mapping is to a lighter color than the pooled mapping, and in the green color, the text mapping is directly pointing at the origin, whereas the pooled mapping is yellower. The lighter mappings of the text sub-context may be due to assimilation effects where the image of the “T” is blended with the background white color, thus making it appear lighter.

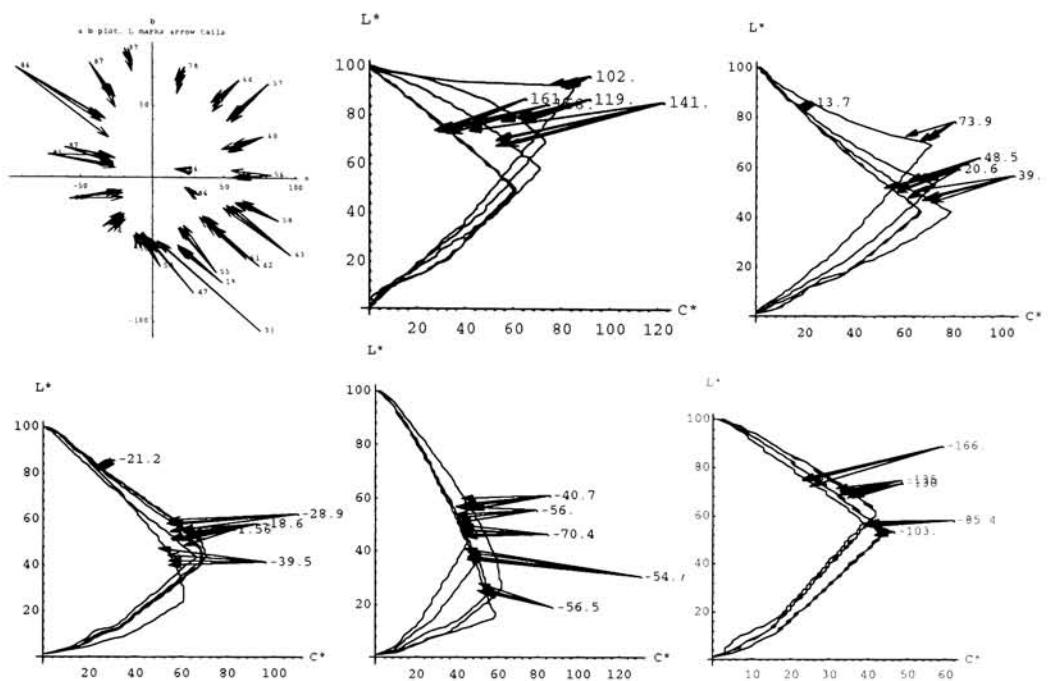


Figure 26. Mean mapping vectors for each of the 5 sub contexts.

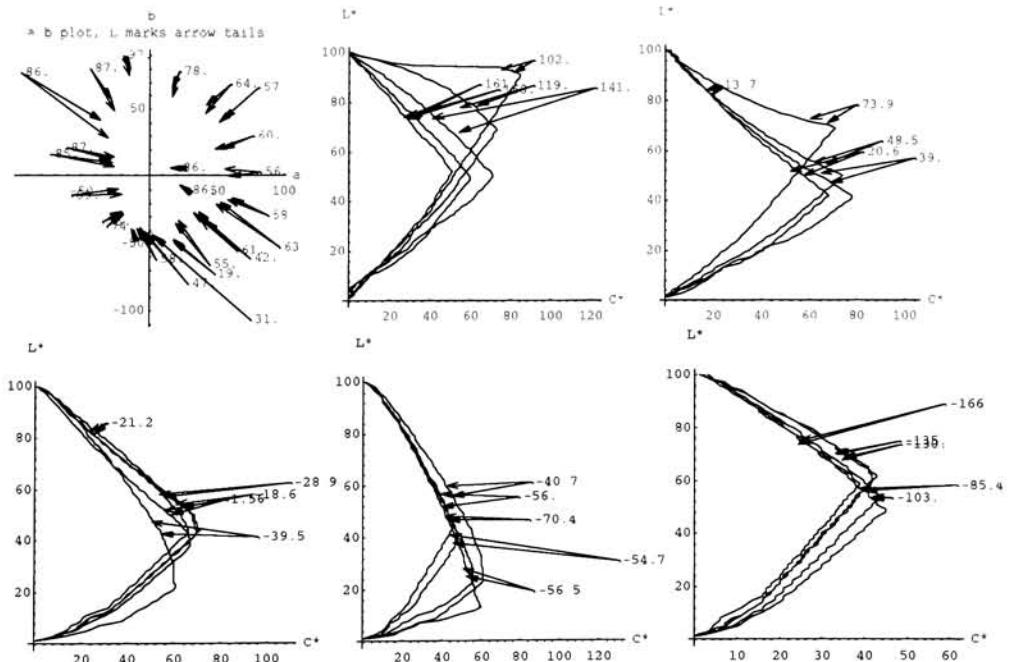


Figure 27. Mean mapping vectors of Text sub-context (open arrow heads) and the mean of all other sub-contexts pooled (filled arrow heads).

2.2.3 Conclusions leading to body of thesis work

Results from these experiments (the P2 experiments) showed that, although color matching is an interesting approach to the gamut mapping problem, it is likely limited in the granularity of the exercise. Restated, this approach was useful to derive a general gamut mapping model and to show that CIELAB color space is very non-uniform in hue in the blue region. However, using the color matching approach was limited in value when trying to derive general and systematic differences in mappings across either image type or task description.

Based on the conclusions drawn from experiments P1 and P2, it was felt that attacking the hue non-uniformity problem had more promise than following the course set out in the proposal. Thus, experiments were designed to more fully explore the nature of hue uniformity in color space, and the meaning and color matching approach the gamut mapping problem was abandoned.

3 Approach and Results

A search on data sets of constant perceptual hue was performed to see if there were bodies of data available that adequately characterized hue uniformity well enough to not have to perform the data set generation. Several sources were researched, but very little data was found that was useful. The most obvious place to start is to investigate the Munsell renotation⁵⁴ colors. Figure 28 shows the Munsell renotation colors of Value 5, plotted on the a^* - b^* plane of the CIELAB color space. Concentric “circles” denote colors of visually equal chroma and curves emanating from the origin like the spokes on a wheel denote colors of visually equal hue. At first glance, this plot indicates that CIELAB color space is fairly uniform in perceived hue.

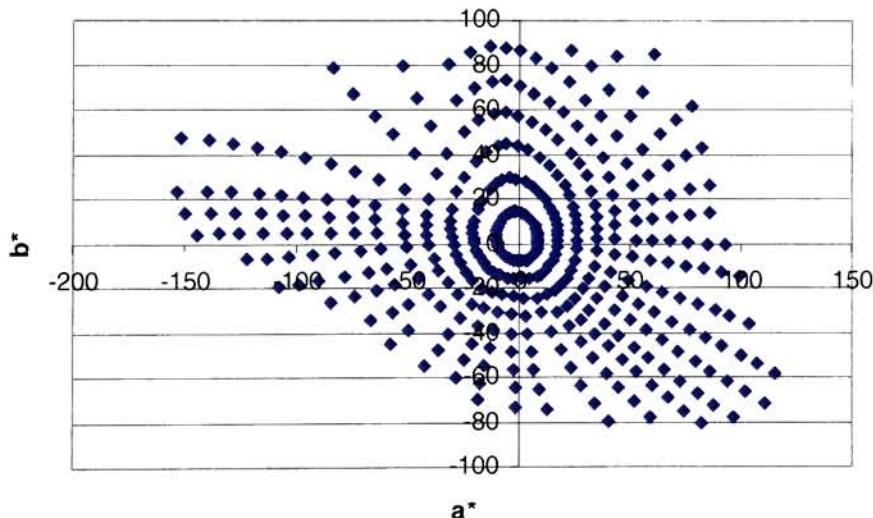


Figure 28. Munsell renotation colors of Value 5 plotted on CIELAB a^* - b^* plane.

However, upon closer inspection, it is seen that a large proportion of the Munsell renotation colors were not scaled by humans, but extrapolated. Figure 29 shows the colors that were actually scaled by observers. The colors outside of the squiggly line are all extrapolated⁵⁵ (in xy chromaticity space!). Additionally, figure 29 shows the projection of a typical CRT gamut onto the CIELAB a^*-b^* plane to illustrate the amount of color volume that is outside of the actual scaled data from the Munsell renotation. From these data, it was concluded that the Munsell data set was deemed insufficient to use to judge perceptual hue uniformity. One explanation for the low chroma limit is that the Munsell color chips were manufactured to be very durable and fade resistant, and were made with a matte finish (which limits chroma). High chroma paints that were available in 1943 were not durable and fade resistant enough to use for the Munsell color chips.

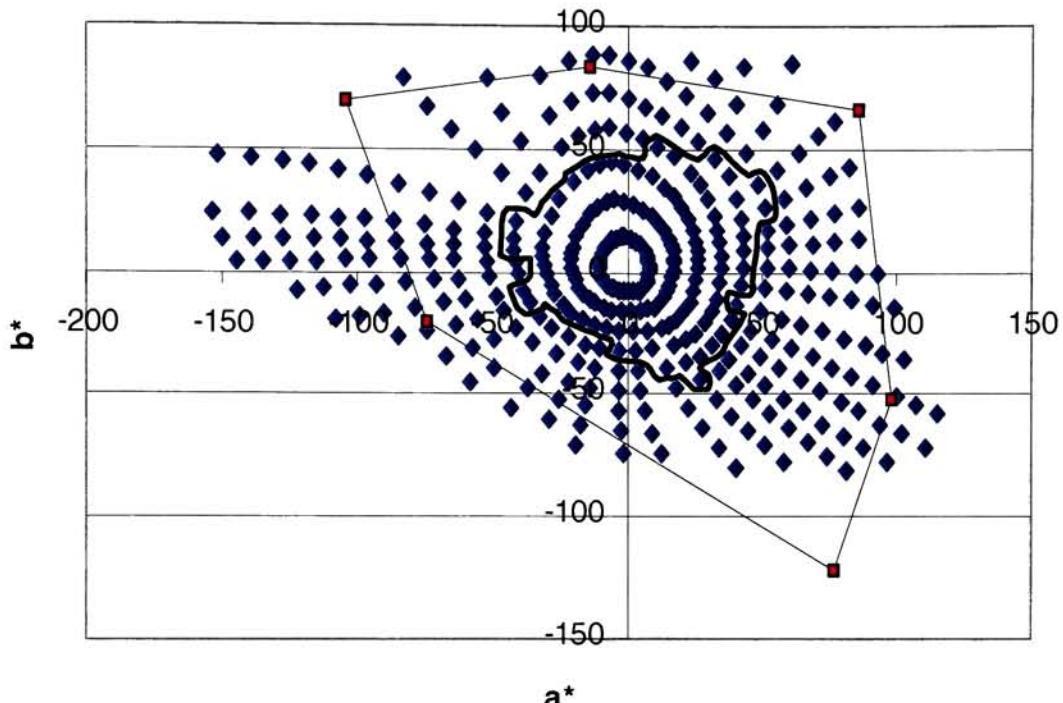


Figure 29. Munsell renotation colors with extrapolation boundary and CRT gamut projection onto CIELAB a^* - b^* plane.

There did exist one data set that was specifically derived to measure hue constancy in CIELUV color space. This was the collection of two data sets from Hung and Berns.⁵³ Two separate experiments were performed to find loci of constant perceived hue in CIELUV color space. The data sets were named CL (constant lightness) and VL (variable lightness). Figure 30 shows two plots of Hung and Berns constant hue data, one for each of the CL and VL data sets. The CL data comprise 4 points per hue angle equally spaced in chroma from max chroma to $\frac{1}{4}$ of max chroma (all of equal CIELUV lightness, with same L^* as the L^* of the

max chroma color). The VL data comprise 9 points per hue angle equally spaced in CIELUV L^* from 20 to 90. In both cases, the reference color was the maximum chroma color of that hue angle.

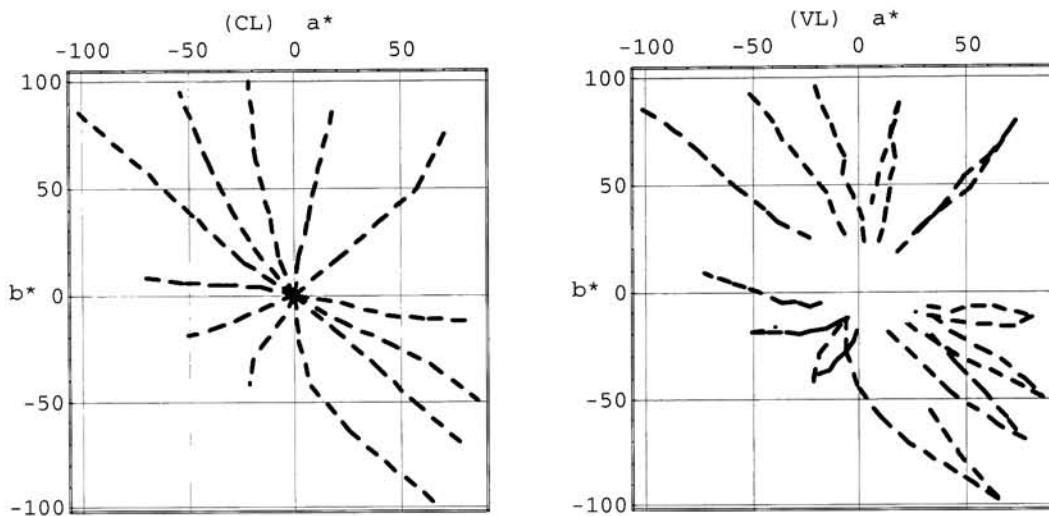


Figure 30. Hung and Berns loci of constant perceived hue. Left plot is CL data, right plot is VL data.

Note that the CL data are smooth, while the VL data appear to be noisy. Notice, in particular, that there is a crossover in the hue loci of blue and cyan-blue (lower left quadrant). This implies that the data found are inconsistent, and were therefore judged to be unusable. This conclusion is supported by the variance measures reported by Hung and Berns in their article. The VL data had about 3 times as much variance as the CL data.

This analysis of available data from which to derive surfaces of constant hue in color space left us with about 48 data points. This is a meager number when considering the volume of color

space occupied by a CRT display. Thus, an experiment was conducted to find surfaces of constant perceived hue. The goal of the experiment was to create a database large enough to be able to derive and test a color space which has a metric hue angle that represents constant perceived hue.

3.1 Experiment E1. Finding surfaces of constant hue in color space

3.1.1 Abstract

A colorimetrically calibrated CRT display was used to measure constant perceptual hue surfaces in color space. 306 points over fifteen equally spaced hue angles (every 24 degrees) in CIELAB color space were sampled. An average of 20 lightness-chroma combinations per reference hue plane was sampled. Thirty observers performed the matching task three times each. Intra-observer variation was used to weight mean observer hue matches for each of 306 colors. Analysis of perceived hue uniformity was performed in CIELAB and CIECAM97s color spaces. Other constant hue experimental results are analyzed and compared to data obtained here.

3.1.2 Introduction

The goal of a color appearance space, among other things, is to provide a structured way to model perception. That is, it should have appearance attributes that accurately reflect the perception of that attribute. Attributes include lightness, brightness, colorfulness, chroma, saturation and hue.⁵⁶

The most studied and best-understood perceptual attributes are lightness and brightness, especially of neutral colors. This is not surprising, as lightness is the perception that defines shapes and textures to the greatest extent and it is the most basic. Indeed, the first two colors in most evolving lexicons are black and white⁵⁷. Because of this attention lightness and

brightness are fairly well understood phenomena. However, other perceptual attributes have not had as much devotion.

All attributes are not equally important. Depending on the application, it may be relatively more critical to accurately model one attribute. For example, when gamut mapping, if one wishes to preserve the perception of hue while altering chroma and lightness, it is quite important not to change the perceptual attributes of that hue. At the very least, the hue name of the color must be maintained.

Hung and Berns⁵³ showed that for four color appearance spaces, none have accurate representations of hue neither with lightness nor chroma. Additionally, color spaces have different nonuniformities. For instance, for CIELAB the cyan-blue to blue-purple region is the least uniform, whereas for Nayatani's space, the red region shows the least uniformity. It has been observed by the author that the nonuniformities in CIELAB color space are severe enough in the blue region to unambiguously cross color name boundaries between blue and purple at a given metric hue angle.

If surfaces of constant perceived hue that correctly modeled human perception existed, presumably a color space could be derived that fit the data. Unfortunately, such a set of data does not exist in the literature. Munsell and NCS color order systems data were scaled only for colors with relatively low chroma because of material limitations. Other experiments have

been done to find constant hue surfaces over a larger gamut, but have been too cumbersome and difficult to get statistically significant results. Hung and Berns give a more detailed analysis of these drawbacks.

While the Hung and Berns' experimental results added more data to the available set, it did not cover enough of the gamut to enable good modeling. Across 12 hue angles, their color sampling was limited to 3 colors of equal lightness as the reference color plus 8 colors of varying lightness along the gamut edge. While the experiment showed that perceived hue is a function of both lightness and chroma, the relationship of hue with the other two dimensions was not fully examined.

In order to complete the description of full surfaces of constant perceived hue, an experiment was performed that sampled an average of 20 colors over 15 equally spaced hue angles in CIELAB space. The experiment was designed to be easy to perform without introducing elements that might sacrifice precision. Thirty observers performed the experiment three times each.

3.1.3 Experimental

The iterative approach of Hung and Berns was deemed to be too time consuming for the number of colors to be tested. An alternate approach was designed that gave the observer freedom to select hue matches using a slider bar that controlled the CIELAB hue angle with a precision of .5 degrees at all lightness and chroma values. This allowed observers to match the 306 colors of one trial run of the experiment in an average of 90 minutes. Observers were

allowed to make as many or as few matches as they wanted in a given sitting. Test stimuli were presented in a shuffled order for each trial run of the experiment.

3.1.3.1 *Script*

Each observer was given a script with the same instructions (it was part of the user interface display so they could refer to it as often as they wanted). The instructions were:

“Make the hue of the test patch the same as the hue of the reference patch. The same hue means the color has the same amount of redness, greenness, yellowness, or blueness.”

For the first session each observer was allowed to get acclimated with the interface and ask questions until they were comfortable with the task. Some observers had difficulty discriminating between a hue matching task and finding the minimum color difference between the test stimuli and the reference. This did not pose a problem in general, as the minimum distance and matching hue usually resulted in the same answer, since the user was confined to the CIELAB hue dimension. There were conditions due to gamut limitations, though, where the minimum color difference meant deviating from a hue match. There were four subjects who had this confusion despite verbal attempts to describe the task and to warn about the difference. These subjects also had large intra-observer variance, so their contribution to the overall mean was quite small.

3.1.3.2 User interface

The experiment was performed on a Sun Sparc Ultra 1 with a 20" Sony GDM-20E20 monitor. The resolution was 1152x900 with a refresh rate of 76 Hz and a 0.31mm-phosphor trio pitch. The CRT display was kept on during the extent of the experiment (which ran from May through July, 1997) to minimize drift from power cycling. The layout of the interface is shown in figure 31. The stimulus squares for both test and reference subtended an angle of 4.2 degrees.

The white border had a luminance of 71 cd/m^2 . The luminance factor of the background gray was .35 (25 cd/m^2). The anchor colors showed what the test stimulus would look like if the slider bar was pulled as far as it could go in either direction. The hue extent of all colors was ± 50 degrees. This amount was a good tradeoff between sensitivity and range of hue angle based on preliminary experiments. The user was instructed to make as good a match as possible, then proceed by pressing the next button. If desired, the user could press the previous button to see matches that were already made. The user could quit at any time and the state would be saved, so that the next session would begin where the last one left off (if the trial was not completed). The entire experiment was conducted using the mouse as the only input device. No keyboard input was used.

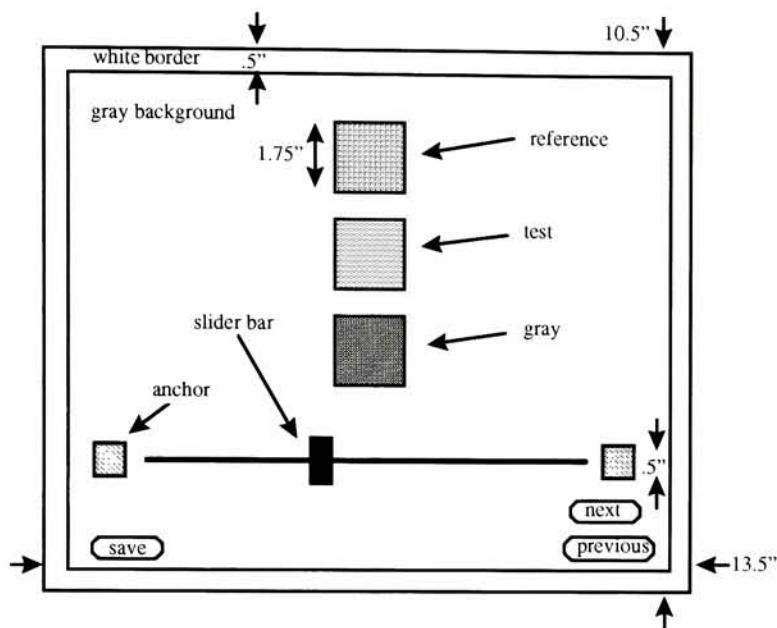


Figure 31. User interface for experiment.

The topmost patch is the reference patch. There were 15 such patches, but they changed at every “next” color presentation because the colors were randomized for each trial. The middle patch is the test color. This is the color that changes when the slider bar is moved. The colors update as soon as the slider bar is moved, and remain where the user leaves the slider. The match decision is recorded when the “next” or “previous” button is pressed. The bottommost patch is a gray reference patch that has the same lightness as the current test patch. This patch was included to be more compatible with the Hung and Berns interface, which also had a gray patch. The purpose of the gray patch was confusing to most observers and would not be included if the experiment were to be rerun.

3.1.3.3 Viewing conditions

The CRT was set up to have a white point near D65, with a luminance of $71\text{cd}/\text{m}^2$. The experiment took place under dark surround conditions. The walls of the room were covered with black felt to eliminate the possibility of reflection from the CRT.

3.1.3.4 Calibration

The CRT was calibrated using a variant of the Gain-Offset-Gamma⁵⁸ model. First the white point was found through a search method that altered the digital count of white (it could not be altered by analog methods). Then the maximum red, green, blue channels were measured to find the chromaticities of the primary colors. The gray scale was then measured to find the luminance-digital count relationship. Ten-bit tone-reproduction curves were made from spline fitting data from 20 points along the neutral axis. Forward and inverse models were created, and the model was tested for inversion error and colorimetric accuracy. The inversion error of the model was less than $.001 \Delta E^*_{ab}$. The average ΔE^*_{ab} for a 5^3 sampling of colors was approximately 1, and the worst case error was $\sim 3 \Delta E^*_{ab}$ units. The CRT was calibrated three times over the course of the experiment when the average color error became greater than 2.5.

3.1.3.5 Color selections

306 colors were selected over 15 equally spaced hue angles in color space. Figure 32 shows an example sampling of a hue leaf at 0 degrees.

The reference color was chosen for each hue leaf to have moderate to high chroma, and to be close to the center of the area of the hue leaf. It was felt that the reference color should not be

too far from any test color. This is in contrast to Hung and Berns who chose the reference color to have the highest chroma. Figure 33 shows a projection onto the a^*-b^* plane of the entire test color set (defined in CIELAB).

Colors were chosen to span as much of the available color gamut as possible. Sampling was done as uniformly as possible, but gamut restrictions made sampling on a uniform grid impossible. Instead, within the gamut of a given hue leaf, the colors are uniformly sampled, but at the edges the samples follow the gamut edge more closely. It was also felt that the density of colors should be consistent around the hue circle, instead of having consistent sample period, with less dense sampling in the cyan region.

