Colour discrimination thresholds and acceptability ratings using simulated Microtile^{TM} displays.

by

Mahalakshmi Ramamurthy

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Vision Science

Waterloo, Ontario, Canada, 2011

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Introduction

Nearly all flat panel video display monitors have luminance and colour variations as the angle of view varies from the monitor's perpendicular. The new Microtile^{TM} displays developed by Christie Digital are no exception to this general finding. A review of any book on colour science will show that there is substantial amount of literature on just noticeable colour differences within various colour spaces. Despite the wealth of data on the topic, there is no general consensus across different industries as to which colour space and colour difference equations are appropriate. Several factors like the background colour, object size, texture of the stimulus are different for different studies; these factors make it very difficult to determine precisely the effect of viewing angle on the perception of coloured images on the Microtiles display based on previous research. Hence, the objective of this thesis was to quantify the measured colour shifts of a Microtile display at different viewing angles, in steps of perceptible thresholds and to evaluate the acceptability of distortions at different viewing angles for complex scenes.

Methods

A preliminary experiment was setup to study the behaviour of Microtile display primaries as a function of viewing angle. The aim was to measure the shift in hue and luminance of the three primaries at different eccentricities (from 0^o to 40^o). The measured trend was used to simulate Microtile shifts on complex images for the rating task.

In the first part of the perceptibility experiment, three reference colours were picked and 12 vectors heading towards the blue-yellow region of the L*a*b* colour space (pertaining to the colour shifts noticed with the Microtile displays). A uniform reference colour was presented in three of the four quadrants on the CRT monitor and one quadrant changed colour in the direction of the sampled vector. An adaptive, four alternate forced choice procedure was employed to determine thresholds for

each of the 3 reference colours. The adaptive technique used was a ZEST paradigm. In the second part of the perceptibility experiment, eighteen directions were sampled around each reference colour.

The rating task was based on simulating the measured attenuations of the Microtile primaries on complex scenes. Subjects rated the images both in terms of acceptability/unacceptability and as percentage image degradation. The simulation was presented on three static complex images, car, landscape and portrait. A total of 60 subjects participated in the study, 20 subjects for each experiment. All subjects were between the age group of 15 to 35 years of age and underwent battery of colour vision tests before being included in the study. All subjects included had average to superior colour discrimination as categorized using the FM-100 Hue discrimination test.

Results

Study1: The preliminary study on Microtile display characteristics as a function of viewing angle showed that all the three primaries decreased in luminance with change in viewing angle. The red primary decreased at a faster rate compared to the other two primaries. The trend presents as a decrease in luminance with the hue shifting towards the blue-green region of the CIE1974 L*a*b* space.

Study 2: Results from both the first and second parts of the perceptibility experiment showed that the vectors sampled in different directions approximated to ellipsoids in the L*a*b* colour space. This finding was consistent with the colour discrimination literature. Vectors on the equi-luminance plane were significantly longer than the vectors on the non equi-Luminance plane. Results showed that the average perceptibility thresholds in the non equi-luminance direction were lower than $1\Delta E_{Lab}$.

Study 3: Results from the rating experiments showed that irrespective of the complexities in the images, distortions greater than five times thresholds were less than 50% acceptable and were rated to be at least 30% degraded. This corresponds to a viewing angle greater than 10^{o} for a Microtile display. The relationship between the stimulus (ΔE) and subjective Image degradation judgements

followed a linear relationship, with the portrait and landscape having similar functions, whereas the car was rated more degraded at lower ΔEs and less degraded at higher ΔEs compared with the other two scenes.

Conclusion

Perceptibility thresholds for different reference colours showed that the conventionally used calibration precision of 1 ΔE_{Lab} is a lenient criterion. Perceptibility thresholds are at least 25% less for the Microtile display reference condition. From the results of the rating data a distortion greater than five times thresholds is less than 50% acceptable and appears to be at least 30% degraded for static complex images. However, the image quality judgments appear to be related to scene context, which requires further study.

Acknowledgements

I am deeply indebted to my supervisor Dr. Vasudevan Lakshminarayanan. Your support and motivation has inspired me throughout and has helped me evolve as a different person. Your approach to science and the time spent under your guidance has provided me an incredible learning experience. I will always remain grateful to you for providing me a great start to my career.