3.1.3.6 Observers

30 observers, 18 males and 12 females, between the ages of 27 and 59 participated in the experiment. The author was pleasantly surprised to learn that the lure of a free lunch would entice so many willing and qualified observers. Each observer was given a brief color test with the Ishihara color vision plates. Using this prescreening, four adult males were identified to have color vision deficiency and were not included in the experiment. Subjects ranged from being very experienced color observers to complete novices. Knowledge of the objective of the experiment should not significantly influence observers' choices although, as expected, experienced observers had much lower intra-trial variability.

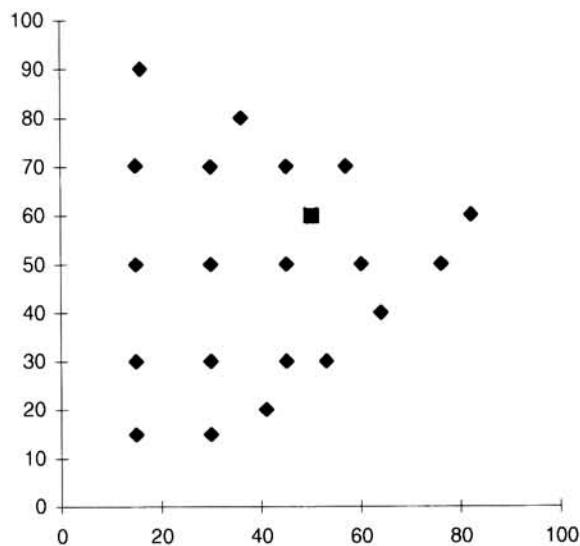


Figure 32. Hue leaf sampling at 0 degrees. The large square is the reference color.

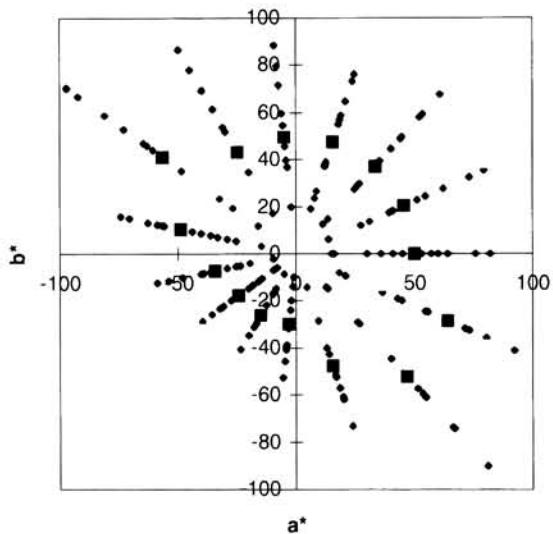


Figure 33. Projection of test colors onto a^* - b^* plane. Large squares indicate reference colors.

3.1.4 Results and Discussion

3.1.4.1 Data set calculation

Each subject performed the hue-matching task three times. The data were analyzed in CIELAB Δh_{ab} space. Histograms of the 90 samples of each color were made. Figure 34 shows histograms for the hue surface at zero degrees.

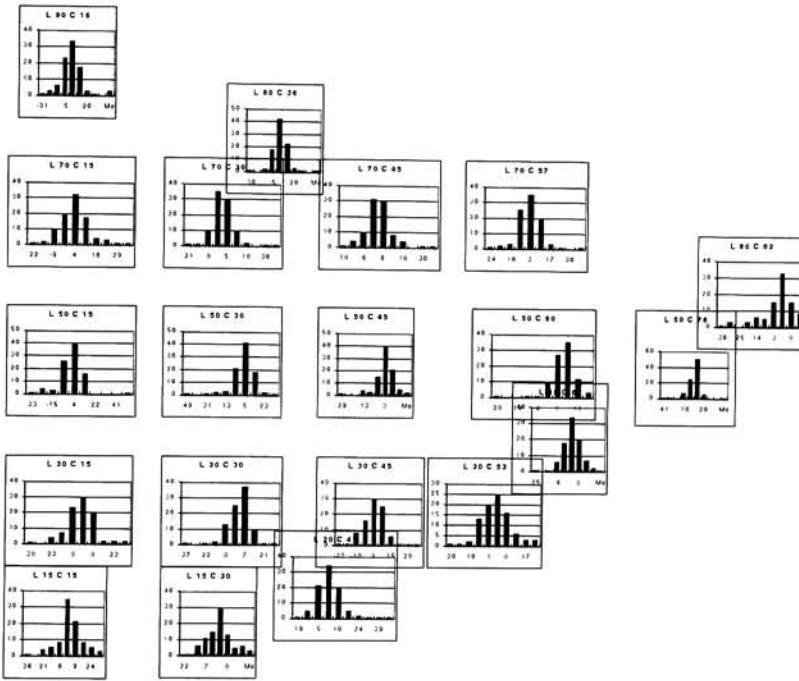


Figure 34. Histograms of hue matches for 90 observations at reference hue angle of 0 degrees. Histograms are plotted in the respective locations in the $L^*-C^*_{ab}$ plane of the reference colors.

The histograms show that as expected, there is a central tendency for color matches, and for the most part, the tails are symmetric. Confidence limits at 95% (1.96 times standard error) were calculated for each color from the entire data set of 90 observations. The standard error of the sample mean is the standard deviation of the sample mean, and is calculated from the sample standard deviation divided by the square root of the number of samples ($\sigma/\sqrt{90}$). This

is an estimate of how different the means would be if the experiment were to be run again. Figure 35 shows the average and standard deviation of 95% confidence limits for each hue surface.

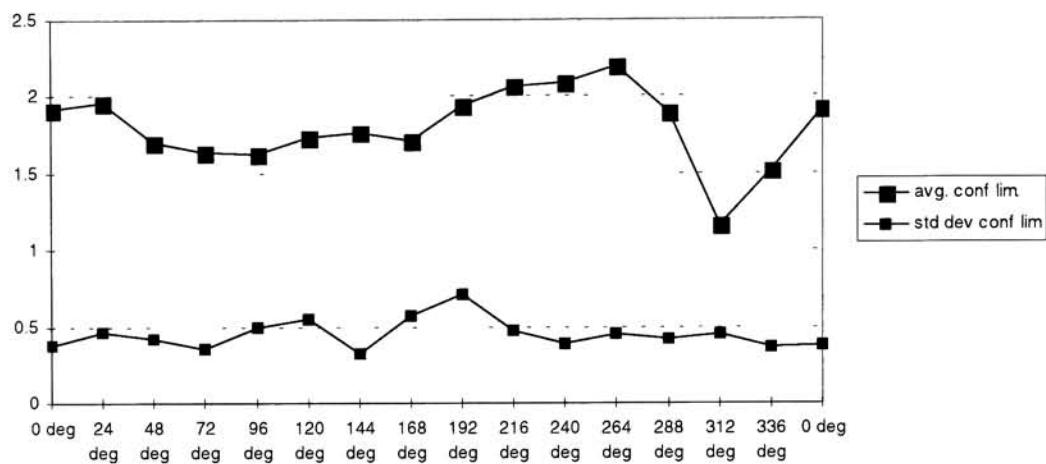


Figure 35. Average and standard deviation of 95% confidence limits of mean hue matches as a function of reference hue angle. Units of y axis are in degrees (CIELAB Δh_{ab}).

Observers noted that for the cyan region (192 degrees – 264 degrees) there was a large degree of uncertainty about where to make a match. Many hue locations looked very similar. Many also commented that the purple-blue colors were very easy to match and they had a high degree of confidence in making those matches. This can be seen from the average confidence limit plot. The large average confidence limit regions show the most variability in the match data, since it was harder to make a match. The 312-degree hue surface is the blue purple in color and has very low variability. Variability of data may also be viewed as a tolerance metric. If there is more variability in the data, there is a larger area over which a match is acceptable.

These data might be useful in helping determine suprathreshold hue tolerances, and is in some ways similar to suprathreshold tolerance data recently reported by Berns, Qiao, and Reniff⁶.

Overall weighted mean data were calculated. Weights were derived from individuals' inverse within-observer variability. A reasonable measure of variability was needed that gave more weight to observers who were more internally consistent without favoring a small minority of observers all the time. Two metrics were investigated, sample variance and absolute difference. A good weighting metric should include more influence from internally consistent observers without completely excluding others. The inverse of absolute difference was chosen over variance, since it was more equitable in weighting observers. Inverse variance weighting gave too few observers (between 1 and 5 for most colors) most of the weight in determining the mean. Figures 36 and 37 show the histograms of weights for inverse absolute difference weighting and inverse variance weighting respectively.

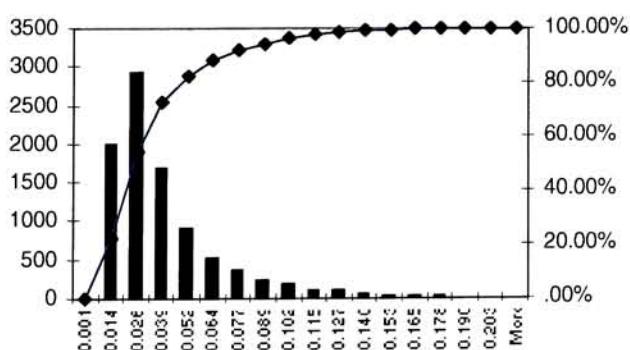


Figure 36. Histogram over the entire data set (306 colors X 30 observation means = 9180) of weights calculated from inverse absolute difference.

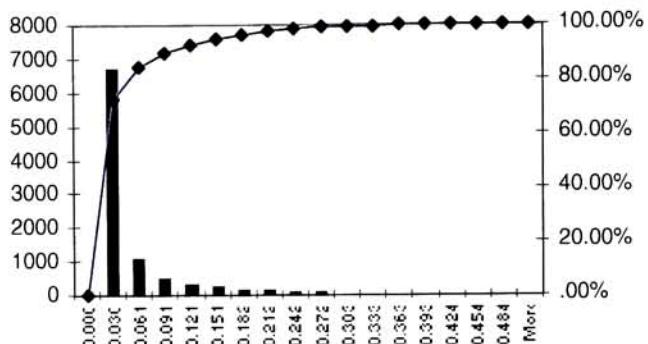


Figure 37. Histogram over the entire data set (306 colors X 30 observation means = 9180) of weights calculated from inverse variance.

If the mean was calculated without weighting, each sample would be normalized by the inverse of the number of samples, in this case 30, since we are taking the mean of the intra-observer mean results. With constant weighting, each observation would be normalized by ~ 0.0333 . For the inverse absolute difference case (figure 36), roughly 45% of weights are above 0.0333, whereas in the inverse variance case (figure 37), only 27% of the weights are above 0.0333.

Weighted mean data with confidence limits are shown in figure 38. The figure is a 2 dimensional representation of the mean hue matches around the hue circle. The x-axis shows reference hue angle, and the y-axis represents the hue match with $\pm 95\%$ confidence limits. The data for each hue is sorted first by lightness (from dark to light), then by increasing chroma. This means that the overall slope of the set of points that make up a constant hue surface represents the variation of hue with lightness. The variation of hue with chroma can be seen as points that make a sawtooth pattern within a set of constant hue points. Note that

the 95% confidence limits are significantly smaller than most of the trends on the constant hue surfaces.

Note at zero degrees the slope of the points is essentially 0. From 0 degrees to 72 degrees, the slope of the points becomes increasingly positive, then decreases back to 0 at 144 degrees. This corresponds to the CIELAB hue angle for constant hue surfaces becoming larger as with increasing lightness. From 144 degrees to 312 degrees the slope of the constant hue points becomes increasingly negative, then tends back to zero as 0 degrees is approached. The constant hue chroma relationships are most apparent at 264 and 288 degrees where the set of constant data points deviate most from a straight line. Chroma relationships with hue are visualized in a more intuitive way later. Figure 38 is primarily for visualization of the variation of lightness with constant hue.

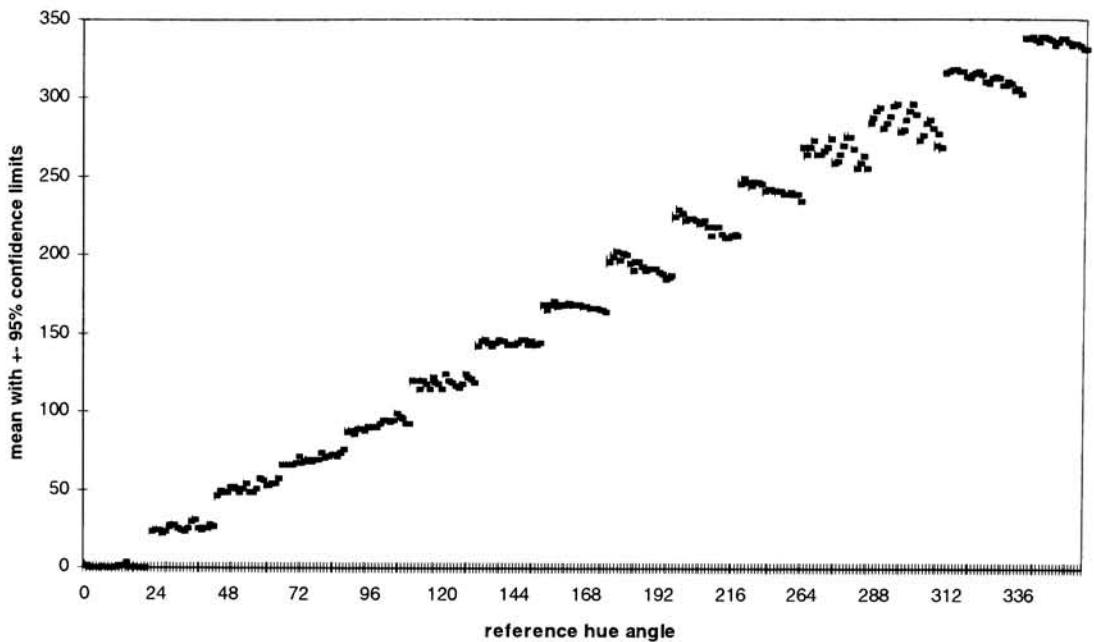


Figure 38. Weighted mean hue matches in CIELAB space with +/- 95% confidence limits.

Figures 39 and 40 show constant hue surfaces on the CIELAB a^* - b^* axes. Figure 39 shows a projection of the surfaces, and figure 40 shows slices at every 10 lightness units from $L^*=10$ to $L^*=90$. The dots on figure 40 are the reference colors.

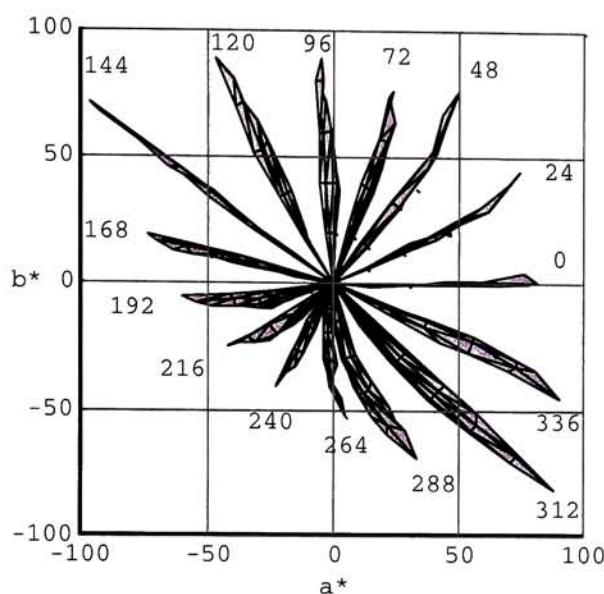


Figure 39. Projection of constant hue surfaces onto CIELAB a^* - b^* plane.

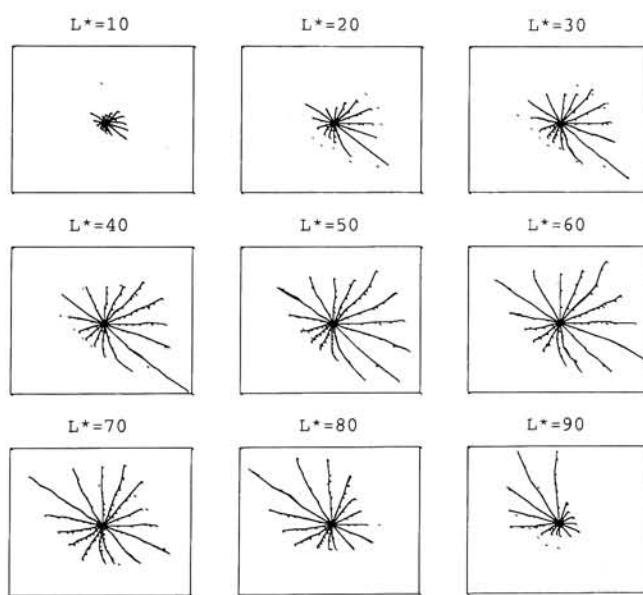


Figure 40. Loci of constant hue in CIELAB a^* - b^* at different lightness levels. Dots represent reference colors.

It is clear that the blue region has the most chroma nonuniformity in CIELAB space. These data support results from Hung and Berns. Figures 41, 42 and 43 show the same data plotted in the recently proposed color appearance model CIECAM97s⁶ (from the CIE TC1-34). The dots on figure 43 denote reference colors.

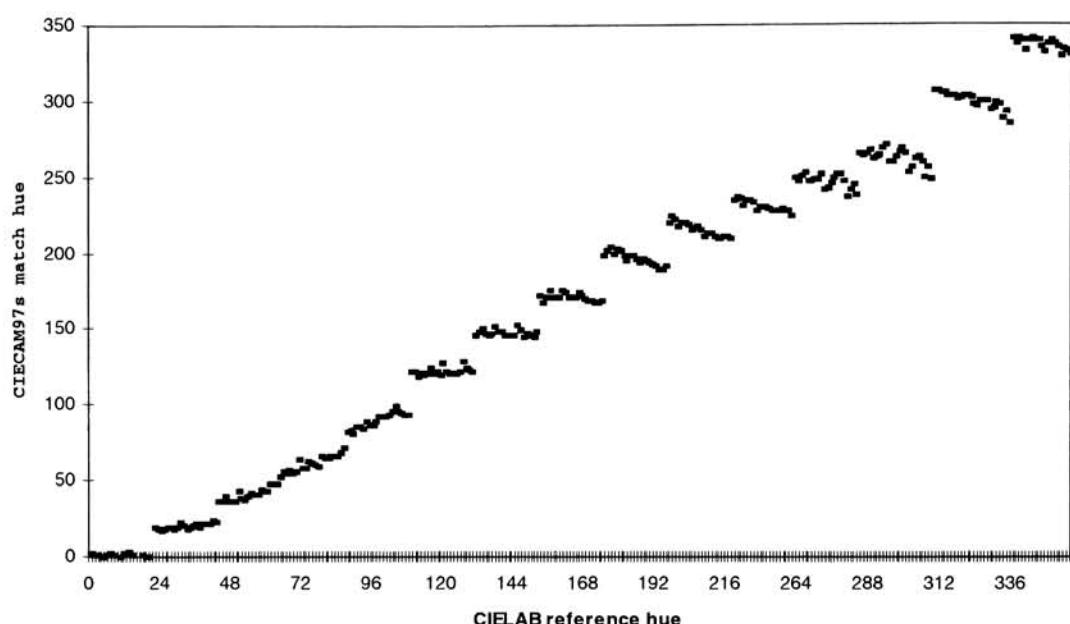


Figure 41. Weighted mean hue matches plotted in CIECAM97s color appearance space. The x-axis is labeled by CIELAB reference angle.

The parameters used in calculating the CIECAM97s coordinates were: $L_A=14.2 \text{ cd/m}^2$ (20% of luminance of white), $Y_b=35$, $X_w=95.01$, $Y_w=100$, $Z_w=108.81$, and the conditions for dark surround: $c=0.525$, $N_c=0.8$, $F_{LL}=1.0$, $F=0.9$. Note that, similar to but not as drastic as CIELAB, the blue region has significant chroma non-uniformity. The lightness relationship with hue in CIECAM97s can be seen to be very similar in trend to the CIELAB data.

The degree of nonuniformity can be quantified in many ways. One way is to calculate the mean color difference from the mean hue angle for each hue surface. This gives a convenient and simple metric for cross color space comparison. This assumes that there is a reasonable similarity between Euclidean distances in both color spaces, and that they are equally discriminative throughout the entire space. While these assumptions are not strictly correct, the metric gives a feel for a first order approximation of non-uniformity. Figure 44 shows the mean distance from the mean hue angle for each hue surface for CIELAB and CIECAM97s color spaces.

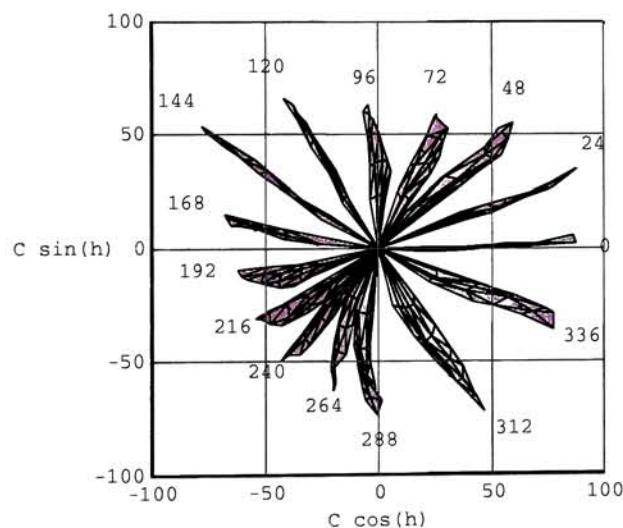


Figure 42. Projection of constant hue surfaces onto CIECAM97s $C \cos(h)$ - $C \sin(h)$ plane. Hue leaves are labeled by the CIELAB reference hue angles for comparison.

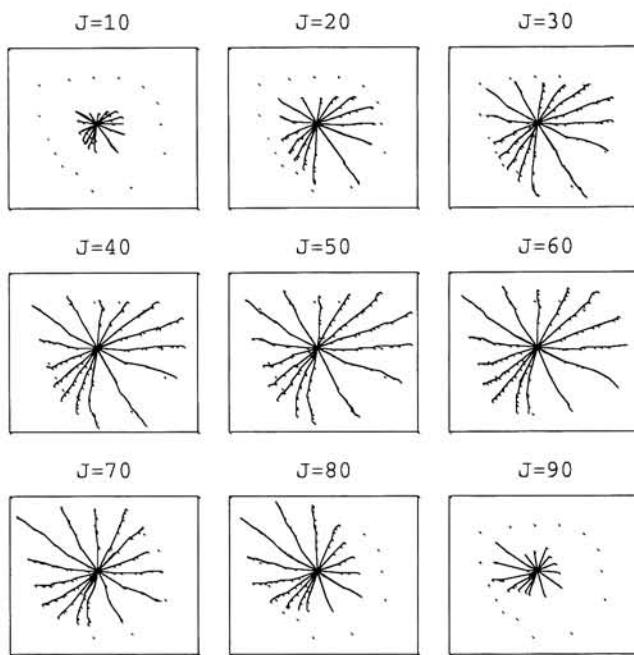


Figure 43. Loci of constant hue in CIECAM97s $C \cos(h)$ - $C \sin(h)$ at different lightness levels. Dots represent reference colors.

The CIECAM mean distance is greater than the CIELAB mean distance for 10 of the 15 hue surfaces. CIELAB distance is greater than CIECAM97s for 2 hue surfaces, and they have approximately the same error on the remaining 3 surfaces. This may lead one to conclude that CIELAB is more uniform, but that would ignore the fact that the 288 degree hue surface has considerably more chroma curvature than CIECAM97s.