I extend my sincere thanks to Dr. Jeff Hovis, for guiding me at every step of this graduate program. Your support and motivation has always been my driving force. You are a great teacher and the experience under your guidance is just invaluable.

I would like to extend my sincere gratitude to Dr. Trefford Simpson. Your valuable comments have always motivated me to think. You are a great inspiration as a teacher and I am thankful for those cheerful words and support you gave me throughout.

I would like to acknowledge my programmers Arthur Smith and Ashish Gupta for your help with MatLab programming.

I would like to acknowledge Christie Digital for sponsoring this project. Special thanks to Delia Zsivanov for guiding and arranging experiments in the Microtile Laboratory. I would also like to thank Jill Tomasson Goodwin for organizing all official meetings.

I would also like to thank the Graduate officers, Dr. Natalie Hutchings and Dr. Trefford Simpson, for your remarkable support and concern throughout. Thanks to the members of Dr. Vengu's lab for setting a great working atmosphere and to the staff & faculty of the School of Optometry.

Thanks to fellow graduate students (GIVS) for providing a very friendly environment.

Finally, I would like to thank my parents, cousins and friends for your unbound love and support. I could not have done this without you.

Dedication

This thesis is dedicated to my parents, and to all my teachers who have been completely inspirational.

I would like to list all my mentors by name:

"Mrs. Arokia Mary, Mrs. Peace Lawrence, Mrs. Emaculate, Mrs. Georgina, Sister Lobo, Mrs. Maria Selvi, Mrs. Anita, Mrs. Sulochana, Mrs. Auxilla Priya, Mrs. Josephine, Mrs. Priya, Mrs. Rose, Dr. Ramaswamy, Dr. L. Srinivasa Varadharajan, Mr. Vishwanathan, Dr. Rashmin Gandhi, Dr. Krishna Kumar, Mrs. Revathy, Dr. Vasudevan Lakshminarayanan, Dr. Jeff Hovis and Dr. Trefford Simpson".

Table of Contents

List of Figures

List of Tables

Introduction

Video Display Terminals have come a long way from Cathode ray tube monitors to today's high definition flat screen displays. Flat screen display designs have a number of attractive features including, relatively compact size and extended colour gamut. The newest generation of displays have resolution that is equal to, or superior to the cathode ray tubes. Nevertheless a common problem with almost all flat panel displays is the change in luminance and hue with different viewing angles. The new Microtile™ display designed by Christie Digital is no exception to this general finding. Given the behaviour of the Microtile displays with viewing angles, the company is interested in quantifying these changes in terms of perceptible colour differences and in determining the acceptability of these shifts as perceived by the observer/customer. The aim of this study is to first translate the measured physical differences at different viewing angles into perceptual threshold units and to have observers rate the acceptability/unacceptability of these differences simulated on complex scenes.

Chapter 1 summarises classic literature on colour difference work, the development of different colour difference equations, and recent progress in the area. The change in luminance and hue of a Microtile™ display as a function of viewing eccentricity is studied in Chapter 2. Chapter 3 discusses the first set of experiments on the perceptibility thresholds based on Microtile™ display characteristics presented in Chapter 2. Chapter 4 is an extension of the perceptibility study, where the thresholds are measured for a greater number of hue and luminance directions for four reference colours. These results are compared to the previous threshold values reported in the literature.

Chapter 5 examines the acceptability of different supra-threshold colour differences on complex images. The images are distorted in a manner similar to what could occur in a Microtile display at different viewing angles. Chapter 6 provides the general conclusions from the above experiments.

Chapter 1
Literature Review

1.1 Introduction:

Historically, colour difference research has been based on two broad purposes. The first was to analyze colour difference data in order to gain information about colour discrimination mechanisms. The second purpose was to use the data as a metric to quantify perceptible colour differences for industrial quality control purposes. This chapter summarizes the colour difference studies that eventually led to the development of different colour spaces and colour difference equations and discusses the general colour difference problem.

1.2 Historic perspective of colour difference

Colour is defined as that characteristic of visible light by which an observer may distinguish differences between two fields of identical contours by just the differences in the spectral composition of the radiant power (Wyszecki, G., & Stiles, W. S. 1982; Kaiser, P.K. 1996). Colour discrimination experiments can be divided into four broad classes of studies based on the different stimulus properties, namely; luminance differences (often referred to as brightness or lightness differences), wavelength (hue) differences, purity (saturation or chroma) differences, and chromaticity (both hue and saturation) differences. After the development of the colour specification system by the Commission Internationale d'Eclairage (CIE), colours have been represented in terms of chromaticity co-ordinates and luminance in a three dimensional colour space. The spaces initially represented the amount of the three primaries required to match a test light, but now the spaces are more appearance based with two orthogonal hue axes, one usually specifying a red-green dimension and the other a blue-yellow dimension. There is a third axis perpendicular to the hue plane usually representing the

luminance or black-white dimension (Shevell, S.K. 2003). These newer spaces are the ones used most frequently in representing colour standards in various industries.

Studies on colour differences can be further classified based on the experimental design. One design measures the perceptibility thresholds directly using psychophysical methods. In these experiments, the task is most often based on asking subjects to compare two stimuli and identify which differs in hue, chroma, luminance or any combination of the three. The second design measures colour discrimination as intra-observer precision (sample covariance) in colour matching experiments. In such designs, colours that fall within 1 to 2 standard deviations of the mean match are considered to be identical to the reference colour, whereas colours outside these boundaries are considered to appear different from the reference colour (Bass, M. 2010).

One of the first systematic studies of colour discrimination, using a standard colour vision space was by Wright (1941). His results showed just noticeable differences (JND) in colour, mapped as unequal line segments in the CIE 1931 x, y, L space (Wright, WD 1941). The unequal line segments indicated that the perceptible colour difference thresholds were not equal distances within a colour space derived from the colour matching experiment. These results also demonstrated that the CIE x, y, L was not a uniform space in terms of colour appearance and discrimination. Wright's results were over shadowed by MacAdam's extensive data on the precision of the colour matching for a single observer. The stimulus was a 2^{o} bipartite field within a dark surround. The task was to adjust the test field's chromaticity, to match the comparison field. Twenty-five standard colours were sampled across the CIE 1931 x, y space. Discrimination thresholds were established as two standard deviations away from the mean match point. On a 2-dimensional plane, MacAdam's data were represented as ellipses of various sizes and orientation. This finding has led to the never-ending quest of developing a uniform colour space and deriving a colour difference equation so that an equal distance anywhere within the space represents an equal perceptual change in the stimulus appearance.

A summary of different colour difference studies and colour equations are given in Table 1.1, which is an expansion and update of Table I (6.5.1) in Wyszecki, G., & Stiles, W. S. 1982.

Colour differences based on observer variability in colour matches is problematic for textile, plastics, and paint industries, as it is difficult to adjust precisely the colour of the materials continuously. Therefore, there was a need to design experiments based on the subjective judgement of differences (acceptability/tolerance) appropriate to the industrial needs. One of the first experiments on the acceptable/unacceptable judgements of colour differences was the study carried out by Davidson in 1953. The experiment was based on a particular production problem in carpet industry. In the study, eight experienced observers made about 300 judgements of acceptability of 15 samples around each of the 19 standards on several occasions. Colour differences that were rated as acceptable were considered to be below observer's threshold, or at least within their tolerance limits. Davidson and his co-workers were able to transform these results into unit circles in a distorted CIE 1931 chromaticity diagram (Wyszecki, G., & Stiles, W. S. 1982).

As Berns (2000) points out in his review, there were several approaches to evaluate colour difference equations, owing to different industrial needs and one of the approaches to evaluate colour difference was by correlating visual data with the calculated magnitude from colour difference equations. Later in 1975, this approach was proved to be misleading as the relationship between the visual assessments and the physical magnitudes were sigmoidal and not linear functions. This correction was implemented in Rich's analysis of pass-fail type subjective data, where he used sigmoidal transformations to predict tolerance limits. However, the ΔE values that produced a given percentage of acceptable matches varied with location within the colour space and varied between several industries. Despite these confounding factors, many colour-related industries have established colour difference tolerances that are based on acceptability/unacceptability judgments (Berns, R.S. 2000).

Table 1.1Timeline of colour difference equations.