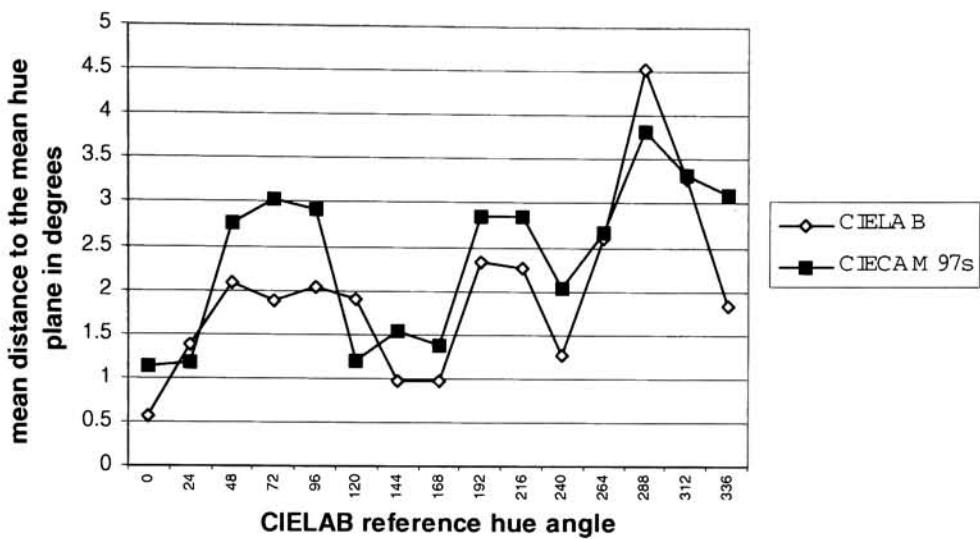


Figure 44. Quantification of average hue non-uniformity for CIELAB and CIECAM97s.

Another way to quantify hue non-uniformity would be to measure the hue angle difference between the greatest hue deviations from the mean in both directions. This would give a worst case metric that might be more interesting to workers deciding which color space to use for hue preserving gamut mapping operations. Figure 45 shows the maximum difference from the mean in hue angle for each hue surface.

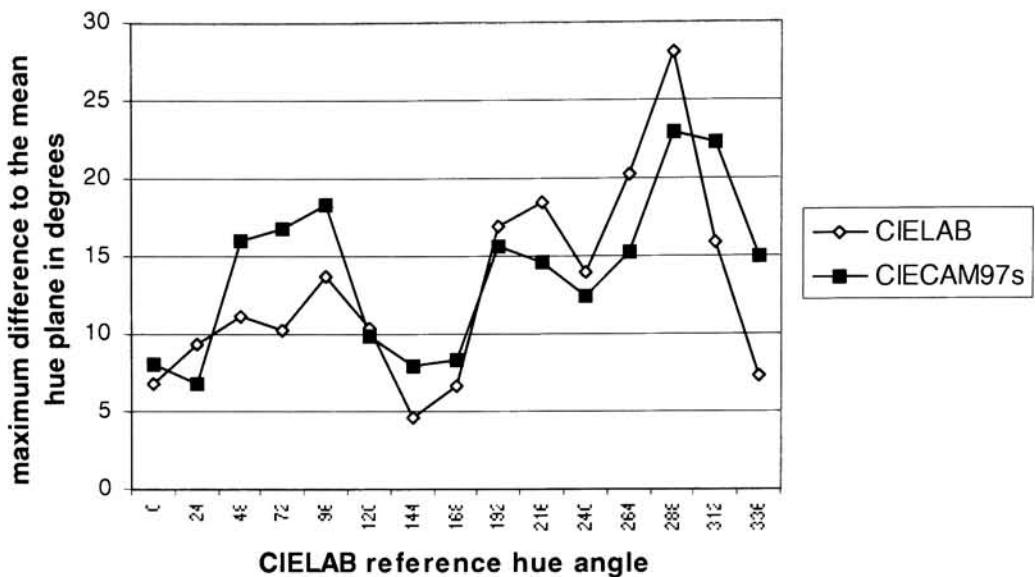


Figure 45. Quantification of maximum hue non-uniformity for CIELAB and CIECAM97s.

Note that in figure 45, the entire region from 192 degrees to 288 degrees has higher maximum error in CIELAB.

Figure 46 shows the projection of hue surfaces onto the opponent axis of linear opponent color space as defined by: lightness axis = CIE Y, opponent axis 1 = X-Y, opponent axis 2 = Y-Z. Notice that the curvature for the blue region is significantly reduced in a non-compressed color space. Although this is interesting, it is not suggested that one work in this type of space just because the curvature of blue is reduced.

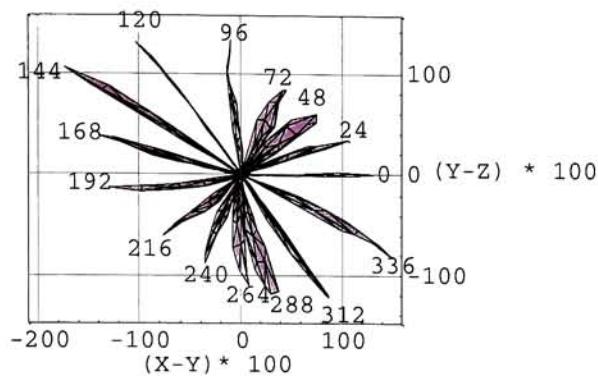


Figure 46. Constant hue surfaces shown in a linear opponent space ($Y, X-Y, Y-Z$). Numbers represent CIELAB reference hue angles.

Figure 47 shows constant hue surfaces plotted in the u^*-v^* plane of CIELUV color space. Notice that the blue curvature is lessened in this space, although visual inspection of the color space shows clear color name change from blue to purple along a constant metric hue angle equal to that of maximum pure blue stimulus for a CRT display.

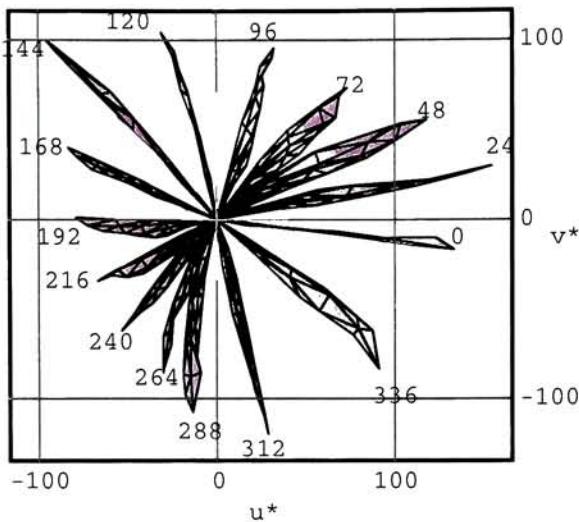


Figure 47. Constant hue surfaces shown plotted in CIELUV color space. Numbers represent CIELAB reference hue angles.

Figure 48 shows Hung and Berns constant hue loci for constant lightness (CL) and variable lightness (VL) experiments superimposed over the results found in this research. For the CL loci, the chroma non-uniformity trends are very similar. The respective CRT gamuts are somewhat different, but the overall agreement appears to be very good. This should not be surprising, as both experimental designs and presentation medium were quite close. The varying lightness hue loci were derived for 8 lightness levels on the edge of the CRT gamut at each reference hue. The VL data match to current findings is not as good as the CL data. In the blue-purple region, there is some degree of overlap between hue surfaces. This may be due to the fact that Hung and Berns only had 9 observers, and as such had a higher degree of uncertainty in the mean result.

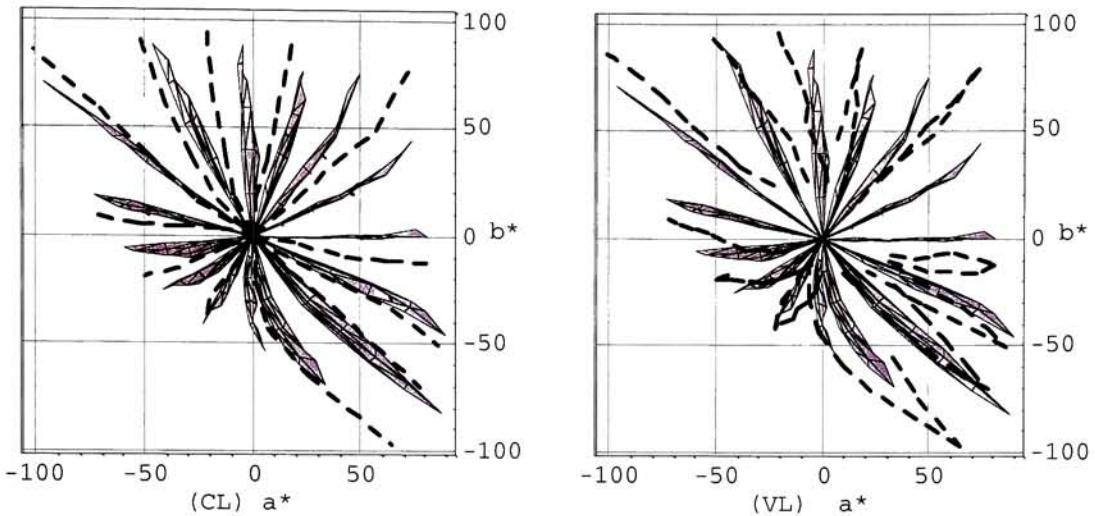


Figure 48. Hung and Berns data plotted on top of experimental results in CIELAB space. Left graph compares CL (constant lightness) data. Right graph compares VL (variable lightness) data. Hung and Berns data are shown in bold dotted lines.

In order to verify that the surfaces derived here are truly constant perceived hue, additional work was done. An experiment was conducted to compare constant metric hue surfaces for perceptual uniformity. These experiments are described in a later section entitled "verification experiments."

The results from this experiment apply strictly to viewing under dark surround conditions on a CRT screen. It is not clear if the data also applies to bright surround conditions and viewing hardcopy output. More experiments should be planned that verify that the constant hue surfaces hold under other viewing conditions.

Another concern has to do with the gamut limitations of the experimental medium, namely the CRT gamut. To make these results general, some method of extrapolation should be done to map the surfaces out to larger chroma values. It would be desirable to extrapolate to chroma values at least as large as can be created by current hardcopy marking technologies. Again, with extrapolation comes the need to verify accuracy of the resultant data set.

3.1.5 Conclusion

An experiment was conducted to find surfaces of constant perceived hue in color space. 306 colors were tested over 15 equally spaced hue angles in CIELAB color space. Results show that there is significant lightness and chroma non-uniformity in both CIELAB and CIECAM97s color spaces. Summary error metrics were developed and shown. Results agree fairly well with Hung and Berns constant lightness data loci of constant perceived hue.

One of the goals of this work was to gather enough data to enable the modeling of a constant perceptual hue color space that is functionally related to CIEXYZ or an equivalent color space. This would allow users to transform their colorimetric data to constant hue color space before gamut mapping, then transform back for color rendering.

3.2 Development of constant hue color spaces

Several approaches were taken to try to develop models of a color transformation that would result in a metric hue angle consistent with perceived hue. The two approaches that yielded useful results were the development of neural net based correction, and the development of a visualization tool that was used to optimize simple, well known transformations. The neural net approach was useful for generating three dimensional lookup tables that were used to create transforms between CIELAB and a constant hue color space. The visualization tool was used to find parameter values for a “good enough” tradeoff between appearance attributes in general and hue uniformity in particular within a well understood structure for color space transformations.

3.2.1 Neural Nets

3.2.1.1 General overview

It was initially thought that neural net training could be used as a mechanism to find a general purpose, invertible color space description that was uniform in perceived hue. This turned out not to be the case, but neural networks were useful in creating table based hue correction approaches for use in verification experiments.

The only completely invertable (with no null space) neural net is one in which the hidden layer has the same number of nodes as the input and output layers. A neural net with one hidden layer with 3 nodes, and an input and output layer with 3 nodes each is shown in figure 49.

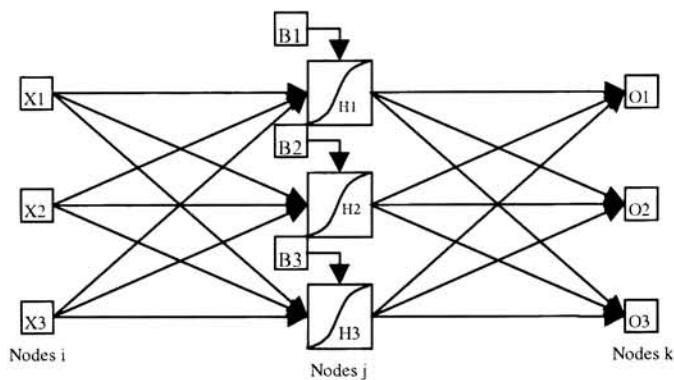


Figure 49. 3 node, 1 hidden layer neural net

The equation for computation of the numerical value for each of the hidden layer nodes is the equation $H_j = f(\sum_i w_{ij} \cdot X_i + \beta_j)$ where i is the index for each of the input nodes, and j is the index for each of the hidden nodes. The function f was chosen to be tanh, a sigmoidal function that has a slope of 1 at the origin. This function can be used to model compressive, expansive, and linear function types depending on where in the domain the input data lies.

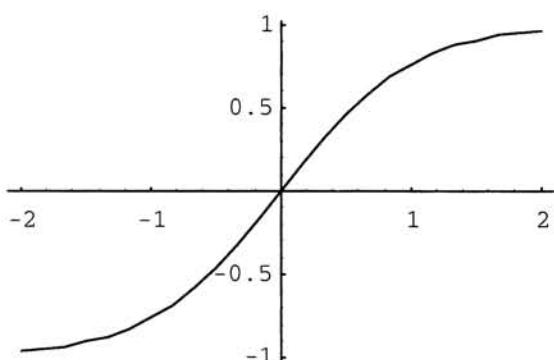


Figure 50. Tanh function is expansive in the negative range, linear near the origin, and compressive in the positive range.

The bias term B can shift the domain to the appropriate part of the curve to match the nature of the nonlinearity. The output node is computed as a weighted sum of hidden layer nodes,

$O_k = \sum_j w_{jk} \bullet H_j$. This is the general model for a neural net functional form, except that the

special case of having 3 hidden layers makes it invertable. This can be shown by modeling the neural net by the equivalent functions in matrix form. The matrix form of the 3 node hidden layer network is shown in equation 4.

$$\begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \bullet \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} w'_{11} & w'_{12} & w'_{13} \\ w'_{21} & w'_{22} & w'_{23} \\ w'_{31} & w'_{32} & w'_{33} \end{bmatrix} \bullet \begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix}$$

Equation 4. Matrix form of forward computation of 3 node hidden layer neural network.

The inverted form is obvious from equation 4, so is not shown. The matrices W and W' must be inverted, then the model is run backward.

3.2.1.2 Modeling the CIELAB transformation

To test whether the neural net approach was viable, it was tested on a known transform set. The 306 color points that were the starting hue colors for the constant hue surfaces experiment (E1) were used to test if the neural network approach could be used to model the transform from CIEXYZ (D65) to CIELAB color space. The same colors were used to train and test the network. After training the network for several minutes, the training was

completed, and the network was tested. The methods used to build the network and training algorithms are discussed in detail in references 59-61.^{59,60,61} The ΔE^*_{ab} error from testing the trained data was 0.07, with a maximum error of 1.52. This shows that a neural network of the type shown in figure 49 can effectively model the CIEXYZ to CIELAB transformation. This should not be surprising since the transformation from CIEXYZ can be generalized to be a 3x3 followed by a nonlinearity followed by another 3x3. The bias term in this case was trained to fall on the part of the tanh curve that models a compressive function.

3.2.1.3 Modeling the constant hue data set

The problem in which one knows the function to use to create the transformation between the tristimulus values and the uniform color space (e.g. CIEXYZ to CIELAB) can be used to test whether the training algorithms can accurately approximate the function. The problem that remains once this is shown is to find the coordinates of the uniform color space that correspond to the tristimulus values of the constant hue data. Immediately the issue arises of trying to pick appropriate coordinate values of the uniform color space. We know that the data that are in the same hue surface should lie on a plane, but the correlates of lightness and chroma are almost arbitrarily decided upon. It was initially thought that some feedback mechanism could be found for the training algorithm to “push” the data toward a convergent solution that had good correlates of lightness, chroma and hue. This was not the case for the 3 node hidden layer network. The nature of the neural network is to try to fit the coefficients (weights) of the model to minimize the error between the coordinates of the input and the coordinates of the output. But there is no mechanism to tell the output coordinates where

they should go to minimize the error. Additionally, there are no extra degrees of freedom to use to force the model to account for errors of the output coordinates, as there would be if the network had more nodes. Many iterative approaches were examined that altered the base hue angle of the hue planes, and used the predicted model to influence the placement of the coordinates for the next iteration. None of these approaches yielded interesting results. The invertable neural network was thus abandoned for use in color space model development.

The more general neural network is not invertable, but is useful in creating a forward transform that can be inverted using table based approaches.⁶² Because neural nets are notorious for their extrapolation craziness, extrapolation was done by hand to $C^*_{ab}=127$ for all hue angles, and for intermediate hue angles. Sixty points were added to the training data set, making it 366 points. The neural network consisted of 6 input nodes, three of which were the input data. The other three input nodes were calculated using Chebyshev polynomials of the input nodes. The network had 1 hidden layer with 10 nodes, and the output layer had 3 nodes. This network was trained and tested. Then a lookup table was made for both the forward and inverse directions. The lookup table was made using techniques similar to those in reference 62. These techniques are well known in the industry, and can be purchased off the shelf with a commercial color correction software package. The ΔE^*_{ab} of the LUTs are shown in table 4.

LUT type	Average ΔE^*_{ab} error	Maximum ΔE^*_{ab} error
Forward transform (CIELAB to hue corrected)	0.31	1.79
Inverse transform (hue corrected to CIELAB)	0.29	1.21
Full circle transform	0.06	0.61

Table 4. Transformation error from hue correction LUTs.

The LUTs were used in experiments E2(A) and E2(B) to transform data from uniformly sampled planes in hue corrected space to CIELAB for display on the CRT monitor.

3.2.2 *Color space visualization and parameter tuning*

The neural network model attempted to solve the color space mapping by hypothesizing the coordinates of the uniform color space, then trying to train a simple network to derive the parameters of the model that fit the transform. The flaw in this approach is that with a simple mapping between CIEXYZ and a uniform color space, all three dimensions of each of the coordinates must be very carefully placed in order to be able to model the transform with no extra degrees of freedom. This means that the body of points that comprise the coordinates in the uniform color space must all move together in a systematic way to conform to the parameters of the transformation.

The alternative approach to solving the problem (creation of a simple, invertable mapping between CIEXYZ and a uniform hue color space) is to hypothesize a reasonable transform, then observe the resultant coordinates after applying the model. If certain parameters are altered and the transformed coordinates are visualized in real time, then intuition can be developed about which parameters make sense to change, and in which directions. This was the approach that was used in the creation of the uniform hue color space. This is a viable approach to use to model many color appearance phenomena and attributes, as it gives immediate feedback, and can be applied to arbitrary sets of transformation data (i.e. data that has some perceptual characteristic such as constant hue, and corresponding fundamental color coordinate specification).

3.2.2.1 Model Descriptions

Two major types of color model were approximated with the visualization tool, the CIELAB type model, and the CIELUV type model. These models are fundamentally different in the calculation of the opponent data, so they must be separated. The CIELAB type model is characterized by successive 3x3 transforms, a nonlinear stage, and more 3x3 transforms. The CIELUV type model is characterized by a 3x3 transform, nonlinear compression of the luminance channel, then chromaticity calculation followed by subtraction for adaptation, and multiplication by the lightness channel.

The CIELAB type model general form is shown in equation 5.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = R \times O \times P \times \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix}$$

$$L' = f(L)$$

$$M' = f(M)$$

$$S' = f(S)$$

$$\begin{bmatrix} P \\ D \\ Q \end{bmatrix} = T \times \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix}$$

Equation 5. General form of CIELAB type color model.

Elements of the model R,O,P, and T are all 3x3 matrices with specific goals. They are separated in the first multiplication for purposes of clarity and to separate functions. The function of the P matrix is to provide a transformation from CIEXYZ into a more fundamental cone tristimulus space. Many of these transformations exist, and most are linear and can be performed using a 3x3 matrix. The O matrix is used to support color models where the opponent step (calculation of luminance and chrominance channels from tristimulus values) is performed before the nonlinearity, as in Guth's ATD color model⁶³. The R matrix allows the opponent transformed space to be rotated through θ degrees through the use of the matrix in equation 6.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \text{Cos}(\theta) & \text{Sin}(\theta) \\ 0 & -\text{Sin}(\theta) & \text{Cos}(\theta) \end{bmatrix}$$

Equation 6. Matrix to rotate second and third row elements for the opponent response.

This assumes that the lightness correlate is made to be in the first coordinate, and the opponent channels are in the second and third coordinates. The nonlinear functions f can be in the form of either f_1 , f_2 , or f_3 , and were implemented as a group, and could take the form of chroma type compression (convert the second and third coordinates to polar coordinates, and compress the radius value, then convert back), a power function, or a hyperbolic function. Equations 7 show these functional forms.

$$f_1 = L^{\sigma_1}; L \geq 0$$

$$f_1 = -(-L)^{\sigma_2}; L < 0$$

$$f_2 = \frac{(\sigma_1+1) \cdot L}{\sigma_1 + L}; L \geq 0$$

$$f_2 = \frac{(\sigma_2+1) \cdot L}{\sigma_2 - L}; L < 0$$

$$f_3 = f_2 \left(\sqrt{M^2 + S^2} \right)$$

Equation 7. Functional forms for nonlinear step.

f_1 and f_2 are applied to all channels equally. f_3 is applied by first computing the chroma from the second and third (M and S) coordinates, compressing with a hyperbolic function, then converting back to M and S by taking $C \sin(\theta)$ and $C \cos(\theta)$ where C is the chroma value after transforming through the nonlinear function. Figure 51 shows example plots of functions f_1 and f_2 . It should be noted that none of the nonlinear transforms can actually model CIELAB nonlinearities, since they have the bias term. This was deemed as too complex for the simple

model, and since RLAB has been successful without the added complexity, it was decided to keep it out.

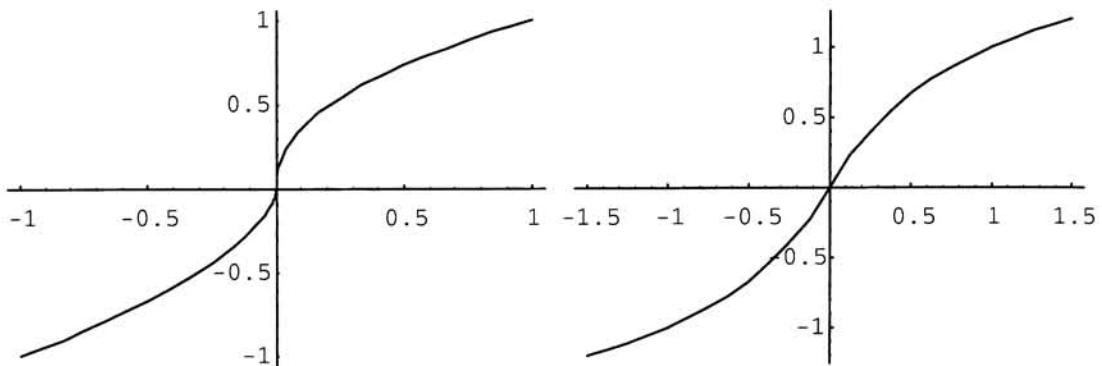


Figure 51. Examples of nonlinear functions. Left plot is function f_1 , with parameters $\sigma_1 = 0.45$, $\sigma_2 = 0.6$. Right plot is f_2 with parameters $\sigma_1 = 1$, $\sigma_2 = 1$. Note that at the origin, f_1 has a slope of ∞ .

This model is made more specific depending on whether the opponent stage is performed after or before the nonlinear compression stage. If the sequence of transforms is like Guth's model, the relevant transform steps are: P O R [nonlinear], thus the tristimulus values can be pretransformed to cone values, then the opponent stage can be calculated, then the system can be rotated (which is not uniform since the nonlinearity follows), and finally the nonlinear stage is applied. One could envision further transforming the output of the nonlinear stage, but no good function was found for this stage, although it is available.

If the sequence of transforms is like the CIELAB model, the relevant transform steps are: P [nonlinear] T, thus the tristimulus values are transformed to primary cone values, then the

nonlinear compression is applied, then the opponent stage is calculated. The opponent, and rotation matrices are not used.

The CIELUV type model general form is shown in equation 8. The constants k_1-k_5 are parameters that can be changed to alter the shape of the resultant space. The P matrix has the same effect as in the CIELAB type model in that it can serve to pretransform the XYZ tristimulus values into a fundamental cone type space. The nonlinearity can be of the form f_1 or f_2 , but not f_3 , as the second 2 elements of the transformed tristimulus values don't correspond to the opponent channel data.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = P \times \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix}$$

$$M' = f(M)$$

$$u' = \frac{k_1 \cdot L}{k_2 \cdot L + k_3 \cdot M + k_4 \cdot S}$$

$$v' = \frac{k_5 \cdot M}{k_2 \cdot L + k_3 \cdot M + k_4 \cdot S}$$

$$u'_n = \frac{k_1}{k_2 + k_3 + k_4}$$

$$v'_n = \frac{k_5}{k_2 + k_3 + k_4}$$

$$u^* = M' \cdot (u' - u'_n)$$

$$v^* = M' \cdot (v' - v'_n)$$

$$L^* = M'$$

Equation 8. General form of CIELUV type color model.

The CIELUV model is somewhat cumbersome from a complexity standpoint since it does not fit cleanly into successive stages of linear and nonlinear transformations. Additionally, it is not a very intuitively pleasing model because of the translation type of adaptation (although in hindsight, this was really not needed, as the adaptation is assumed to have been done previous to the transformation to the uniform color space coordinate system). The inversion of the CIELUV general equation is shown in equation 9.

$$\begin{aligned}
 M' &= L^* \\
 M &= f^{-1}(M') \\
 v' &= \frac{v^*}{M'} + v'_n \\
 u' &= \frac{u^*}{M'} + u'_n \\
 S &= \frac{M(k_1 \cdot k_5 - k_2 \cdot k_5 \cdot u' - k_1 \cdot k_3 \cdot v')}{k_1 \cdot k_4 \cdot v'} \\
 L &= \frac{k_5 \cdot M \cdot u'}{k_1 \cdot v'} \\
 \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix} &= P^{-1} \times \begin{bmatrix} L \\ M \\ S \end{bmatrix}
 \end{aligned}$$

Equation 9. Inverse CIELUV type equation

3.2.2.2 Visualization tool development

The visualization tool served three purposes. Firstly, it allowed colorimetrically accurate viewing of the uniformity of the color space within the gamut of the CRT used. For this to work, both the forward and the inverse models described in equations 5 and 8 were

implemented. Secondly, it allowed visualization of several psychophysical data sets. Various data sets were used in order to ensure that the creation of a uniform hue color space did not adversely affect other attributes of a general purpose uniform color space. Along with visualization of both constant hue data sets, quantitative data were calculated. This gave a more objective way to compare different prospective spaces. Thirdly, the visualization tool used a multitude of controls to alter various parameters of the model. The control of parameters was designed to give intuitive ways to change coefficients, since there is a large number of them in a given model.

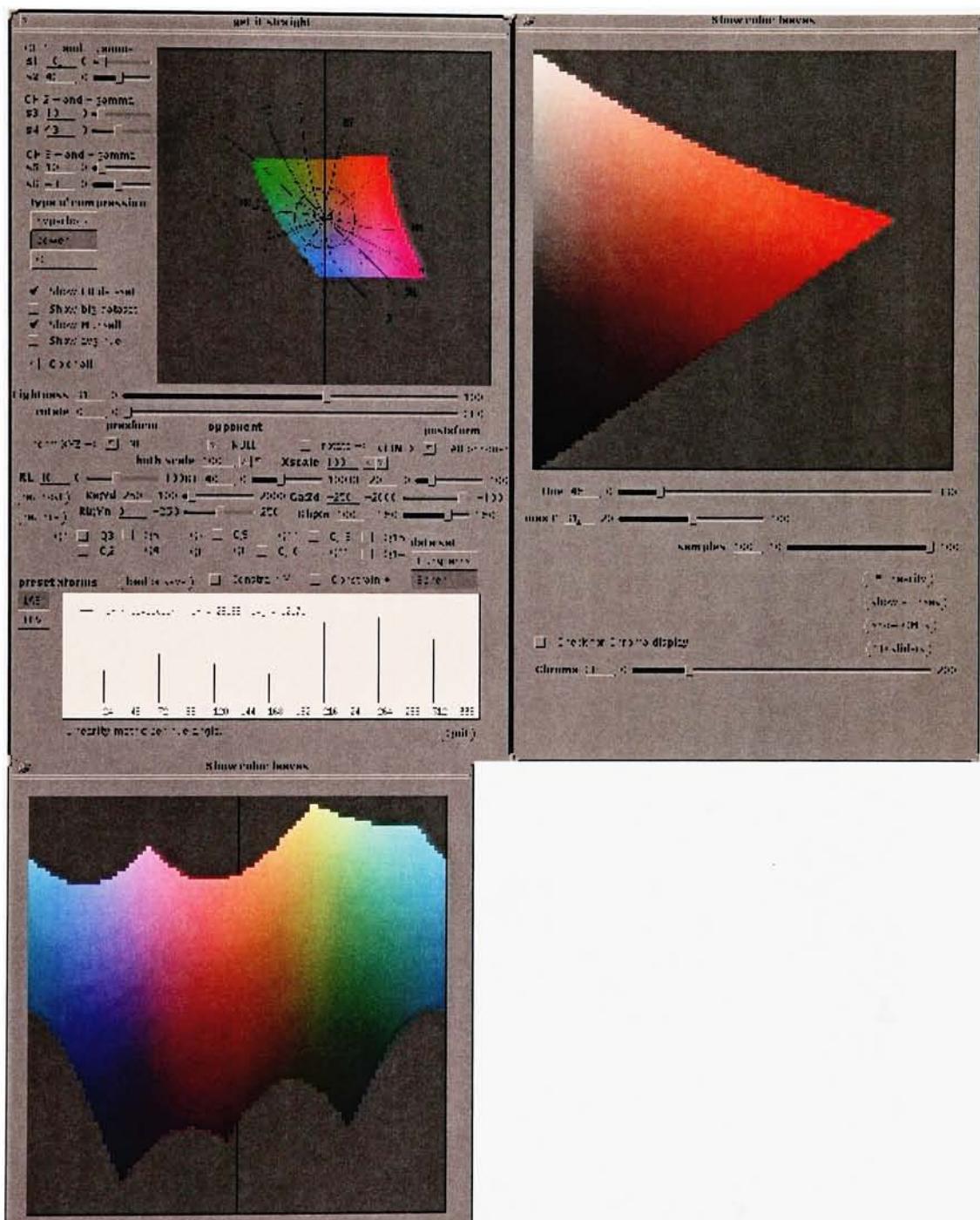
3.2.2.2.1 Visualization of color spaces

The visualization tool user interface offered three separate 2-dimensional views of color spaces: Slices of the opponent channels plane at varying lightness levels, slices of the chroma-lightness plane at varying hue angles, and slices of the hue-lightness cylinder mapped onto a plane at various chroma levels. The display updates everything in real time for any change in any parameter value for the entire model, thus one can watch the space warp as a function of parameter values. Figure 52 shows examples of the three visualization modes of the interface. The constant lightness plane can be seen concurrently with either the constant hue plane, or the constant chroma cylinder. A checkbox is used to control which plane is displayed.

Note, on the constant lightness slice, that there is a circle whose center lies at the origin and a line that projects from the origin at a 45 degree angle. These are indicators that show at which hue and chroma values the other 2 planes cut through the space. There are slider bars to control the lightness level, the chroma value, and the hue angle in increments of 1/100 for

chroma and lightness, and $1/360$ for hue. There is another slider bar that controls the number of samples used to display the images. In figure 51, it is set to the maximum, which is 100 square. This does not update quickly in real time, so the slider bar allows the number to be varied between 100 and 10 samples per display plane. There is also an option to completely turn off the displaying of color planes altogether.

Figures 53 and 54 show close-ups of the control frames. The constant chroma frame is not shown in close-up, because it resides in the same space as the constant hue frame. A checkbox allows the user to toggle between constant hue and constant chroma.



Figures 52. Visualization tool user interface showing the three slices of color dimension, constant lightness, constant hue, and constant chroma.

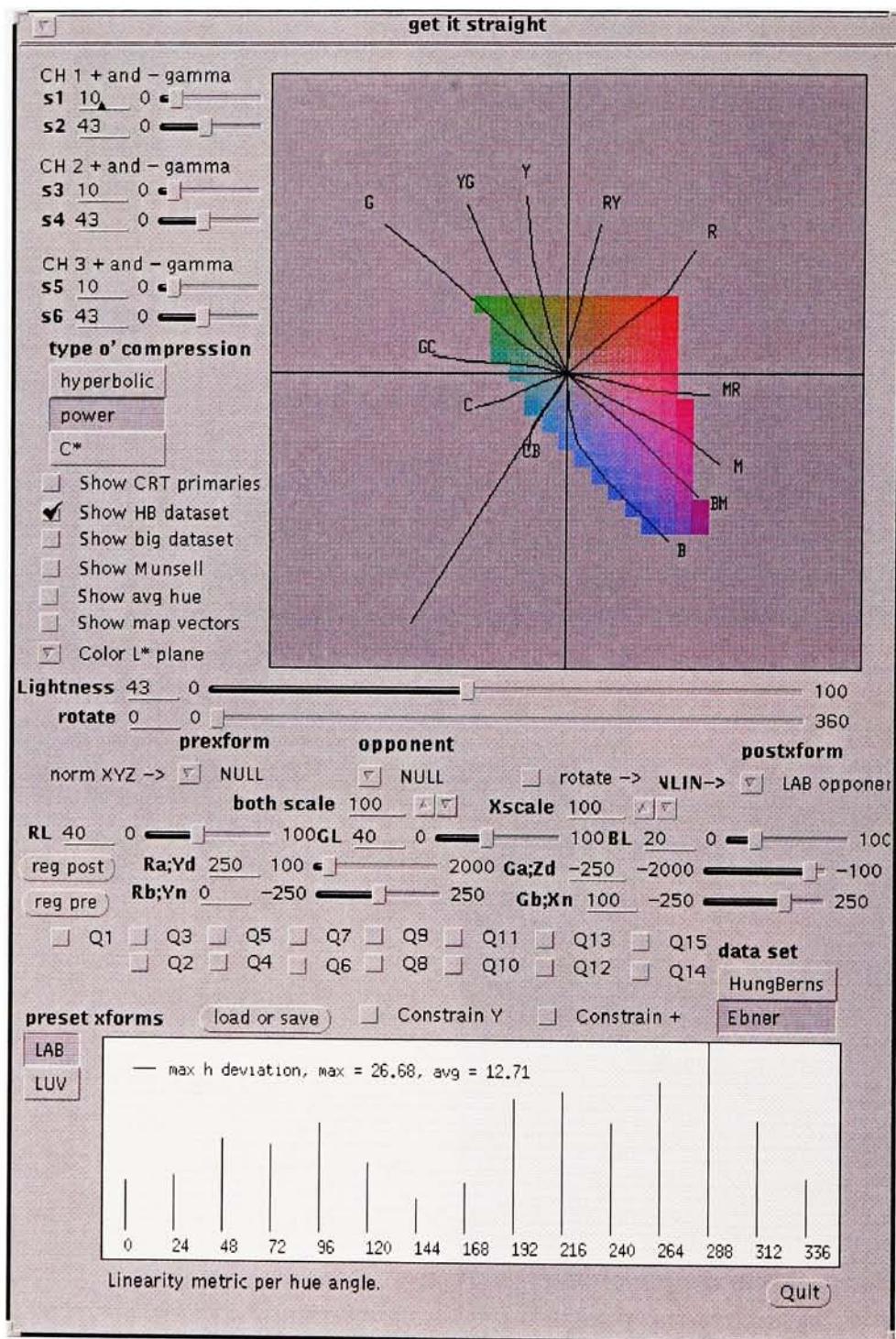


Figure 53. Close up of the controls and constant lightness display.

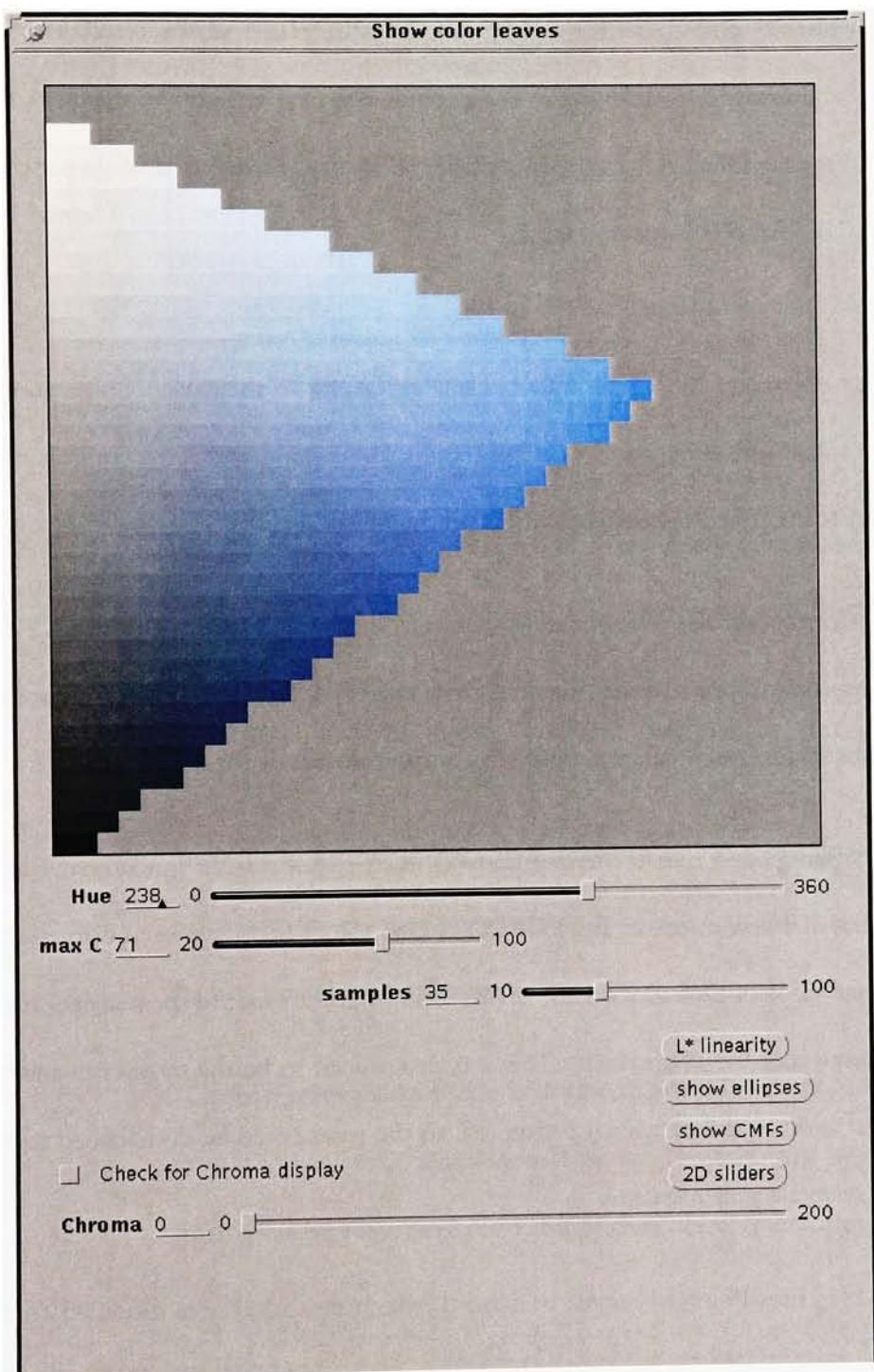


Figure 54. Close-up of the constant hue display.

The color planes displays were created by sampling the planes in the resultant color space, assuming the domain 0;1, -1;1, -1;1. The coordinates that lay on those planes were then transformed through the inverse model to CIELAB and displayed. Out of gamut colors were mapped to CIELAB 50,0,0 (neutral gray).

3.2.2.2 Controls for parameter tuning

Controls for parameter tuning fall into buckets according to their use. Some controls help tune the pretransform, some controls affect the nonlinear steps, and some controls can change the posttransform. Each control is described in the following list:

1. Pretransform controls. There is a list of hardcoded pretransforms that are XYZ to LMS linear approximations to cone primaries from several different authors. Additionally, a set of 2 dimensional slider bars were used to control parameters for this 3x3 matrix.
 - 1.1. The hardcoded transforms are derived from popular cone fundamental references. Each matrix assumes as input CIEXYZ D65 values that have been normalized by the white value of D65 (0.9504 1.0 1.0889) such that the input to the matrices had a range from 0 to 1 for all channels. This was determined to be the most convenient way to deal with the pretransform matrix (P), so the rows could be constrained to sum to 1 to maintain gray response.
 - 1.1.1. Hunt-Pointer-Estevez transform, which was used in a modified form in the final choice for the final color space, was derived from the D65 normalized 3x3 matrix from the RLAB model.⁶⁴ The normalized form of the matrix used in the

model was derived from multiplying the D65 matrix by a matrix whose diagonal elements comprised the D65 white data. Shown in matrix form, this is:

$$\begin{bmatrix} .3804 & .7076 & -0.0879 \\ -0.2151 & 1.1653 & 0.0497 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} = \begin{bmatrix} .4002 & .7076 & -0.0808 \\ -0.2263 & 1.1653 & 0.0457 \\ 0.0 & 0.0 & 0.9182 \end{bmatrix} \times \begin{bmatrix} 0.9504 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0889 \end{bmatrix}$$

The resultant effect on D65 CIEXYZ data is identical. This is simply to show how the visualization tool interpreted the data.

1.1.2. Müller and Judd primaries were found in Wyszecki and Stiles⁶⁵ on page 634.

For this matrix, all rows summed to 5, so the resultant matrix used was simply to divide each element by 5, so the rows summed to 1. This matrix is:

$$\begin{bmatrix} 0.6391 & 0.4896 & -0.128 \\ -0.5091 & 1.4098 & 0.0993 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}. \quad \text{This may not be strictly correct, since the}$$

normalized matrix was not derived from a D65 normalized matrix. However, strictness was not an issue, as these hardcoded matrices were intended to act as starting points for parameter tuning.

1.1.3. Guth primaries were found in Wyszecki and Stiles on page 646. The matrix was treated to have rows that sum to 1 by dividing each row by the sum of the rows. Again, this may not have been strictly the correct transformation to represent the

data properly, but it was a reasonable transform when visualized. This matrix is:

$$\begin{bmatrix} 0.2332 & 0.8162 & -0.0494 \\ -0.4638 & 1.3656 & 0.0982 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

1.1.4. Vos and Walraven primaries were found in Wyszecki and Stiles on page 615.

This matrix was treated the same way as the Guth matrix. This matrix is:

$$\begin{bmatrix} 0.2346 & 0.8213 & -0.0559 \\ -0.4681 & 1.3785 & 0.0896 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

1.1.5. Smith and Pokorny primaries were found in Wyszecki and Stiles on page 615 as

well. Unfortunately, they were incorrectly calculated and seem to come from a mix of Vos and Walraven and the proper Smith and Pokorny data. Nevertheless,

the matrix used was: $\begin{bmatrix} 0.2346 & 0.8213 & -0.0558 \\ -0.4637 & 1.3655 & 0.0982 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}.$

1.2. Two dimensional slider bars were implemented to give fine and intuitive control over the transformation that altered the fundamental LMS space. For each row of the pretransformation matrix, there are two degrees of freedom, since the rows were constrained to sum to 1 to preserve gray response. Each of the slider bars allowed the changing of one of the LMS tristimulus values (after the 3x3 matrix). Thus, with this control one can independently alter the shape of the color matching function of the

“primary” cone response. Figure 55 shows an example picture of the 2 D slider bars. Figure 56 shows an example of the transformed color matching function, first through the identity matrix (thus the result are CIE color matching functions), then through the Hunt pretransform. Very fine control can be exercised over the shape of the transformed curves with the 2 D slider bar controls.

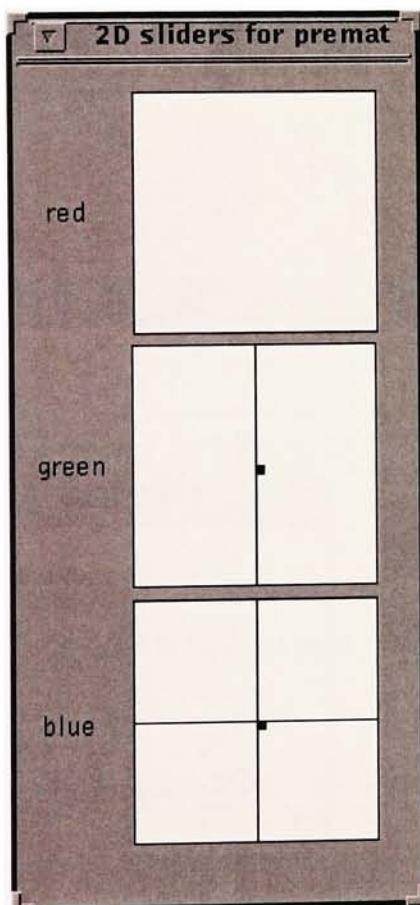


Figure 55. Two dimensional “slider bars” that controlled the pretransformation matrix.

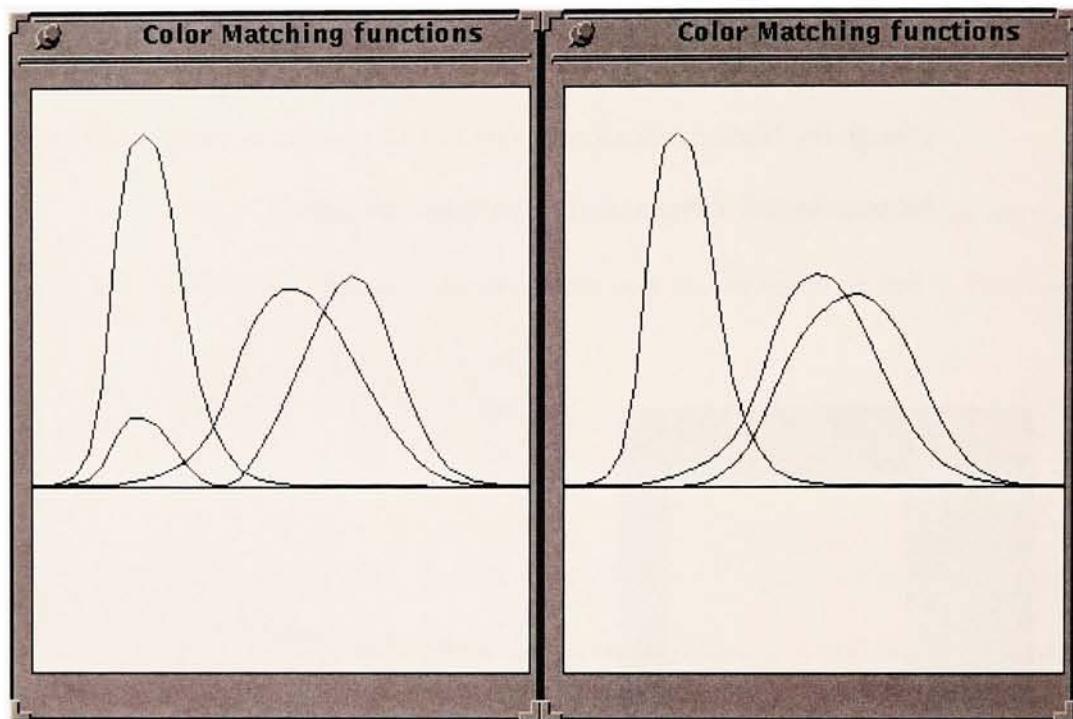


Figure 56. Examples of CIE 1931 color matching functions transformed through pretransformation matrix. Left image is through an identity matrix. The right image is through Hunt-Pointer-Estevez matrix.