Colour Difference Equation proposed by	Year	Data Set	Colour difference formula (Δ E)	Application
Nickerson index of fading[5]	1936	Munsell data	$\Delta E = \frac{2}{5}C\Delta H + 6\Delta V + 3\Delta C$ *where H, C, V are Munsell hue, chroma and value*	
Balinkin[5]	1941	Munsell data	$\Delta E = \left(\frac{2}{5}C\Delta H\right)^2 + (6\Delta V)^2 + \left(\frac{20}{\pi}\Delta C\right)^2 \quad ^{1/2}$	Modification of the Nickerson index to match Euclidean geometry.
Adams-Nickerson (ANLAB 40)[5]	1942	Munsell data	$\Delta E = \quad 40 \quad (0.23\Delta V_y)^2 + \quad \Delta \quad V_x + V_y \quad ^2 + \quad 0.4\Delta(V_y + V_z)^2 \quad ^{1/2}$ *V is the Munsell value function.*	
Judd Hunter[5] NBS unit	1942	Judd's data	$\Delta E = \quad f_g \quad 221Y^{\frac{1}{4}} \quad \Delta\alpha^2 + \Delta\beta^2 \quad ^{\frac{1}{2}} \quad ^2 + \quad k\Delta(Y^{\frac{1}{2}}) \quad ^2 \quad ^{1/2}$ f_g accounts for masking influence of the glossy surface	Glossy surface
Cube root[4]	1957	Modified Adams data	$\Delta E = \quad \Delta L^2 + \Delta a^2 + \Delta b^2 \quad ^{1/2}$ $\Delta L = 25.29\ \Delta(G^{1/3})$ $\Delta a = 106.0\ \Delta\ (R^{1/3}) - \Delta\ (G^{1/3})$ $\Delta b = 42.34 \quad \Delta \quad G^{\frac{1}{3}} \quad - \quad \Delta \quad B^{\frac{1}{3}}$ *The $\Delta R^{1/3}$, $\Delta G^{1/3}$, $\Delta B^{1/3}$ denote the cube root of the difference in reflectance of the sample and the standard.*	The simplicity of the expression makes it flexible for practical colour matching purposes.

Simon-Goodwin[5]	1958	MacAdam's data	Colour difference chart – combination of Brown- MacAdam's data and modified Munsell-type lightness.	Used for rapid hand calculation of colour difference
CIE 1964[3]	1964	Five data sets (Wright's and Sugiyama's data)	$W = 25Y^{\frac{1}{2}} - 17$ $U = 13W(u - u_0)$ $V = 13W(v - v_0)$ $u = \frac{4X}{X + 15Y + 3Z}$ $v = \frac{6Y}{X + 15Y + 3Z}$ $\Delta E = \left(U_1 - U_2\right)^2 + \left(V_1 - V_2\right)^2 + \left(W_1 - W_2\right)^{2\ 1/2}$ *Subscripts 1 and 2 represent the comparison and test colours.*	Colour control in industry
Friele-MacAdam-Chickering (FMC 1 and 2)[1, 2]	1967	MacAdam's data	FMC 1- $\Delta E=((\Delta C)^2+(\Delta L)^2)^{1/2}$ *where, ΔC represents the total chromatic difference and ΔL represents the luminance difference.* FMC 2- the chromatic difference is multiplied by a factor *K1*.	FMC 2- for paper, paint, plastic, textile industries.
McDonald[5]	1976	J & P Coat's data	$\Delta E_{JPC79} = \left(\frac{\Delta L}{S_L}\right)^2 + \left(\frac{\Delta C}{S_C}\right)^2 + \left(\frac{\Delta H}{S_H}\right)^{2\ \frac{1}{2}}$	Dyes
[7]CIE76 L*a*b*	1976	Adapted from Adams-Nickerson equation	$\boldsymbol{L^* = 116\ Y\ Y_0^{\frac{1}{3}} - 16}$ $\boldsymbol{a^* = 500\ \left(X\ X_0^{1\ 3} - Y\ Y_0^{1\ 3}\right)}$ $\boldsymbol{b^* = 200\ \left(Y\ Y_0^{1\ 3} - Z\ Z_0^{1\ 3}\right)}$ $\Delta E^*_{ab} = \left(\Delta L^{\ 2} + \Delta a^{*\ 2} + \Delta b^{*\ 2}\right)^{\frac{1}{2}}$	Generally applied in surface colour industries.