The choice of pretransform matrix has a significant effect over the resultant shape of the color space. Take, for example, the following series of figures, together labeled figure 57. Each of the color matching functions is shown next to the resultant color space shape. The only difference in each of the color spaces is the pretransform matrix, which in turn affects the shape of the CMF and color space. Appendix C shows the corresponding parameter set for each of the figures respectively.

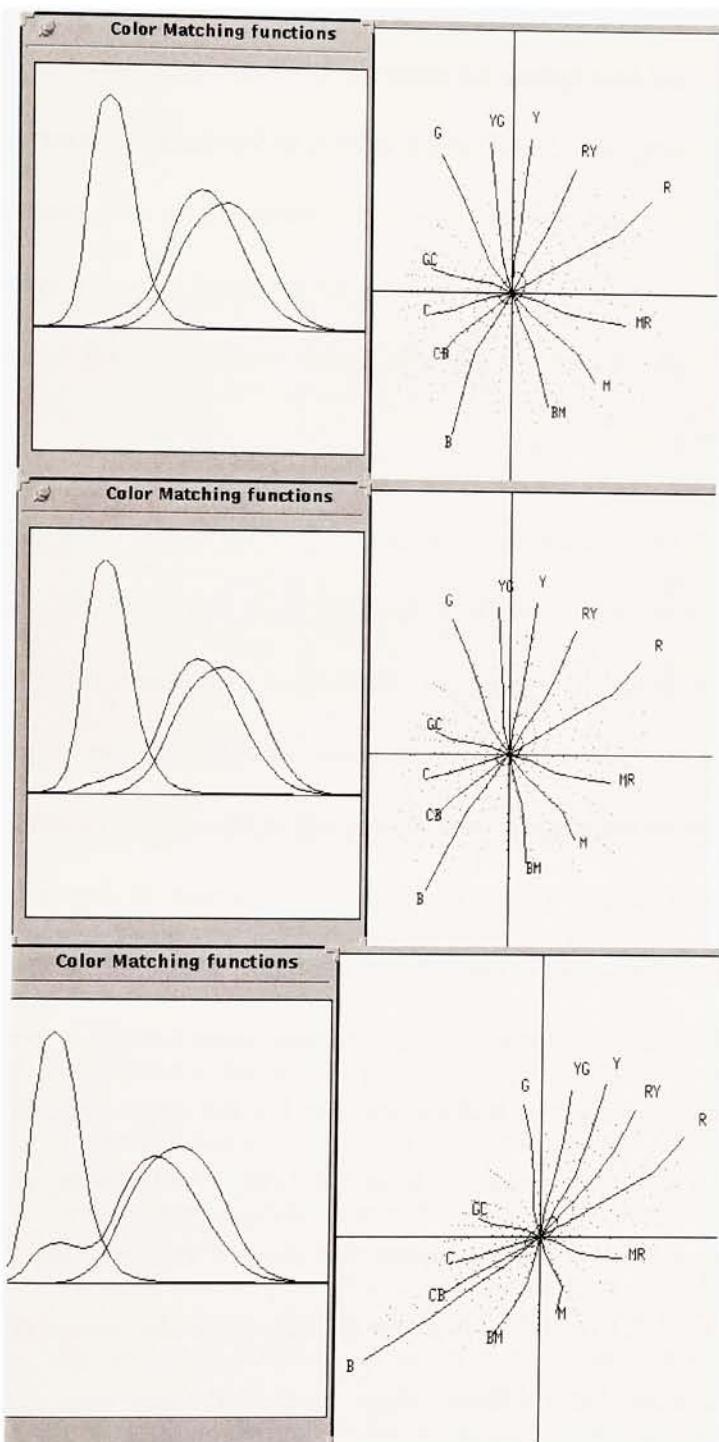


Figure 57. Three examples of CMFs along with their influence on the shape of the color space. Color space images show Hung and Berns CL data set and Munsell V5.

2. The O matrix controlled the opponent step for the Guth type model. This type of model performs the opponent calculation before the nonlinear compression stage. The opponent transform was calculated using only 2 rows of the matrix, as the third was reserved for the lightness response. Since, for opponent response, the corresponding rows must sum to 0 to ensure grays lie at the origin, only 2 elements of each row need to be controlled. This control was offered through the use of 4 slider bars that changed the first 2 elements of each of the opponent rows.
3. The R matrix controlled the rotation step for the Guth type model. Rotation matrix elements are shown in equation 6. One parameter, the angle of rotation, is needed, and was provided for through the use of a slider bar. Rotation of the domain of the Guth type model is not uniform, since the nonlinear compression step follows the opponent and rotation steps. The effect on the shape of the color space in this type of model is to warp hue lines away from the axes and toward the diagonals (axis plus 45 degrees). This is because the slope of the nonlinearity is greatest at the origin. This effect is visualized in figure 58. This type of model was not given much consideration because it worked only with a nonlinear compression stage that had a slope near 1 at the origin (hyperbolic type), thus the lightness response deviate quite far from CIELAB. Additionally, it was very difficult to find rotation and opponent parameters that showed improved hue constancy over current spaces (CIELAB, CIELUV). Note, in the figure, the blue constant hue locus (for Hung and Berns data set) has a different shape on the left figure than on the right.

This is due only to rotating the opponent channels through an angle of 45 degrees. The shape deformation is due to the following nonlinear functions being applied.

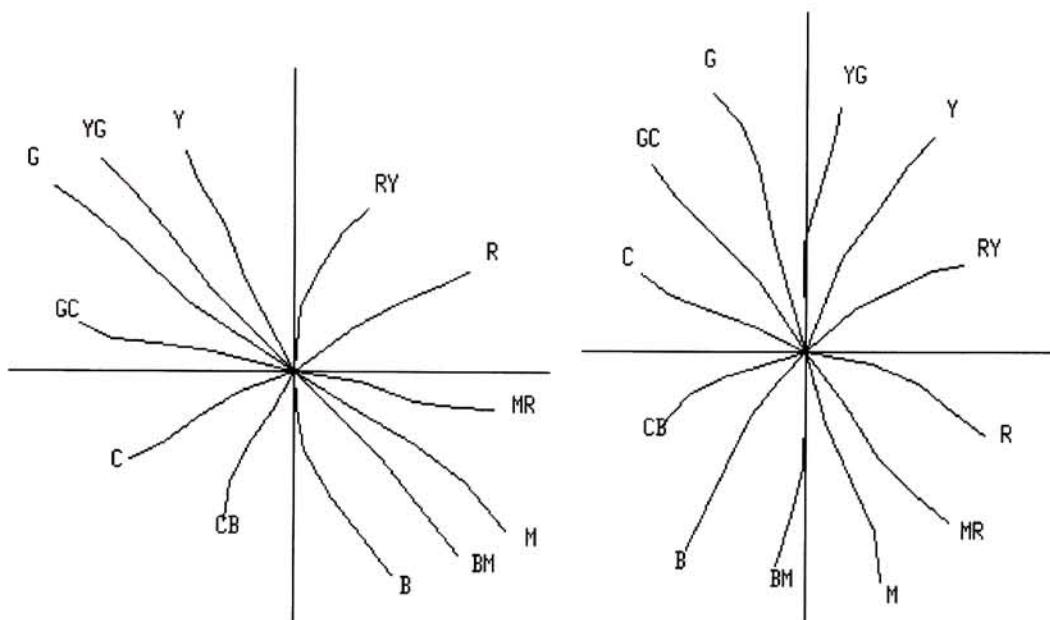


Figure 58. Example hue lines from Guth type model. The first image is rotated at an angle of 0 degrees, the second is at 45 degrees. Note the shape difference of the constant hue loci.

4. Nonlinear controls were presented using a series of 6 slider bars, each controlling the positive and negative parts of the curves respectively. Figure 59 shows an image of the nonlinearity controls. Figure 51 shows an illustration of how the controls can make various curve shapes. For the power function, the parameters would be 45 and 60 (both are divided by 100) and for the hyperbolic function, the parameters would be 10 and 10 (both args are divided by 10 to make a reasonable scale range). The selection range on the

interface was 0 to 100.

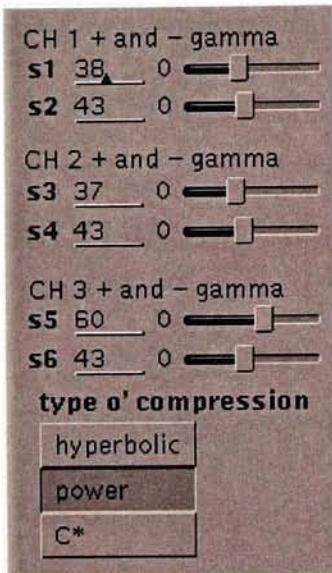


Figure 59. Nonlinear controls for visualization tool. Six parameters correspond to the positive and negative parts of each of the three channels. Functions take the form of hyperbolic, power function, or chroma compression.

5. Posttransform controls included 4 opponent channel controls, and 3 lightness controls.

The opponent channel controls are intended to influence the shape of a CIELAB type model where the opponent calculation is done subsequent to the nonlinear function. Since rows of the opponent channels sum to 0 to make grays lie on the origin, only 2 parameters are necessary. The 3 lightness controls allowed experimentation with how much of the three channels should be used to create the lightness response off the gray axis. By forcing the sum to equal 1, the neutral response can be constrained to be simply a function of the nonlinearity. Thus, the RL^AB response can be exactly mimicked on the neutral axis as

long as the nonlinearity is power function with gamma = 0.43, and rows sum to 1 for the pretransform, and the lightness row sums to 1 in the posttransform.

6. The last control that was offered through the user interface was the ability to apply linear regression to either the pretransform matrix, or the posttransform matrix. The data used was the Ebner and Fairchild data set (306 points of constant hue on 15 different hue surfaces). One could use the interface to include or exclude any of the 15 hue surfaces through checkboxes. Of course at least one surface is needed to provide data to the regression engine. The regression model data needs pairs of triplets to perform the optimization on. The first set of triplets was made by taking the CIEXYZ input data, transforming it to the appropriate coordinate system for input. For the pretransform case, this was simply the CIEXYZ values. For the posttransform case, the data had to be processed through the pretransform and the nonlinearity. The second set of triplets was made by processing the input data through the entire model, then calculating the average hue angle for each hue surface. The average hue angle then replaced the actual hue angle for each color, then the data was processed through the inverse model to the appropriate point. In the pretransform case, the data was transformed through the inverse posttransform and the inverse nonlinearity. In the posttransform case, the data was used directly.

Multiple linear regression was applied to the input-output triplet pairs to generate new 3x3 matrices for the model. This technique worked reasonably well if it was applied only once or twice, but the 3x3 matrices diverged when the regression was applied successive times.

3.2.2.2.3 Quantitative metrics for hue uniformity and psychophysical data set visualization

Several data sets were investigated in the development of the color space. These data sets were used to verify that alterations made to the color space did not adversely affect other attributes of the space. This also makes the decision of which parameters are “best” much more complex, since it is not easy, or even feasible to assign quantitative metrics to all the data, let alone optimize a space with respect to all attributes. Each of the data sets will be discussed here along with the intent and the source. Only the constant hue data sets and the chromatic lightness data set were quantified (Hung and Berns, Ebner and Fairchild, Pirrotta and Fairchild). Quantification of the constant hue data sets took the form of three different metrics. For the visualization tool, the metric used was the most severe: For each set of constant hue data, the absolute difference between the minimum and maximum hue angle was calculated. The other 2 quantifications are holdovers from the calculation method used in the publication of reference 2. These quantifications found the mean and maximum absolute difference (in hue angle) from the mean hue angle for each constant hue set. Thus there are 3 ways to compute a hue uniformity metric.

1. The Ebner and Fairchild² data set was used to derive one of the two uniformity metrics, and the 306 points were displayed on the interface. The data were originally derived in native form as CIELAB D65 data, so they were converted to CIEXYZ D65, then normalized by the white point coordinates (0.9504, 1.0, 1.0889). This data set can be found on the MCSL web site at
<http://www.cis.rit.edu/people/faculty/fairchild/CAM.html>.

2. The Hung and Berns⁵³ constant lightness (CL) data set was used to derive the other uniformity metric. The 48 points were displayed on the interface, an example of which is shown in figure 58. The data were copied out of the paper (table III), then normalized by the white point coordinates for use in the tool. The original data were derived under D65 surround.
3. Munsell Value 5 data were obtained from the RIT VAX computer. The Munsell data was used to scale the x and y axes, and to minimize eccentricity of the chroma “circles” in the destination color space. These are the Munsell renotation data that can be found in Wyszecki and Stiles⁶⁵ on pages 840-852 as well as in the original reference⁶⁶. These data are originally in CIE xyY chromaticity coordinates under Illuminant C. The data were converted to CIEXYZ D65 using the most recent RLAB⁶⁷ equations assuming average surround, and hardcopy viewing ($D=1$, $Y_n=200 \text{ cd/m}^2$, $\sigma=1/2.3$). The data were then normalized by the white point for use in the tool.
4. Lightness response along the neutral axis was chosen to be nearly identical to the RLAB color space, which is very nearly the same response as CIELAB. Lightness response was plotted as a function of CIELAB L* in the visualization tool.
5. Lightness of chromatic object color data were used to make the new color space more accurately represent lightness of colors off the neutral axis. These data were copied from the reference⁶⁸ (Table I). These 36 points were reported in CIELAB L*C*_{ab}h space, with an illuminant near D65. The data were converted to CIEXYZ D65 assuming D65

illuminant. These data quantify the Helmholtz-Kohlrausch effect, which states that chromatic colors appear lighter than neutral colors of the same luminance factor. Since CIELAB uses only the luminance information to derive the lightness response, it cannot, by definition, predict the effect. By combining the responses from other channels, certain aspects of the effect can be reduced, but it was found that this simple model does not predict the Helmholtz-Kohlrausch dependence on chroma, but can reduce the hue dependence that CIELAB exhibits.

6. MacAdam Ellipses were calculated for observer PGN's data. These data are discrimination data that show threshold sensitivity to the dimensions of CIE xyY chromaticity space. In a perfectly uniform color space, the ellipses should resolve to circles, although threshold information is less interesting in a uniform color space than it is in a color difference space. These data are from Wyszecki and Stiles⁶⁵ on page 309. These data are defined in CIE xyY space under illuminant C. The ellipses are defined using major and minor axis lengths, center coordinates (x and y, all data were equiluminant), and rotation in degrees. For each ellipse, 20 samples in CIE xyY of the perimeter were made. A point on an ellipse can be represented by $x = a \cdot \text{Cos}(\phi)$, and, $y = b \cdot \text{Sin}(\phi)$, where a and b are the axes of the ellipse. The point (x',y') can be found by rotating point (x,y) through θ degrees by applying the rotation rule $x' = x \cdot \text{Cos}(\theta) + y \cdot \text{Sin}(\theta)$; $y' = -x \cdot \text{Sin}(\theta) + y \cdot \text{Cos}(\theta)$. The resultant points can be translated to sit around the origin of the ellipse by applying the translation rule $x'' = O_x -$

$x', y'' = O_y - y'$ By sampling the ellipse, then rotating to θ degrees then translating to the origin of the ellipse, samples can be made in the color space. The sampled points are then transformed to CIEXYZ D65. The data were also normalized by the white point for use in the tool.

7. Suprathreshold color difference ellipses⁶⁹ were copied from the reference (Table II, Cross Section for $\Delta L^* = 0$). These data are derived from color differences that are more than just noticeable, so are possibly more important to get closer to circular than are the MacAdam ellipses. These 19 ellipses were sampled in a manner similar to the MacAdam ellipses. Since the data were already in CIELAB using Illuminant D65, they were simply converted to CIEXYZ D65. The data were also normalized by the white point for use in the tool.
8. OSA Color System medium gray ($L = 0$) constant lightness planc scaling data is shown. These data were copied from Wyszecki and Stiles (reference 65), pages 870-871. The data were expressed in CIE xyY (D65), and converted to normalized XYZ for use in the tool.
9. In addition to data sets, the “color matching functions” of the CIE color matching functions are visualized. What are called color matching functions are the CIE color matching functions after being transformed through the first 3x3 matrix. This response function can be thought of as the response of some cone fundamentals that are a 3x3 matrix linear transform away from the CIE 1931 (2⁹) color matching functions.
10. The spectral locus is plotted in the color space for reference.

3.2.2.3 Properties of the chosen color space (IPT)

The IPT color space is named such that its coordinates have some degree of relationship to the meaning of the dimensions. The lightness dimension is denoted as I, which can be loosely related to the word intensity which provides a clue to its meaning. The red-green dimension is denoted as P, which can be related to the fact that it is "dominated" by the red response (protan) and is the dimension lost by protanopes. The yellow-blue dimension is denoted as T, using the same argument for the tritan response. IPT is also short for Image Processing Transform since it is useful for transformations such as gamut mapping.

3.2.2.3.1 Engineering Considerations

The IPT color space was chosen for a number of reasons. The primary reason is that it has a metric hue angle that more closely represents perceived hue. Additionally, it has been designed with some engineering constraints in mind. The biggest engineering constraint is that it should, for the popular CRT gamut, not be discontinuous in hue. That is, it should not, for a given hue angle, for a CRT type gamut, have in-gamut colors at a larger chroma than out-of-gamut colors. This situation can happen if the color space is warped appropriately, and it can happen with a space that accurately models constant perceived hue. It also happens to occur in the blue region of the color space (I think coincidentally). The following scenario illustrates the point.

If one were to create a gradient sweep from neutral gray to device blue, this would result in a straight line in chromaticity space. Consider a typical primary phosphor set with chromaticity coordinates of Red: 0.630 0.340, Green: 0.310 0.595, and Blue: 0.155 0.070 with a white point

of D65. Figure 60 shows a line connecting the blue primary to a point on the neutral axis, lets say a point that is 0.2 multiplied by the white point (near $L^*=50$). If one creates a color gradient sweep from device blue to a neutral gray, or to white, it will be immediately apparent that the middle chroma colors are decidedly purple in color. This is illustrated in figure 61 by plotting the path of the gradient sweep in CIELAB color space along with lines of constant hue. Here, the Hung and Berns loci are used because they are simpler to visualize.

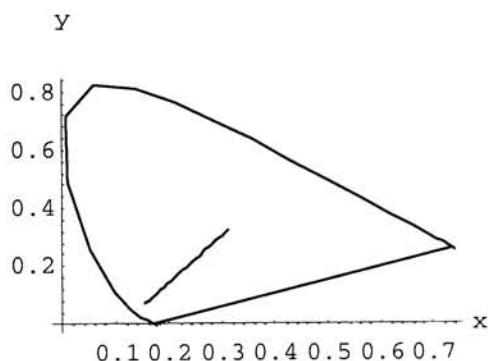


Figure 60. A line showing the path of a gradient sweep from gray to the blue primary (in CIE xyY chromaticity space).

Figure 60 establishes that the path from neutral gray to primary blue using a linear mixing rule (as evidenced by the fact that it makes a straight line in chromaticity space) makes colors that appear purple in the mid-chroma region. The explanation of why this can cause discontinuity in the CRT gamut description in a uniform hue color space is illustrated in figure 62. This shows the path from neutral gray to primary CRT blue in the new IPT color space (to be described in full shortly).

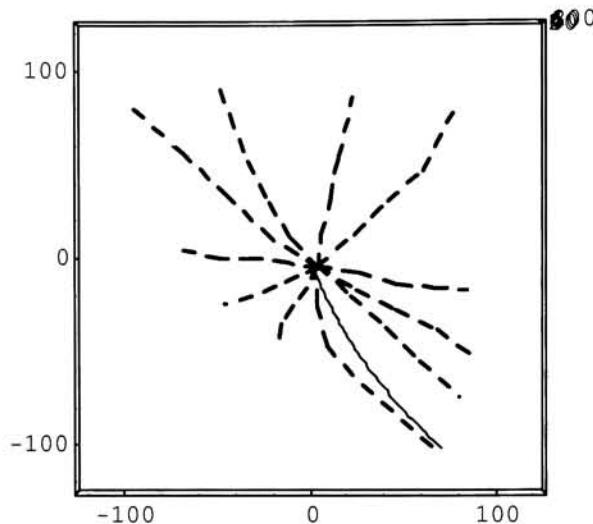


Figure 61. Path of a gradient sweep from gray to blue primary in CIELAB a^* b^* space. Notice that the mid chroma colors are more purple than the constant hue locus. Dotted lines are Hung and Berns constant hue loci.

Note the prominent leftward curvature of this path. Note also, that it accurately describes the perception that the path is more purple in the mid-chroma region, whereas if one would believe that CIELAB was uniform, it would appear that the CRT primary actually becomes *bluer* in the mid-chroma region. From Figure 62, it is not hard to imagine a case where the curvature of the primary blue signal would become so extreme in certain color spaces, that, at a line projected from the origin toward a direction near the primary blue signal, the colors could go out of gamut, then come back into gamut as chroma is increased. This is, in fact, exactly what happens when a color space is designed that exactly straightens the Hung and Berns CL blue locus. Because of the danger of discontinuity of gamut at a constant metric hue

angle, it was decided that the color space should be explicitly designed to not exhibit the artifact.

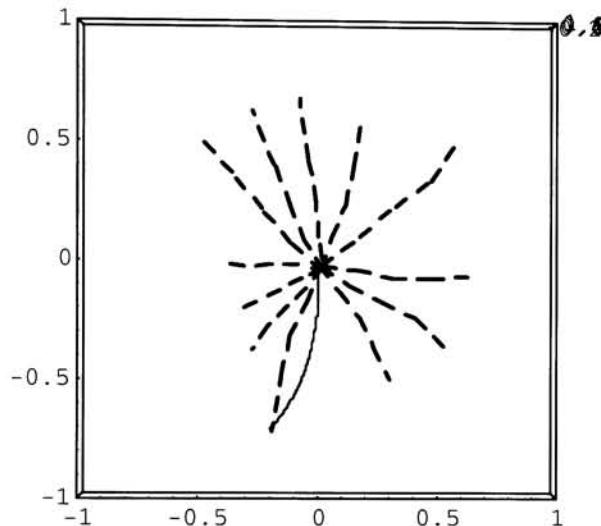


Figure 62. Path of a gradient sweep from gray to blue primary in IPT P-T space. Notice that the mid chroma colors are more purple than the constant hue locus. Dotted lines are Hung and Berns constant hue loci.

Thus, the IPT color space does not exhibit hue discontinuity at specific metric hue angles. It does this at the expense of getting the Hung and Berns constant blue locus exactly straight. However, as will be discussed later, it was judged more uniform than either of the constant hue *data sets*, so the slight bend in blue is not observable.

Other design criteria were that the color space should be as simple as possible and still do a good job. The final model that was decided upon is extremely simple. It is also very amenable to transformation by general transformation engines, such as the PostScript imaging model,

which supports conversion using cascaded sequences of 3x3 matrices and one dimensional non linear steps. The decision to use a power function, as in RLAB, simplifies the model when used in a high accuracy scenario (data not cached into quantized lookup tables), as it needs no conditional statements to treat low luminance colors.

All the following visualizations and quantifications are shown with respect to CIELAB color space. This is because CIELAB is the most popular gamut mapping color space, and the order of complexity between the two spaces is nearly equivalent.

3.2.2.3.2 Model and Coefficients

The IPT forward model is shown in equation 9. Equation 10 shows the inverse model.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.4002 & 0.7075 & -0.0807 \\ -0.2280 & 1.1500 & 0.0612 \\ 0.0 & 0.0 & 0.9184 \end{bmatrix} \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix}$$

$$L' = L^{0.43}; L \geq 0$$

$$L' = -(-L)^{0.43}; L < 0$$

$$M' = M^{0.43}; M \geq 0$$

$$M' = -(-M)^{0.43}; M < 0$$

$$S' = S^{0.43}; S \geq 0$$

$$S' = -(-S)^{0.43}; S < 0$$

$$\begin{bmatrix} I \\ P \\ T \end{bmatrix} = \begin{bmatrix} 0.4000 & 0.4000 & 0.2000 \\ 4.4550 & -4.8510 & 0.3960 \\ 0.8056 & 0.3572 & -1.1628 \end{bmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix}$$

Equation 10. Forward IPT model and coefficients.

$$\begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} = \begin{bmatrix} 1.8501 & -1.1383 & 0.2385 \\ 0.3668 & 0.6439 & -0.0107 \\ 0.0 & 0.0 & 1.0889 \end{bmatrix} \begin{bmatrix} I \\ P \\ T \end{bmatrix}$$

$$L = L'^{2.3256}; L' \geq 0$$

$$L = -(-L')^{2.3256}; L' < 0$$

$$M = M'^{2.3256}; M' \geq 0$$

$$M = -(-M')^{2.3256}; M' < 0$$

$$S = S'^{2.3256}; S' \geq 0$$

$$S = -(-S')^{2.3256}; S' < 0$$

$$\begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.0976 & 0.2052 \\ 1.0 & -0.1139 & 0.1332 \\ 1.0 & 0.0326 & -0.6769 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}$$

Equation 11. IPT inverse model and coefficients.

3.2.2.3.3 Color Matching Functions

What are called color matching functions of the color space are the CIE color matching functions (xbar, ybar, zbar) transformed through the first 3x3 matrix of the model as a function of wavelength. These can be thought of as the fundamental signals that are then compressed and coded into opponent signals. The color matching functions of the IPT model deviate only slightly from the Hunt-Pointer-Estevez primaries⁶⁴ that are described on page 120. Figure 63 shows the IPT color matching functions overlaid on top of the Hunt-Pointer-Estevez primaries. Note the only difference is in the middle wavelength (green) response function. This function has more responsivity at the left tail, and less responsivity at the peak than does the Hunt-Pointer-Estevez function. Note that the resultant yellow-blue response is probably very sensitive to the amount of energy in the tail of the green response because it is

very near the wavelength of the peak of the blue response. This results might be evidence that the functions for transformation to the primary cone response might be improved using an alteration of this nature, as the conversion to a (metric) uniform color space is likely to mimic the human response function to the degree that it is perceptually uniform.

The alteration was made to the pretransform matrix because it noticeably straightened the constant hue loci for the blue region in both constant hue data sets. For example, the Hung and Berns absolute maximum deviation metric for the Hunt primaries (all else remaining constant in the model) was 10.3, while the metric was 7.9 for the IPT color space (altered primaries). Although this may not seem like a large difference from the metric, the visual stimulus clearly contained purple in mid-chroma with max chroma at the blue primary, while with the IPT parameters, no purple was observed (by the author). This result is corroborated through the visual verification experimental results.

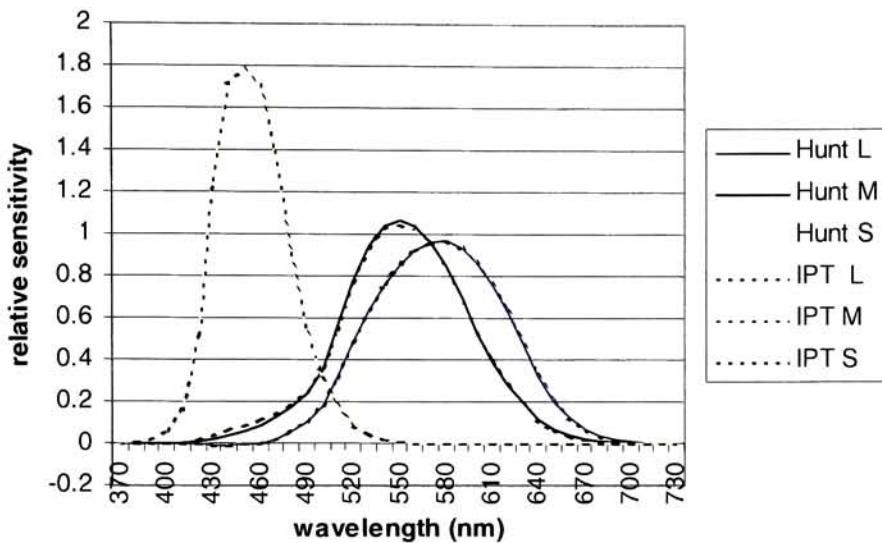


Figure 63. Color matching functions of the IPT color space compared to the Hunt-Pointer-Estevez primaries. Note the larger response on the left tail of the M (green) response function.

3.2.2.3.4 Constant Hue data sets

Figure 64 shows an overlay of both constant hue data sets. The left image shows constant hue data sets in IPT, and the right image shows the surfaces in CIELAB. Hung and Berns constant lightness data is shown as bold dotted lines. The IPT plot appears to have different scale factors for the blue (lower left) Hung and Berns line vs. the Ebner and Fairchild surfaces. This is an artifact of the way the sampling was done in CIELAB space for the constant hue experiment. Notice on the CIELAB plot, that the Hung and Berns line for blue (the most curved line in the lower right quadrant) lies over, but extends much farther than the Ebner and Fairchild surface at the same location. These colors were initially sampled in a more cyan position than the device blue, so were limited by the gamut of the CRT display.

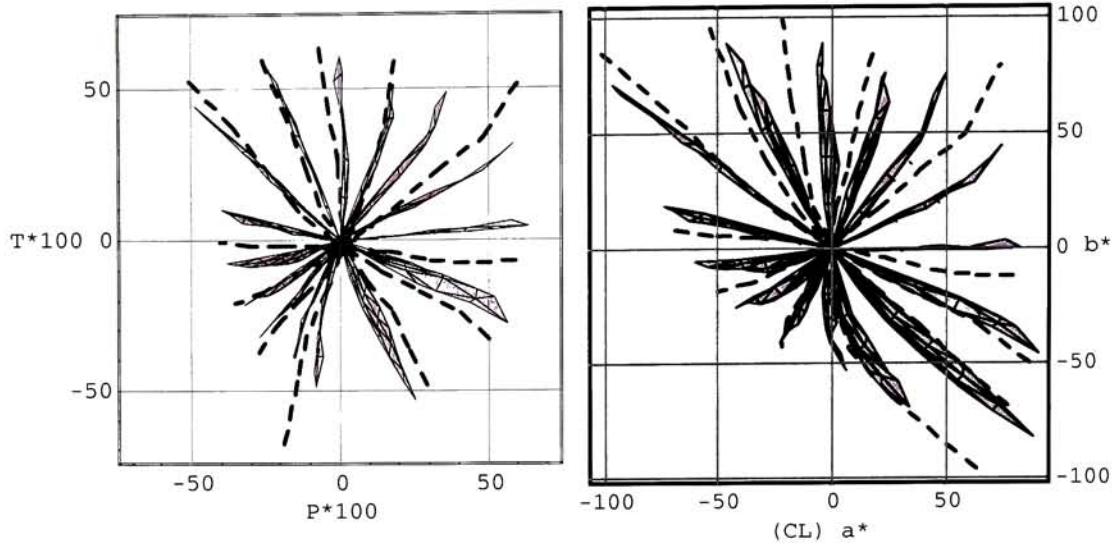


Figure 64. Constant perceived data sets plotted in IPT (left) and CIELAB (right). Hung and Berns⁵³ Constant Lightness (CL) data set is in dotted lines. IPT is plotted with P and T scaled by 100.

Notice that, in the blue region, the hue loci are straighter in the IPT space than in CIELAB. Furthermore, the hue loci in the rest of the space appear to be as straight in IPT as in CIELAB. Quantification is shown for CIELAB, CIECAM97s (just because it was already done, so I didn't remove it), and IPT for the mean distance from the mean hue and the maximum absolute distance from the mean hue in figure 65. Clearly, the mean distance from the mean hue is smaller over the entire space than is either CIELAB, or CIECAM97s. The maximum absolute deviation from the mean hue angle is significantly smaller in IPT than in either CIECAM97s or CIELAB in the blue region. There are 4 reference hue angles where the IPT color space has a larger maximum deviation from the mean hue angle. These reference hue angles are 144 (green), 168 (greenish cyan), 312 (magenta), and 336 (magenta-

red). The absolute magnitude of hue nonuniformity is quite small for 144 and 168, and while the magnitude of the metric is larger at 312 and 336, hue nonuniformity was not observed. Interestingly, these two reference hue angles had the lowest average confidence limits (see figure 35). As will be seen later, IPT was not judged less uniform than either constant hue data set at any of the reference hue angles.

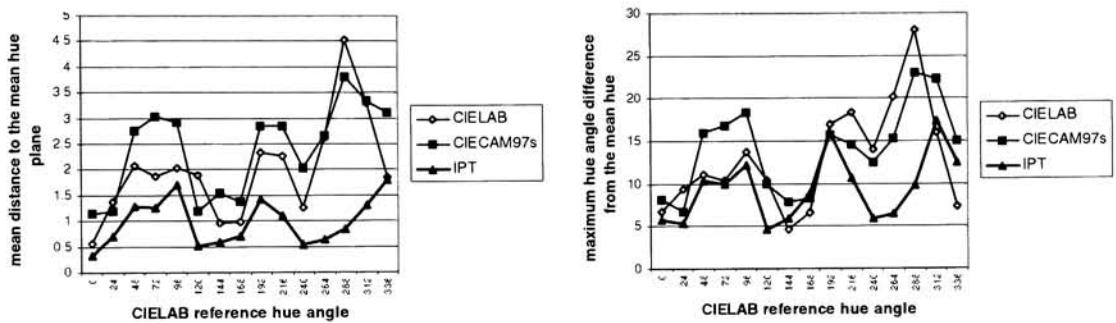


Figure 65. Quantification of hue nonuniformity for CIELAB, CIECAM97s, and IPT.

The visualization tool used a different metric to calculate hue uniformity. It used the most stringent metric, the maximum absolute deviation of hue between all points in a locus. The calculation was done both on Ebner and Fairchild's data set (as in figure 65), and on Hung and Berns (CL) data set. Figure 66 shows those quantifications. Because the visualization tool did not exactly implement the CIELAB model, it is approximated (and labeled CIELABEQ, for CIELAB equivalent) by using a nonlinear power of 0.43 instead of the cube root and conditional statement for low luminance.

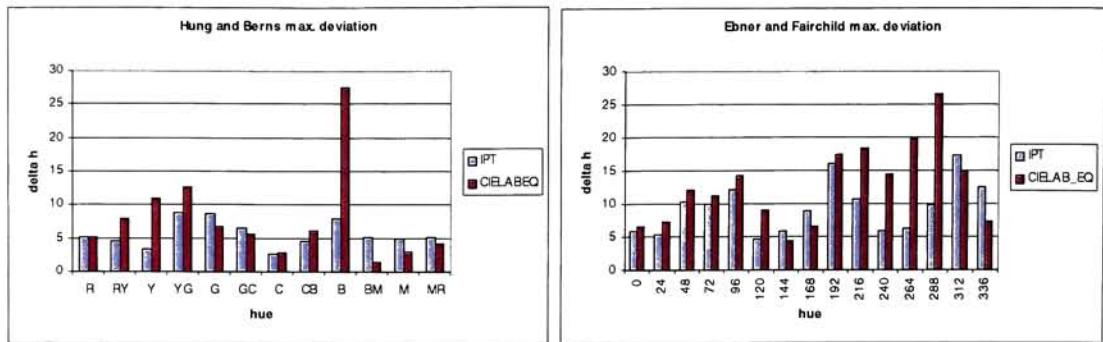


Figure 66. Maximum deviation quantification used in the visualization tool.

This quantification also clearly shows IPT's improved hue uniformity in the blue region, especially with Hung and Berns' (CL) data.

3.2.2.3.5 Munsell Value 5 data

Munsell renotation data is plotted to show that the chroma response of the space is not terribly different than that of CIELAB. No quantification has been done of the eccentricities of the constant Munsell chroma loci. Figure 67 shows the Munsell value 5 data plotted in IPT space on the left and in CIELAB on the right. The CIELAB plot appears to be slightly more uniform in the low chroma circles, but neither appears to be perfect, and neither appears to be grossly bad. This is a possible area of improvement of the color space. However, one would need to be careful to only use the non-extrapolated data to tune the chroma response of the space. Other chroma data sets would be interesting to find as well to corroborate the Munsell data base.

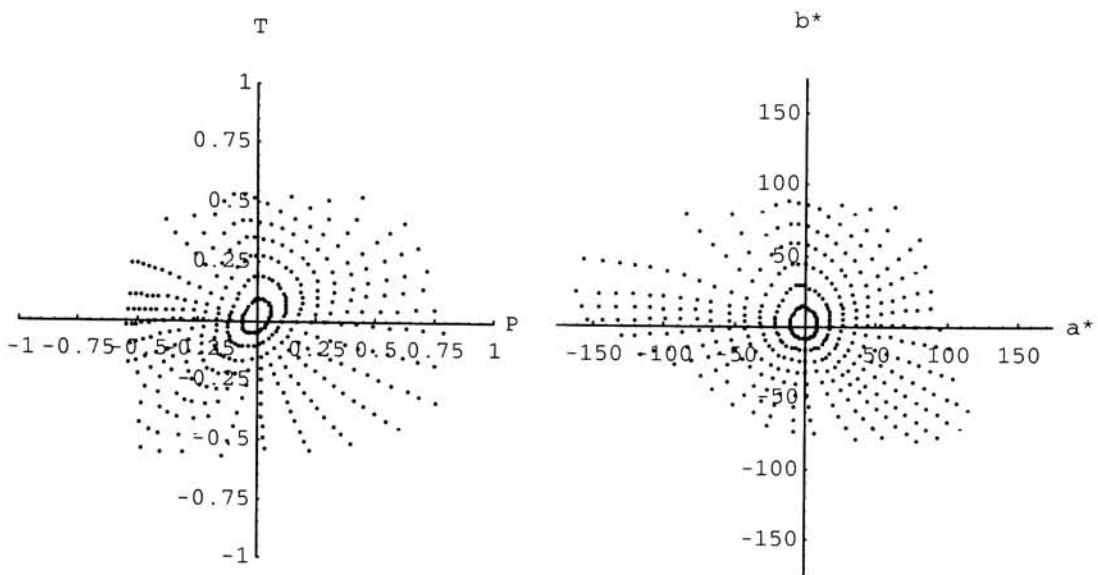


Figure 67. Munsell Value 5 data plotted in IPT (left) and CIELAB (right)

3.2.2.3.6 Neutral Lightness Response

The CIELAB lightness function was used as the comparison metric for judging appropriate lightness response. CIELAB L* is known to be very representative of the lightness response of humans under average surround conditions. The CIELAB L* function was mapped as closely as possible while maintaining a simple power function. The final response exponent, 0.43 is very close to, but not exactly the same as the RLAB response for average surround, which is $1/(2.3)$, or 0.43478. If desired, for coherence, the exponent could be changed to $1/(2.3)$ with virtually no effect on the response of the color space. Figure 68 shows the IPT lightness response as a function of CIELAB L*. Note the major difference in the response is at the dark end of the curve. The difference is due to the fact that the slope of a power function (power between 0 and 1) tends toward infinity as the value goes to 0.

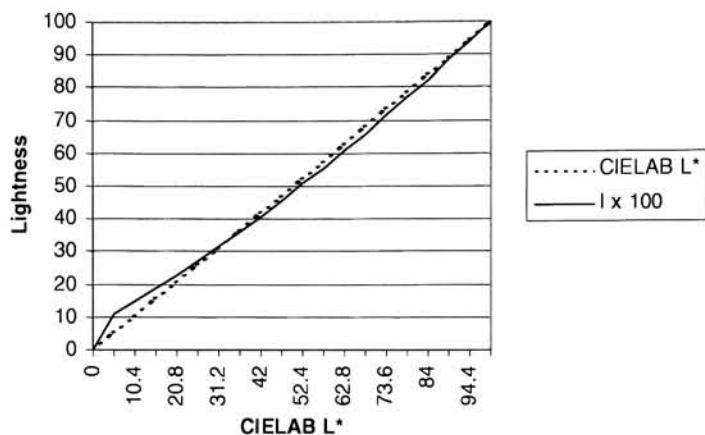


Figure 68. IPT lightness correlate as a function of CIELAB L*

3.2.2.3.7 Chromatic Lightness Response

The chromatic lightness response is modeled on data derived from Fairchild and Pirrotta. The Wyszecki data set referenced in that paper was not investigated, but might be used in further refinements of the model. The observed lightness response of chromatic colors was measured by having observers make lightness matches between chromatic colors and neutral colors. The chromatic lightness response is a measure of how well the color space predicts the lightness of chromatic colors. To show this, a comparison is done between the lightness response (in the color space of interest) of the chromatic colors to the lightness response of the neutral colors that were judged as matches. Thus, in a color space that had perfect chromatic lightness response, the metric lightness of chromatic and neutral colors (which were judged by observers as having equal lightness) would be the same. This is best explained along with an illustration. In figure 69 the CIELAB L* of a neutral color that matched a chromatic stimulus

is plotted as a function of the CIELAB L* of the chromatic stimulus. If CIELAB had a perfect chromatic lightness response, the data would fall on a straight line with a slope of 1.

Neutral L*

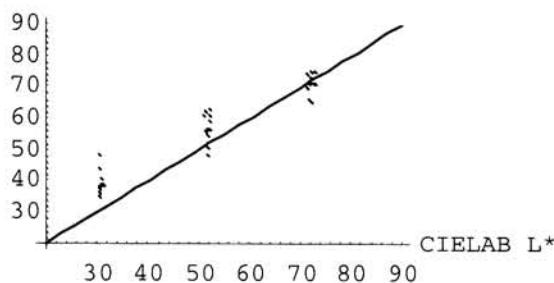


Figure 69. Chromatic lightness response of CIELAB. x axis is L* of chromatic stimulus, y axis is L* of neutral stimulus that matched in lightness (mean observer judgement). RMS error between observed lightness and L* of chromatic color is 7.12.

The RMS error between the observed lightness of gray stimulus and the CIELAB L* of matching lightness is 7.12. When a simple model (which is a function of L*) is used to predict observed lightness, the RMS error can be reduced.

Neutral L*

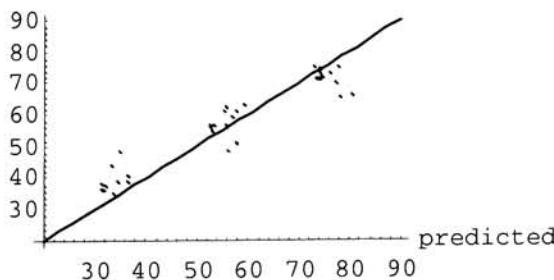


Figure 70. CIELAB matching L* response as a function of predicted L*_{new} = L* + 0.098 C*_{ab}. RMS error between observed lightness and predicted is 5.97.

If a linear regression is applied to the model $L^*_{\text{new}} = L^* + k C^*_{ab}$, the constant k is found to be 0.098. The plot of this simple prediction function is shown in figure 70. Note that the match points are closer to the straight line. The RMS error for this simple prediction is 5.97. By adding enhancements to the predicted Lightness model, the RMS error can be reduced to 4.2. The form of this most elaborate model is denoted L^{**} , and is a function of lightness, chroma, and hue of the chromatic color. Equation 11 shows the entire lightness prediction model and its parameters.

$$L^{**} = L^* + f2(L^*) \cdot f1(h) \cdot C^*$$

$$f1(h) = 0.116 \left| \sin \left(\frac{h - 90}{2} \right) \right| + 0.085$$

$$f2(L^*) = 2.5 - 0.025 \cdot L^*$$

Equation 12. Final lightness predictor for chromatic colors from Fairchild and Pirrotta⁶⁸. RMS error between observed lightness and $L^{**}(L^*, C^*_{ab}, h)$ is 4.2.

The same data is shown plotted in IPT in figure 71. Although the RMS error is only slightly smaller than the RMS error in CIELAB space, notice that the layout of the points falls along a path that is straighter than the CIELAB data points. This allows the creation of a more simple prediction model for lightness of chromatic colors. Specifically, the RMS error can be reduced to below the error of the best CIELAB predictor without the function being dependent on hue.

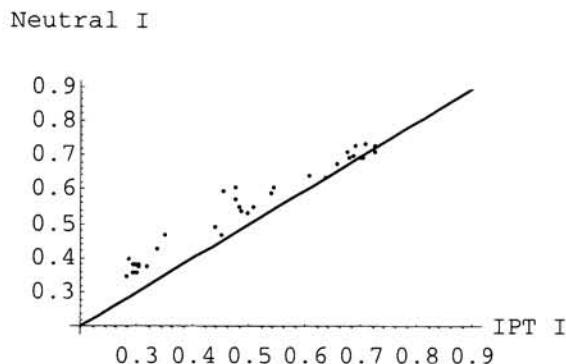


Figure 71. Chromatic lightness response of IPT. x axis is L^* of chromatic stimulus, y axis is L^* of neutral stimulus that matched in lightness (mean observer judgement).). RMS error between observed lightness and L^* of chromatic color is 6.73.

Using the model $I_{\text{new}} = I + k C^*_{\text{IPT}}$, a regression was done to find the value of $k = 0.202$. The RMS error of this predictor is 4.35, which is nearly the same value as the CIELAB predictor that is a function of L^*, C^*_{ab} , and h (RMS error was 4.2). The equivalent of $f_2(L^*)$ was used to enhance the prediction, which is shown in its full form in equation 12.

$$I_{\text{pred}} = I + f(I) \cdot 0.2024 \cdot C^*_{\text{IPT}}$$

$$f(I) = 2.5 - 2.5 \cdot I$$

$$C^*_{\text{IPT}} = \sqrt{P^2 + T^2}$$

Equation 13. Final lightness predictor for IPT color space for chromatic colors from Fairchild and Pirrotta⁶⁸. RMS error between observed lightness and predicted is 3.5.

Figure 72 shows the plot of the final model of a lightness predictor for chromatic colors in IPT. Note that the RMS error is lower, and the model is simpler because it has no hue angle dependency.

Neutral I

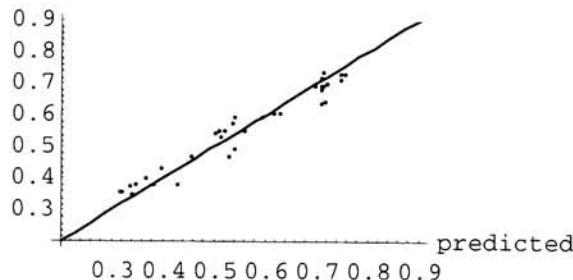


Figure 72. Plot of final lightness predictor for IPT color space for chromatic colors. RMS error between observed lightness and predicted is 3.5.

3.2.2.3.8 MacAdam (PGN) Discrimination Ellipses and Spectral Locus

Although the intention of a uniform color space is to accurately model appearance attributes (not color differences), it is nonetheless interesting to see how small color differences are mapped in the color space. Therefore, both discrimination (at or near color difference threshold) and suprathreshold color difference data are investigated. Figure 73 shows the MacAdam ellipses plotted on IPT space (left), and CIELAB space (right). Along with the MacAdam ellipses, the spectral locus, monochromatic equienergy signals, are plotted (assuming D65 adaptation).