[7]CIE76 L*u*v*	1976		$L^* = 116\ Y\ Y_0\ ^{\frac{1}{3}} - 16$ $u^* = 13L^*\ u' - u_0{}'$ $v^* = 13L^*\ v' - v_0{}'$ $u' = \frac{4X}{X + 15Y + 3Z}\quad v' = \frac{9X}{X + 15Y + 3Z}$ $\Delta E^*{}_{uv} = \quad \Delta L\ ^2 + \ \Delta u^*\ ^2 + \ \Delta v^*\ ^2\ ^{\frac{1}{2}}$	For Television and illumination industry. In general to specify just perceptible colour differences
CMC[5]	1984	Based on CIE Lab	$\Delta E_{CMC} = \quad \frac{\Delta L^*}{lS_L}\ ^2 + \ \frac{\Delta C_{ab}{}^*}{cS_C}\ ^2 + \ \frac{\Delta H_{ab}{}^*}{S_H}\ ^2\ ^{\frac{1}{2}}$ *S_L, S_C, S_H, l, c are weighting factors*	British and American standard for textile industries
BFD[5]	1987	Luo-Rigg data set	$\Delta E_{BFD} = \quad \frac{\Delta L^*}{l}\ ^2 + \ \frac{\Delta C_{ab}{}^*}{cD_C}\ ^2 + \ \frac{\Delta H_{ab}{}^*}{D_H}\ ^2 + R_T\ \frac{\Delta C_{ab}{}^*}{D_C}\ \ \frac{\Delta H_{ab}{}^*}{D_H}\ ^{\frac{1}{2}}$	
CIE 94[5]	1994	Luo-Rigg[8] and RIT-DuPont data set.	$\Delta E_{94} = \quad \frac{\Delta L^*}{k_L S_L}\ ^2 + \ \frac{\Delta C_{ab}{}^*}{k_C S_C}\ ^2 + \ \frac{\Delta H_{ab}{}^*}{k_H S_H}\ ^2\ ^{\frac{1}{2}}$ *S_L, S_C, S_H, are weighting factor*	
CIE 00[6]	2000	BFD, Rit-DuPont, Luo &Rigg – combined data sets	$\Delta E_{00} = \quad \frac{\Delta L'}{k_l\, S_L}\ ^2 + \ \frac{\Delta C'}{k_C\, S_C}\ ^2 + \ \frac{\Delta H'}{k_H\, S_H}\ ^2 + R_T\ \frac{\Delta C^*}{k_C\, S_C}\ \ \frac{\Delta H'}{k_H\, S_H}\ ^{\frac{1}{2}}$ *S_L, S_C, S_H, are weighting factor and R_T is the rotational factor.*	Most recent

[1]Chickering, KD 1971; [2]Chickering, KD 1969; [3]Wyszecki, G. 1965; [4]Glasses, LG 1958; [5]Berns, R.S. 2000; [6]Luo, M.R. 2001; [7]Robertson, A. 1977; [8]Luo, MR 1986.

1.3 Recent Colour difference studies

One of the major drawbacks in comparing several classic colour difference equations is the diversity in the stimulus parameters used in different studies. As a solution to this problem, the CIE established certain guidelines to co-ordinate research on colour difference experiments (Robertson, A. 1977). Since then, efforts have been made to adhere to the suggested criteria and to define each parameter tested. General guidelines established by the CIE are listed in Table 1.2.

Recent colour discrimination studies use computers to generate the stimulus and therefore, most of the targets used now are CRT and LCD generated colours, mainly due to the ease in producing desired colours coupled with good accuracy (Berns, R.S. 1993, Brainard, D.H. 2002). In fact, several studies suggest that CRT and LCD displays are a convenient and useful tool in colour science and can replace surface colour data (Xu, H. 2005; Baribeau, R. 2005; Urban, P. 2011).

Table 1.2 CIE guidelines for studying colour differences (Robertson, A. 1978).