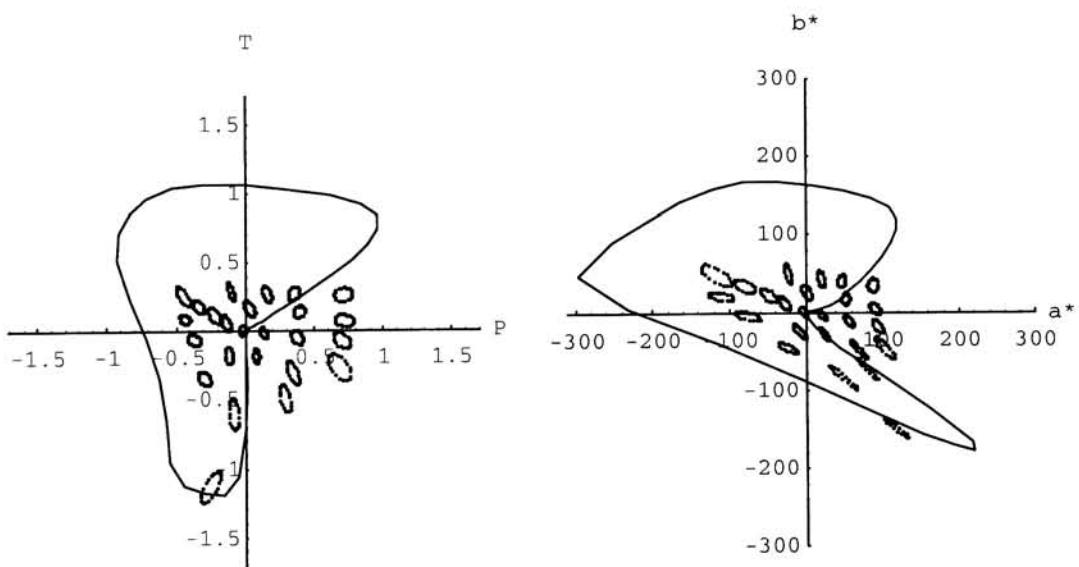


Figure 73. MacAdam ellipses plotted in IPT (left) and CIELAB (right) along with the spectral locus.

Note that the IPT color space, especially for the blue region, models the ellipses with less difference between major and minor axes. The spectral locus for both of the color spaces

approaches chroma of 0 as the stimuli reach the edge of the spectral sensitivity of the color matching functions.

3.2.2.3.9 Suprathreshold Color Difference Ellipses

Supra threshold color difference ellipses are shown plotted in IPT and CIELAB in figure 74. These color ellipses model a color difference on the order of 1 delta E^*_{ab} unit. Because the unit difference of the reference was near gray, the color difference was definitely above threshold, but is still considered a small color difference. Notice that these plots exhibit a similar behavior in the blue region as did the MacAdam ellipses. Most notably, the eccentricity of the ellipses in the blue region is reduced in IPT space.

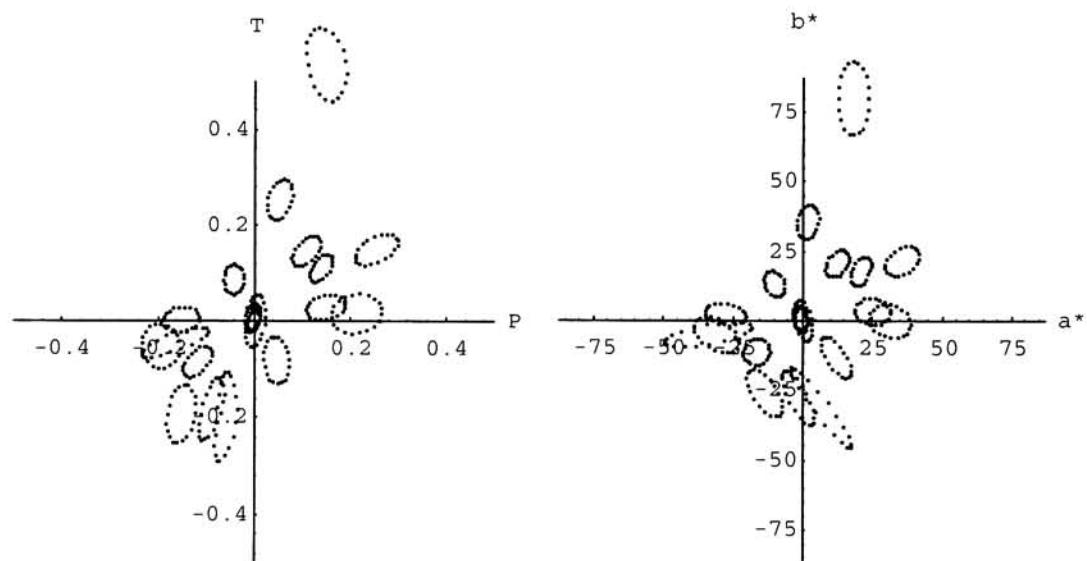


Figure 74. RIT-DuPont visual color-difference ellipses plotted in IPT (left) and CIELAB (right).

3.2.2.3.10 OSA Color System medium gray ($L = 0$) constant lightness plane

Constant lightness plane from the OSA color system is shown in figure 75. Data are plotted in IPT on the left and CIELAB on the right. This data set is included in addition to Munsell V5 data because the scaling was done on a rectangular grid, thus all colors in the plane are equally spaced. These data were scaled using colored paint chips. The color difference between any two colors is more than 20 times just noticeable.

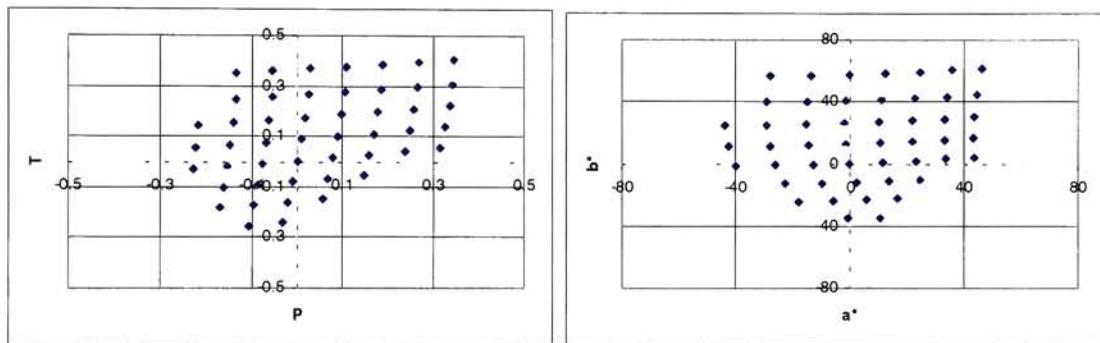


Figure 75. OSA color system medium gray constant lightness plane plotted in IPT (left) and CIELAB (right).

Note that in CIELAB, the vertical loci data are curved dramatically, whereas in IPT, though the data are not exactly evenly spaced, the uniformity is much improved. It appears that the curvature is slightly reversed in IPT with respect to CIELAB. This may suggest that the color space is slightly overcompensating for the blue constant hue locus for object colors. This may point to a fundamental difference in perception between object colors and self luminous (CRT displayed) colors.

3.3 Experiments E2. Verification Experiments

The constant hue data sets and the IPT color space were tested to measure the difference in uniformity when entire hue leaves were viewed. This task is fundamentally different from the task that was performed in deriving either of the constant hue data sets in that there was a large amount of color context (the entire CRT gamut at the “same” hue) to judge uniformity within. Both experiments A and B were designed to be identical in procedure. The only difference between the experiments was that the first one compared CIELAB, hue correction based on Hung and Berns, and hue correction based on Ebner and Fairchild, and the second experiment compared the IPT color space to the 2 constant hue data sets.

3.3.1 Comparing CIELAB to Hung & Berns and Ebner data sets

3.3.1.1 Introduction

A paired-comparison experiment was conducted to test the hue uniformity difference between two uniform perceived hue data sets and CIELAB color space. A colorimetrically characterized and calibrated CRT display was used to present the stimuli. Viewing conditions were set to dark surround, near D65 white point, with a maximum luminance of $70\text{cd}/\text{m}^2$.

Hung and Berns’ constant lightness data set was used by Braun⁷⁰ to derive a two dimensional table lookup scheme that enabled forward and inverse transformation between CIELAB color space and the Hung and Berns uniform hue data space. Ebner and Fairchild’s data set of constant perceived hue was used by the author to derive a three dimensional table to enable transformation between CIELAB color space and the respective uniform hue data space.

3.3.1.2 Experimental

Fifteen hue angles were uniformly sampled in lightness and chroma within the respective color spaces, then transformed to CIELAB for display on the calibrated CRT display. Out of gamut colors were converted to CIELAB coordinates 50,0,0. For each hue angle sampled, the three color spaces were compared. For each hue angle, three presentations were required to compare each color space to the other ($N^*(N-1)/2$). For each observation, two repeats of the data set were presented to the user to account for screen non-uniformity. Each observation session required 90 judgements (15 hues X 3 Images X 2 repeats). The list of pairs of hue comparisons was pseudo-randomly shuffled for each observation session. The user interface used to present the stimuli is shown in the figure 76.

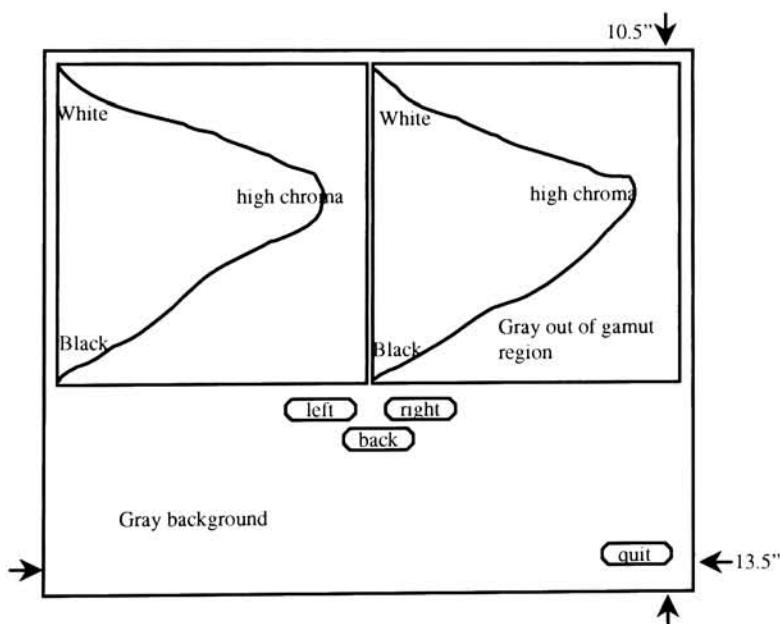


Figure 76. User interface for verification experiment.

The task was as follows:

"You will be shown pairs of images. For each pair of images shown, pick the image that has the best hue uniformity. The maximum chroma color for each image is the same color. Compare the colors in each image separately, don't compare colors between images."

Only same base hue images were compared to each other, never were a mix of hues presented to the observer.

Thirty observations of the entire data set were made. Nine observers took part in the experiment. All observers had experience with color, and were familiar with the terminology, and the concept of hue uniformity. Using Thurstone's law of comparative judgement⁷¹ (case V), fifteen interval scales were derived, one for each hue sampled, that compared relative hue uniformity of the three color spaces. Confidence limits were calculated based on the number of observations of the entire data set. The confidence limits were calculated using the

following equation: $\pm \frac{1.96 \cdot \sqrt{2}}{\sqrt{N}}$, where N = 30, the number of observations (not number of

observers). This formula is based on the assumption that each scale value is a mean of a normal distribution where the units of the scale are $\sqrt{2}$ times the standard deviation of the distribution. Thus, a 95% confidence interval is calculated from ± 1.96 times the standard

error of the sample mean, which is 1.96 times $\frac{S}{\sqrt{N}}$, where S is arbitrarily assigned a value of 1,

as is done in Thurstone's case V. This seems inconsistent with the calculation of the 95% confidence limit by a factor of $\text{sqrt}(2)$.

3.3.1.3 Results

Figure 77 shows the results from the experiment. Higher bars mean that the color space was judged to be more uniform.

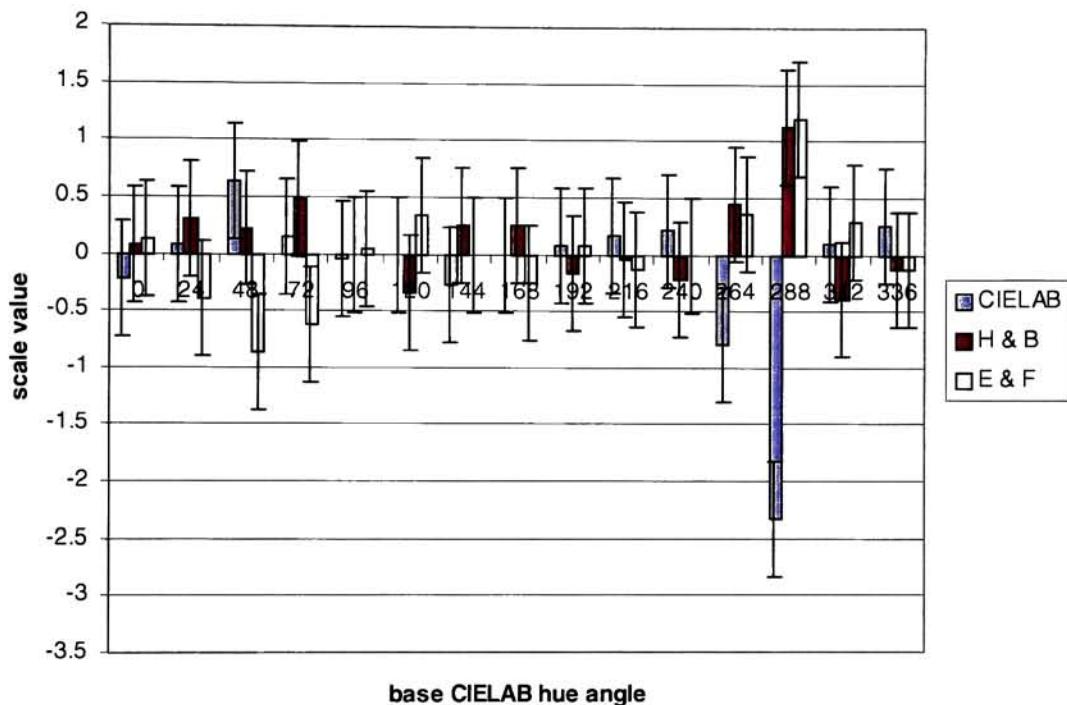


Figure 77. Scale values for judged uniformity between CIELAB and the 2 constant hue data sets. H&B are Hung and Berns data, E&F are Ebner and Fairchild's data.

The most prominent feature of this figure is the significant difference in judged uniformity at the 264 and 288 degrees reference hue angles (these represent the blue region). Hung and Berns and Ebner and Fairchild data sets were judged to be not significantly different in the blue colors. Most of the other reference hue angles seem to have very close scale values,

which leads one to believe that they are equally uniform at those hue angles. There is an interesting trend observed at the reference hues 24, 48, and 72, where the Ebner and Fairchild hue surfaces are judged to be less uniform than both CIELAB and the Hung and Berns hue correction. This trend was also observed in the gamut mapping experiment conducted by Braun⁷⁰, where the red color of the parrot's head was mapped to a color that was too magenta when using correction based on either constant hue data set. The results from Braun's experiment led him to make a new hue correction LUT that only affected the blue region of CIELAB space. This solved an immediate problem, but is much less general than a functional color space approach to the constant hue problem. The root of the judged non-uniformity of is not well understood. While the Hung and Berns results are not, from this result, judged to be statistically less uniform, the shape of the E&F curve (at 48 degrees) and the shape of the red stimulus of H&B have the same trend (see figure 48, left plot). This bend in both loci of the red stimulus may be more apparent in spot colors, and there may be preference influences. Regardless of the cause, it is a systematic and repeatable effect.

3.3.2 Comparing IPT color space to Hung & Berns and Ebner data sets

The experiment described in E2, part A was repeated using hue planes from the new IPT color space in place of the CIELAB color space. The identical user interface, script, data preparation, and data analysis was performed. The number of observers for this experiment was 10, and 30 observations were made over the entire data set. Confidence limits were calculated the same way as in experiment E2 A. which resulted in 95% values that were the

same (since the number of observations was the same). Figure 78 shows the results for the mean uniformity judgements.

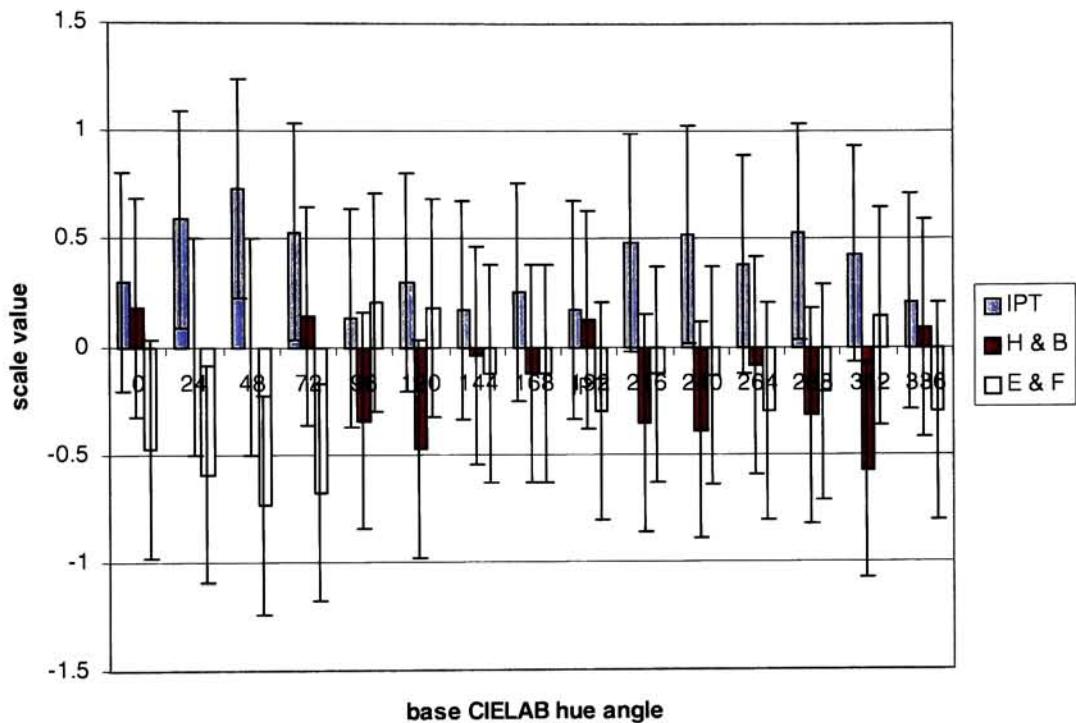


Figure 78. Scale values for judged uniformity between IPT and the 2 constant hue data sets. H&B are Hung and Berns data, E&F are Ebner and Fairchild's data.

The first characteristic to notice about this plot is that the extent of the scale values is of the same order as that of figure 77, if the blue region is discounted. The next thing to notice is that the scale values for IPT color space are uniformly greater than either of the constant hue data set scale values (although not all significantly different statistically).

Clearly, the IPT color space is judged either more uniform (when the confidence limits are beyond the mean of the other two space, such as with reference hue 48), or no less uniform (e.g. reference hue 144) than the constant hue data sets. Experiment E2 A. has shown that for the reference hue of 48 degrees, CIELAB was judged more uniform than either of the constant hue data sets. The magnitude of uniformity difference between CIELAB and the next judgement was about 0.5 units, which is contrasted to roughly 0.75 units in figure 78 (for reference hue angle 48). From this we can conclude that the IPT space is roughly as uniform as CIELAB at the reference hue of 48 degrees, and more uniform than CIELAB everywhere else. Additionally, the IPT color space is judged more uniform than the constant hue data sets from which it was derived.

The last statement is a bit confusing. That a derived color space is more uniform than data sets that were derived to be perceptually uniform seems to be wrong. But one must consider that in making lookup tables that model the constant hue color space, and transform between CIELAB and the constant hue data, some error will be introduced. Even when the data used to create the transformations are very accurately modeled in the LUT transformations (which it is in both cases), there will be uncertainty in accuracy of the transformation in other parts of the space. It seems that it is always better to have an analytical description of a transformation than to use lookup table based approaches because of the above issues (noise and accuracy between sample points). Additionally, perceptually derived data bases are always calculated as a central tendency of a dispersion of individual results. The average response should reflect the nature of the tendency, but there are many ways to calculate a central tendency, none of

which will be “accurate” all the time (accurate is in quotes because its judgement relies on individual’s responses which are not the same as the mean in almost all cases).

4 Conclusions

This work began with an in depth literature search that describes aspects of color such as color order systems, color harmony, color names, color meaning, and preferred color reproduction of pictorial images. From the literature search, a set of experiments was defined that sought to find differences in gamut mapping decisions based on color context. Several contexts were defined that represented a taxonomy of graphical (non-pictorial) image types. Experiments were conducted that sought to determine the influence of image context (in business graphics type images), and task description on gamut mapping decisions. Additional color meaning experiments were conducted to try to find dimensions of meaning of color in the context of business graphics.

A general color gamut mapping model was derived from the above work that has many applications. A paper describing this and the gamut mapping experiment was published in *Color Research and Application*.

Based on the above work and other's experiments, it was decided that gamut mapping in CIELAB was not sufficiently accurate with hue, and there were not sufficient data in the community to generate a better color model. Therefore, an experiment was conducted that found 306 points that spanned the gamut of a CRT display that found 15 surfaces of constant hue in color space.

Based on the above experimental results, a general tool for visualization of color space transformations in real time was developed. The techniques used for visualization and the controls that altered the parameters of color models were described.

A new model of color space named IPT was formulated from the use of the visualization tool. Many popular color appearance data sets were used in the development of the new color space. The IPT color space is a simple transformation from CIEXYZ D65, average surround conditions (reference viewing conditions). The model consists of a 3x3 matrix transform, a nonlinear power function, and another 3x3 matrix transform. This model is a generalization of the CIELAB color space that takes advantage of a transformation to cone fundamental tristimulus space before signal compression, then a linear transform to create the opponent and lightness channels.

The constant hue data sets were judged against the CIELAB color space, and the IPT color space in 2 subsequent experiments. The CIELAB color space was judged much worse in blue and slightly better in red than the constant hue data sets, while the IPT color space was judged superior in hue uniformity to both constant hue data sets at all hue angles tested.

The IPT color space is the culmination of all the above work, either directly, or indirectly. The IPT color space has many appearance attributes that are superior to the CIELAB color space which it was compared against, most notably in the blue region, and in the dependence on hue angle for lightness prediction of chromatic colors.

There are several areas where the IPT color space can be utilized. The color space is both easy to implement for Image Processing Transformations (IPT's), and is a much better predictor of color appearance attributes than is CIELAB, and it is much more uniform in hue than any of CIELAB, CIELUV, or CIECAM97s.

5 Future Work

Additional work needs to be done in two areas. Firstly, further verification must be done to test whether the IPT color space is the preferred color space to use for gamut mapping applications. This can only be done by testing the space under realistic gamut mapping conditions. This remaining work may conclude that there are other similar, but better color spaces to gamut map within, and much of this may depend heavily on observer preference as well as psychometric uniformity. The second area of work is involved with refinement of the color space to test and create more perceptually uniform representation of human color vision. Specifically, the relationship of rate of change of perceived hue as a function of hue angle should be tested to give better data with which to model hue uniformity. These goals are complementary, but need to be approached from different directions.