Minimum number of Reference colours	Grey: x_{10}: 0.314, y_{10}: 0.331, L_{10}: 30 Red: x_{10}: 0.484, y_{10}: 0.342, L_{10}: 14.1 Yellow: x_{10}: 0.388, y_{10}: 0.428, L_{10}: 69.3 Green: x_{10}: 0.248, y_{10}: 0.362, L_{10}: 24 Blue: x_{10}: 0.219, y_{10}: 0.216 , L_{10}: 8.8
Parametric factors to be defined for colour discrimination experiments.	Sample size Illumination level Sample separation Texture Luminance factor Size of ΔE

	Colour of surround Observer variability Duration of observation Monocular or binocular viewing

The most recent version of the CIE colour difference equation is the CIE 2000 equation developed by Luo et al, (2001) using a combination of five extensive datasets. This equation consist of weighting factors for the hue, chroma and luminance terms in addition to a rotation factor to align ellipses in the blue region of the colour space. The CIE 2000 has been evaluated against different data sets. Cui et al (2001) collected a large set of colour discrimination data using CRT generated colours. Besides investigating the effects of different stimulus parameters (e.g. the interaction between stimulus and background colours, gap width, frame size, sample size), the data was used to test colour difference equations. The evaluation of different equations was based on a measure called the performance factor (PF/3) (adapted from, Guan, S.S. 1999). This measures the percentage disagreement between two comparison sets, so that a higher PF/3 indicates larger difference between the calculated colour difference and the actual data. Cui et al's results showed that the CIE 2000 outperformed five other L*a*b* based colour difference formulas CIELAB, CMC (Colour Measurement Committee), BFD, CIE94, LCD (CIE Colour Appearance Model02, from large colour difference data set abbreviated as LCD; Luo, M.R. 2006). Part of the reason for CIE 2000 outperforming other equations was due to its relatively accurate predictions in the blue region where the tilt in ellipses is different from other regions in space (Cui, G. 2001).

Later, Xu et al (2004) designed a visual evaluation task to compare threshold (the extent to which a test colour appears the same) and supra threshold (perceivable difference between pairs) colour differences using CRT colours. Their results showed that the CIE 2000 and CMC out

performed CIE94 and CIE Lab in predicting both threshold and supra-threshold data. CIE 2000 was particularly good in the blue and grey regions of the colour space. Results from hue discrimination study carried out by Baribeau and Robertson (Baribeau, R. 2005); however, showed that the CIE2000 cannot fully account for hue dependent changes in the blue-purple region of the colour space, because of the dependency of the weighting factors on age. These results suggest that the acceptance of the CIE2000 equation is still in its testing stage and the development of a uniform colour space is still a challenge (Melgosa, M. 2000).

The review of the colour difference equations also helps to determine the variability in the perceptible colour difference thresholds within a common colour space for different subjects and different procedures. The variation in colour discrimination thresholds range from a 0.5 ΔE_{Lab} units based on the colour difference of 2 standard deviations in CIE1931 xyL space from Brown's data (Brown, W.R.J. 1957) to about $3\Delta E_{Lab}$ units as reported in Xu's data (Xu, H. 2005). Xu's colour difference thresholds are equivalent to about 4.5 standard deviations from the mean colour match in a CIE 1931 xy space. This information on the variability in the thresholds is helpful in evaluating future data sets to get a sense of its stands in the pool colour discrimination studies.

1.4 Conclusion

Despite the wealth of colour difference data in the past, deriving new threshold based data is still required. This is due to the highly variable range of thresholds with different stimuli and procedures/tasks (Poirson, A.B. 1990). Although the CIE recommended parameters have set guidelines in colour difference research and have improved the efficacy with which data sets are organized, it is sometimes not practically applicable to adhere to all of the parameters (Robertson, A. 1977). The CIE standards are definite guidelines in setting up an experiment, but the reference colours and reference viewing conditions are amendable to the aim of the experiment. Since, the

different data sets discussed above are restricted to implicit or explicit viewing conditions; there is a need to define reference viewing conditions and to construct the stimulus based on the Microtile™ display characteristics. Chapter 2 details on the Microtile™ display characteristics.