6 Appendices

6.1 Appendix A: Data plots from experiment P1.

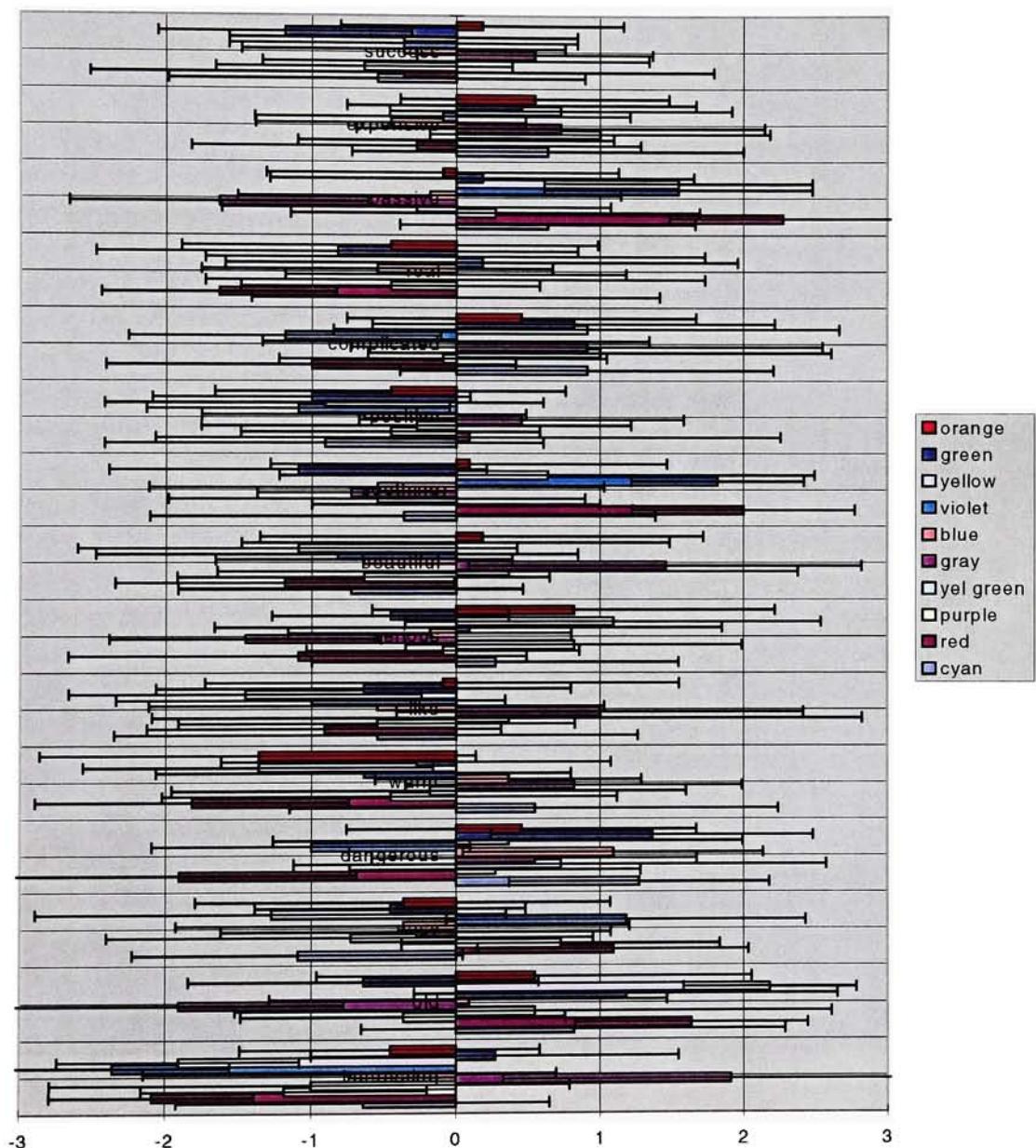


Figure 79. Mean scale values for colors and descriptive terms for abstract representative image type.

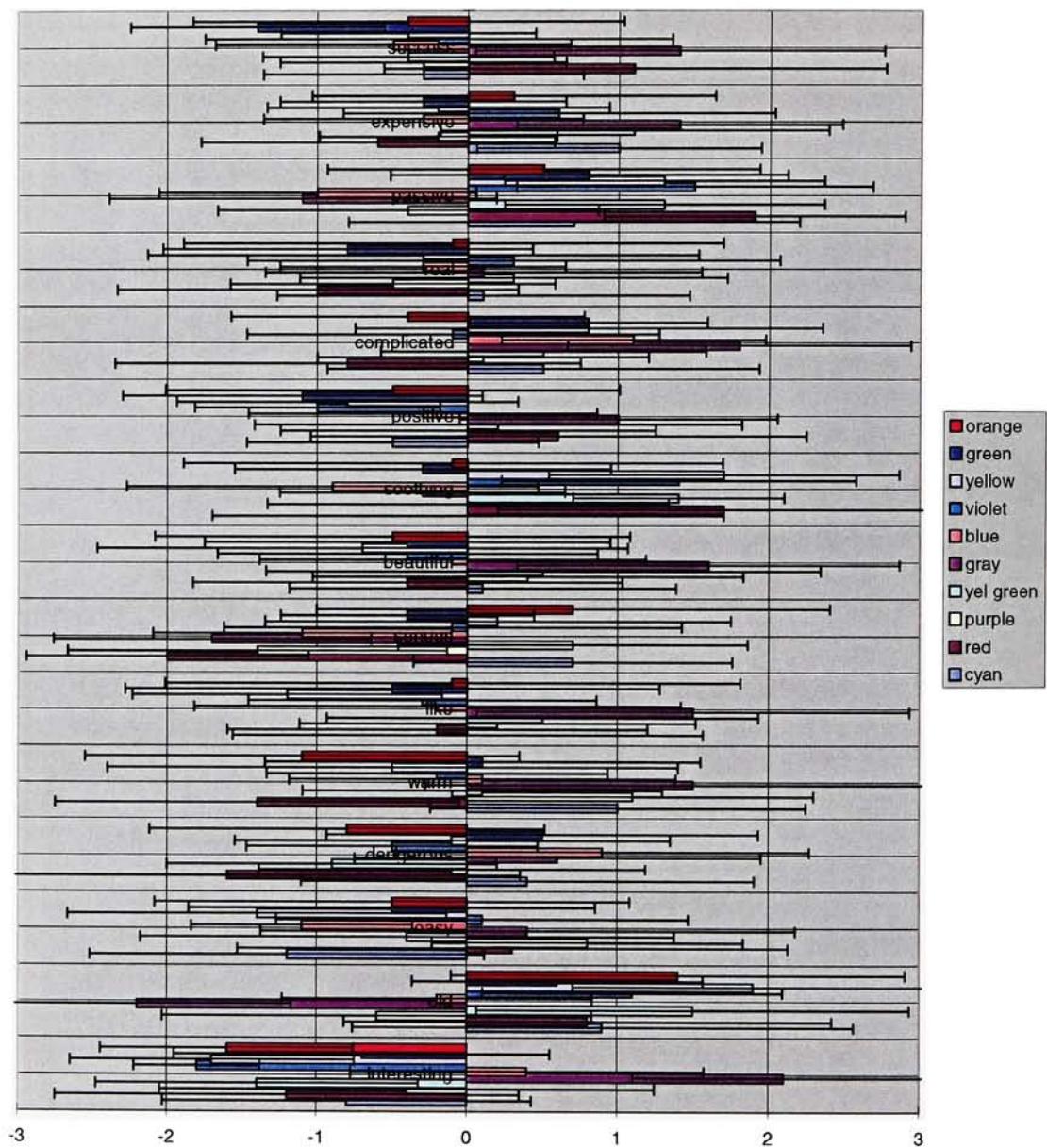


Figure 80. Mean scale values for colors and descriptive terms for data representative image type.

6.2 Appendix B: Data from experiment P2.

Table 5. Luminance and chromaticities of viewing setup.

	x	y	Y(cd/m ²)
D65 white point	.313	.329	1
Luminance of Halon under source	.320	.333	90.6
Luminance of Paper under source	.320	.333	80.3
Background gray matte under source	.307	.321	13.3
CRT Workspace background	.306	.322	
CRT Window background	.305	.322	
Root color	.306	.322	
CRT calibrated white (RGB)	255	240	220
CRT Center calibrated white lights on	.305	.321	79.1
CRT Left calibrated white lights on	.301	.316	
CRT Right calibrated white lights on	.307	.322	
CRT calibrated white lights off	.305	.314	75.0
Flare at center with rear 2 banks on	.320	.340	3.96

Table 6. Colors chosen for gamut mapping color match.

L*	a*	b*	C* _{ab}	h degrees	color name
65.15	49.51	67.93	84.06	53.91	orange
63.19	60.67	-55.49	82.22	-42.44	violet
56.72	74.36	1.86	74.38	1.43	maroon
20.56	48.12	-71.30	86.02	-55.98	dark blue
87.73	23.26	-8.80	24.87	-20.72	reddish pink
75.63	-25.54	-36.77	44.77	-124.78	sky blue
42.65	70.86	-60.76	93.34	-40.61	purple
89.17	-48.15	-16.86	51.02	-160.69	bright cyan
85.64	-89.15	79.44	119.41	138.29	bright green
57.51	71.26	67.01	97.82	43.23	red
88.65	-57.56	23.25	62.08	158.00	light green
55.20	-7.30	-43.57	44.18	-99.51	medium blue
87.73	16.31	7.29	17.86	24.09	skin tone
97.11	-21.92	92.24	94.81	103.36	bright yellow
64.02	90.43	-54.42	105.54	-31.03	bright violet
32.64	78.94	-106.64	132.68	-53.49	deep blue
60.14	8.97	-60.98	61.64	-81.62	medium blue
60.86	66.92	28.69	72.82	23.20	light red
86.59	-67.48	18.19	69.89	164.91	light green
48.60	30.06	-76.90	82.57	-68.64	blue
76.74	-30.65	-31.15	43.70	-134.53	light blue
58.85	80.44	-28.84	85.46	-19.72	magenta
79.64	14.69	79.05	80.40	79.47	tannish orange
57.61	43.68	-64.74	78.11	-55.99	light purple
88.55	-46.26	83.74	95.67	118.91	pea green

6.3 Appendix C: Three parameter files corresponding to Figure 57

Notice that for each of the parameter files, only the second row of the entry named “premat” is different. Compare these differences to the change in color matching function and corresponding shape of the color space in figure 57.

Parameter file for top figure:

```
premat
0.380400 0.707600 -0.088000
-0.215700 1.165300 0.049800
0.000000 0.000000 1.000000
oppmat
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 0.000000 1.000000
rotmat
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 -0.000000 1.000000
postmat
0.000000 1.000000 0.000000
3.500000 -3.500000 -0.000000
0.000000 1.006356 -1.006356
nlinvals
10.000000 43.000000 10.000000 43.000000 10.000000 43.000000
Opponentargs
3.500000 -3.500000 0.000000 1.006356
UIargs
2 1 1 0 4
```

Parameter file for middle figure:

```
premat
0.380400 0.707600 -0.088000
-0.166667 1.100000 0.066667
0.000000 0.000000 1.000000
oppmat
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 0.000000 1.000000
rotmat
1.000000 0.000000 0.000000
```

```

0.000000 1.000000 0.000000
0.000000 -0.000000 1.000000
postmat
0.000000 1.000000 0.000000
3.500000 -3.500000 -0.000000
0.000000 1.006356 -1.006356
nlinvals
10.000000 43.000000 10.000000 43.000000 10.000000 43.000000
Opponentargs
3.500000 -3.500000 0.000000 1.006356
UIargs
2 1 1 0 4

```

Parameter file for bottom figure:

```

premat
0.380400 0.707600 -0.088000
-0.083333 0.933333 0.150000
0.000000 0.000000 1.000000
oppmat
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 0.000000 1.000000
rotmat
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 -0.000000 1.000000
postmat
0.000000 1.000000 0.000000
3.500000 -3.500000 -0.000000
0.000000 1.006356 -1.006356
nlinvals
10.000000 43.000000 10.000000 43.000000 10.000000 43.000000
Opponentargs
3.500000 -3.500000 0.000000 1.006356
UIargs
2 1 1 0 4

```

7 References

- ¹ Ebner, F., and Fairchild, M., Gamut Mapping From Below: Finding minimum perceptual distances for colors outside the gamut volume, *Color Research and Application*, VOL, NO, PP. (1997)
- ² Ebner, F., and Fairchild, M., Finding constant hue surfaces in color space, *Proceedings of SPIE, Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III*, **3300**, 107-117, (1998)
- ³ Munsell, A. H., A Color Notation, Munsell Color Co. 12th ed. 1971
- ⁴ Munsell, A. H., A Grammar of Color, Van Nostrand Reinhold Co., 1969
- ⁵ Birren, F., Application of the Ostwald Color System to the Design of Consumer Goods, *J. Opt. Soc. Am.*, **34**, No. 7, 396-399 (1944)
- ⁶ Zeishold, H., Philosophy of the Ostwald Color System, *J. Opt. Soc. Am.*, **34**, No. 7, 355-360 (1944)
- ⁷ Foss, C. and Nickerson, D., Analysis of the Ostwald Color System, *J. Opt. Soc. Am.*, **34**, No. 7, 361-381 (1944)
- ⁸ Stevens, J.C., and Stevens, S.S., Brightness Function: Effects of Adaptation, , *J. Opt. Soc. Am.*, **53**, No. 3, 375-385 (1963)
- ⁹ Hard, A. and Sivik, L., NCS-Natural Color System: A Swedish Standard for Color Notation, *Color Res. and Appl.*, **6**, No. 3, (1981)
- ¹⁰ Hard, A. and Sivik, L., Some Reflections on Studying Colour Combinations, *Color Res. and Appl.*, **19**, No. 4, (1994)
- ¹¹ Hard, A. and Sivik, L., Outlines of a Theory of Colors in Combination, *Man-Environment Systems*, 9:217-228, (1979)
- ¹² Hard, A. and Sivik, L., Distinctness of Border: An Alternative Concept for a Uniform Color Space, *Color Res. and Appl.*, **11**, No. 2, (1986)
- ¹³ Moon, P. and Spencer, D., Geometric Formulation of Classical Color Harmony, , *J. Opt. Soc. Am.*, **34**, No. 1, 46-59 ,(1944)
- ¹⁴ Moon, P. and Spencer, D., Area in Color Harmony, *J. Opt. Soc. Am.*, **34**, No. 2, 93-103, (1944)
- ¹⁵ Pope, A., Notes on the Problem of Color Harmony and the Geometry of Color Space, *J. Opt. Soc. Am.*, **34**, No. 12, 759-765, (1944)
- ¹⁶ Moon, P. and Spencer, D., Aesthetic Measure Applied to Color Harmony, *J. Opt. Soc. Am.*, **34**, No. 4, 234-242, (1944)
- ¹⁷ Granger, G.W., Area Balance in Color Harmony: An Experimental Study, *Science*, **17**, 59-61, Jan. (1953)
- ¹⁸ Nayatani, Y. et. al., An Appraisal of Two-Color Harmony by Paired Comparison Method, *Acta Chromatica*, **2**, No. 1 (1969)
- ¹⁹ Berlin and Kay, Basic Color Terms, U. of California Press, (1969)
- ²⁰ Kelly, K. and Judd, B., The ISCC-NBS Method of Designating Colors and A Dictionary of Color Names, *National Bureau of Standards Circular 553*, (1955)
- ²¹ Birren, F., Color and Human Response, Van Nostrand Reinhold, (1978)
- ²² Sivik, L., Studies of Color Meaning, *Man-Environment Systems*, **5**, no. 3, 155-160, (1975)
- ²³ Taft, C., The Stability and Variability of Color Meaning Associations Across Time and Cultures, *Colour 89, Proceedings of the 6th AIC Congress*, (1989)
- ²⁴ Sivik, L., Color Meaning and perceptual color dimensions: A study of color samples, *Goteborg Psychological Reports*, **4**, No. 1, (1974)
- ²⁵ Taft, C. and Sivik, L., Cross National Comparisons of Color Meaning, *Goteborg Psychological Reports*, **22**, No. 3, (1992)

-
- ²⁶ Stvik, L., Color Meaning and perceptual color dimensions: A study of Exterior Colors, *Goteborg Psychological Reports*, **4**, No. 11, (1974)
- ²⁷ Kobayashi, S., The Aim and Method of the Color Image Scale, *Color Research and Application*, **6**, No.2, (1981)
- ²⁸ Buck, G.B., and Froelich, H.C., Color Characteristics of Human Complexions, *Illuminating Engineering*, Jan (1948)
- ²⁹ Bartleson, C.J., Memory Colors of Familiar Objects, *Journal of Opt. Soc. of America*, **50**, No. 1, (1960)
- ³⁰ Newhall, S.M., Burnham, R.W., and Clark,J.R., Comparison of Successive with Simultaneous Color Matching, *Journal of Opt. Soc. of America*, **47**, No. 1, (1957)
- ³¹ Bartleson, C.J., Some Observations on the Reproduction of Flesh Colors, *Photographic Science and Engineering*, **3**, No. 3, (1959)
- ³² Bartleson, C.J. and Bray, C.P., On the Preferred Reproduction of Flesh, Blue-Sky, and Green-Grass Colors, *Photographic Science and Engineering*, **6**, No. 1, (1962)
- ³³ Hunt,R.W.G, Pitt, I.T., Winter, L.M., The Preferred Reproduction of Blue Sky, Green Grass, and Caucasian Skin in Colour Photography, *The Journal of Photographic Science*, **22**, (1974)
- ³⁴ Asada, T., Sakaue, K., Haneishi, H., and Miyake, Y., Preferred Skin Color Reproduction of Hardcopy, WWW page <http://www.icsd6.tj.chiba-u.ac.jp/ICSD6/papers/asada/skin.html>, (1994)
- ³⁵ Fedorovskaya, E., Blommaert, F., de Ridder, H., Perceptual Quality of Color Images of Natural Scenes Transformed in CIELUV Color Space, *Proceedings from the First IS&T and SID's Color Imaging Conference*, (1993)
- ³⁶ Jones, L., Psychophysics and Photography, *Journal of Opt. Soc. of America*, **34**, No. 2, (1944)
- ³⁷ MacAdam, D.L., Quality of Color Reproduction, *Journal of the SMPTE*, **56**, (1951)
- ³⁸ Roufs, J.A.J., Perceptual Image Quality: Concept and Measurement, *Philips J. Res.*, **47**, (1992)
- ³⁹ R.S. Gentile, E. Walowit, and J.P. Allebach, A Comparison of Techniques for Color Gamut Mismatch Compensation, *J. of Imaging Technology*, **16**, No. 4, Oct. 1990
- ⁴⁰ M.C. Stone, W.B. Cowan, J.C. Beatty, Color Gamut Mapping and the Printing of Digital Color Images, *ACM Transactions on Graphics*, **7**, No. 4, Oct. 1988
- ⁴¹ J. Meyer, B. Barth, Color Gamut Matching for Hard Copy, *SID 89 Digest pp 86-89*, 1989
- ⁴² Osgood, Suci, and Tannenbaum, *The Measurement of Meaning*, University of Illinois Press, 1978
- ⁴³ Ibid, page 73. Relative importance of semantic space dimensions.
- ⁴⁴ Same as ref. 24, page 10.
- ⁴⁵ Stone and J. Beatty, Color Gamut Mapping and the Printing of Digital Color Images, *ACM Transactions on Graphics*, **7**, No. 4, 249-292 (1988).
- ⁴⁶ W. Wallace and M. Stone, Gamut Mapping Computer Generated Imagery, *SPIE Vol. 1460 Image Handling and Reproduction Systems Integration*, 20-28, (1991).
- ⁴⁷ R. S. Gentile, E. Walowit, and J. P. Allebach, A Comparison of Color Gamut Mismatch Compensation, *J. of Imaging Technology*, **16**, No. 5, 176-181, (1990).
- ⁴⁸ M. Wollski, J. P. Allebach, C. A. Bouman, Gamut Mapping, Squeezing the Most out of Your Color System, *IS&T and SID's 2nd Color Imaging Conference: Color Science, Systems, and Applications*, 89-92, (1994).
- ⁴⁹ T. Hoshino and R. Berns, Color gamut mapping techniques for color hard copy images, *SPIE Vol. 1909*, 152-165, (1993).
- ⁵⁰ Montag, E.D. and Fairchild, M.D., Simulated Color Gamut Mapping Using Simple Rendered Images, *Proc. SPIE 2658*, San Jose, 316-325(1996).

-
- ⁵¹ N. Katoh and M. Ito, Gamut Mapping for Computer Generated Images (II), *IS&T and SID's 4th Color Imaging Conference: Color Science, Systems, and Applications*, 126-129, (1996).
- ⁵² William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery, Numerical Recipes in C, Second Edition, *Cambridge University Press*, ISBN 0-521-43108-5, pp.408-412, 1992.
- ⁵³ A. P. Hung and R. Berns, Determination of Constant Hue Loci for a CRT Gamut and Their Predictions Using Color Appearance Spaces, *Color Res. & Appl.*, **20**, No. 5, 285-295, (1995).
- ⁵⁴ Newhall, Nickerson, Judd, Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors, *J. Opt. Soc. Am.*, **33**, No. 7,385-418 (1943).
- ⁵⁵ R.W.G. Hunt, Measuring Color, Second Edition, *Ellis Horwood*, ISBN 0-13-567686-X, pp.142, (1991).
- ⁵⁶ R. W .G. Hunt, "Colour Terminology", *Col. Res. & Appl.*, **3**, No. 2, pp. 79-87, 1978.
- ⁵⁷ B. Berlin and P. Kay, "Basic Color Terms", pp. 17-23, University of California Press, Berkeley, 1969.
- ⁵⁸ R. S. Berns, R. J. Motta, and M. E. Gorzynski, "CRT colorimetry, part I: theory and practice", *Col. Res. & Appl.*, **18**, pp. 299-314, 1993.
- ⁵⁹ Timothy Masters, Practical Neural Network Recipes in C++, 1993 Academic Press
- ⁶⁰ Timothy Masters, Signal and Image Processing with Neural Networks, 1994 Wiley
- ⁶¹ Timothy Masters, Advanced Algorithms for Neural Networks: A C++ Sourcebook, 1995 Wiley
- ⁶² R. Rolleston, "Using Shepard's Interpolation to Build Color Transformation Tables", *IS&T and SID's 2nd Color Imaging Conference: Color Science, Systems and Applications*, pp.74-77 (1994).
- ⁶³ S. L. Guth, "Model for color vision and light adaptation", *J. Opt. Soc. Am. A*, **8**, No. 6, pp. 976-993 (1991)
- ⁶⁴ M.D. Fairchild, and R.S. Berns, "Image Color-Appearance Specification Through Extension of CIELAB", *Col. Res. & Appl.*, **18**, No. 3, pp. 178-189, 1993.
- ⁶⁵ G. Wyszecki and W.S. Stiles, "Color Science: Concepts and Methods, Quantitative Data and Formulae", Second Edition, Wiley Press, ISBN 0-471-02106-7, (1982).
- ⁶⁶ S.M. Newhall, D. Nickerson, D.B. Judd, "Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors", *J. Opt. Soc. Am.*, **33**, No 7, pp. 385-418, (1943).
- ⁶⁷ M.D. Fairchild, "Refinement of the RLAB Color Space", *Col. Res. & Appl.*, **21**, No. 5, pp. 338-346, 1996.
- ⁶⁸ M.D. Fairchild, E. Pirrotta, "Predicting the Lightness of Chromatic Object Colors Using CIELAB", *Col. Res. & Appl.*, **16**, No. 6, pp. 385-393, 1991.
- ⁶⁹ M. Melgosa, E. Hita, A.J. Poza, D.H. Alman, R.S. Berns, "Suprathreshold Color-Difference Ellipsoids for Surface Colors", *Col. Res. & Appl.*, **22**, No. 3, pp. 148-155, 1997.
- ⁷⁰ Personal communication with Gus Braun, and an abstract submission by Gus Braun, the author, and Mark Fairchild for CIC'98, "Color Gamut mapping in a Hue-Linearized CIELAB Color Space".
- ⁷¹ G.A. Gescheider, "Psychophysics, Method, Theory, and Application", Second Edition, Lawrence Erlbaum Associates, ISBN 0-89859-375-1, pp. 152 (1985).