Chapter 2

Preliminary study on Microtile display characteristics as a function of viewing eccentricity

2.1 Objective

The purpose of this study was to measure the changes in the luminance and hue of a Microtile display (Christie Digital, Ontario) at different viewing eccentricities. This data provides information on the physical attenuation of the three primaries at different horizontal viewing angles and is used in simulating these shifts on complex images for observer rating.

2.2 Background

2.2.1 Microtile- Screen technology

Individual Microtile displays are 40.7 by 30.5cms rear projection displays that can be assembled to form larger displays in various configurations. A three colour Light Emitting Diode (LED) is used as the source and is placed at the focal length of the Fresnel lens layer. The Fresnel lens serves to distribute luminance evenly on to the next layer, which is the lenticular lens layer. The lenticular lens in turn distributes the light onto the third layer, which is the actual viewing screen (from internal reports, Christie Digital). A schematic representation of the optical rear projection technology is shown in Figure 2.1. Despite the highly sophisticated screen technology, there are perceivable differences in hue and brightness as viewing angle changes.

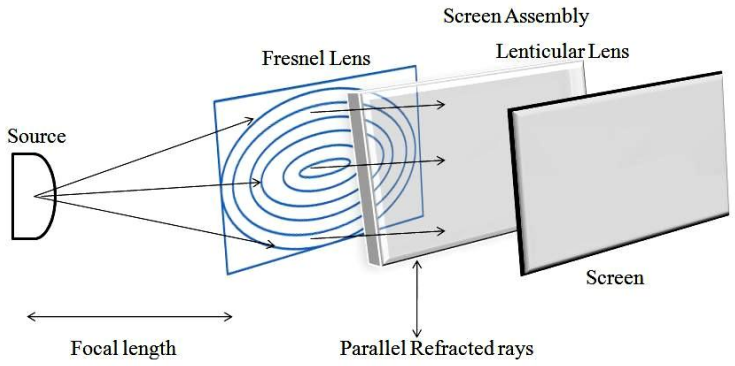


Figure 2.1 Schematic representation of the optical rear projection design.

2.3 Experimental set up

In order to quantify the changes in hue and luminance with changes in viewing angle, a 3x3 Microtile array was assembled. Measurements were first carried out on a single Microtile (center tile of the 3x3 array) at a viewing distance of 158.5cms. The changes in luminance were then confirmed for a larger display (a 3x3 array), measured from a viewing distance of 475.5cms. These values were based on the industry convention where a typical viewing distance is set to be 3 times the diagonal of the display under consideration. The Microtiles were calibrated and colour matched using internal software. The colour of the centre Microtile was measured at different viewing angles using a SpecBos 1201 spectrophotometer (Jeti Instruments, Germany). Because the displays were large, the 1964 CIE 10 degree colour-matching functions were used to convert the radiance to the tristrimulus values. The only light source in the room was the Microtile display. Figure 2.2 shows a schematic representation of the set up.

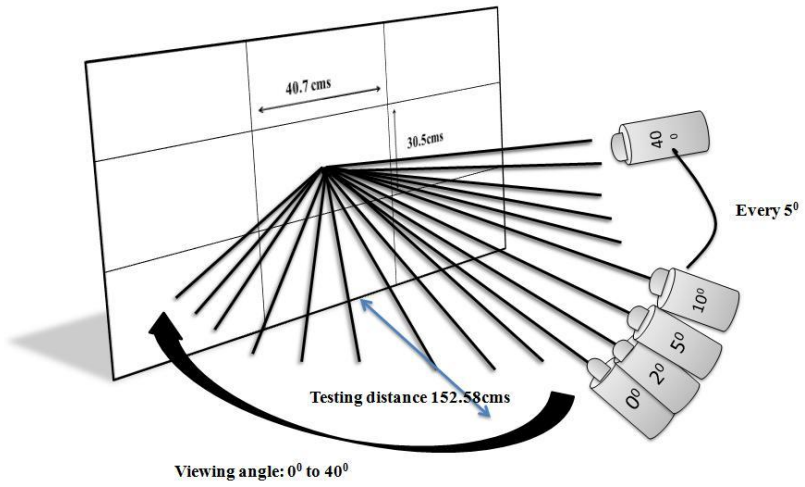


Figure 2.2 Experimental setup for measuring changes in colour as a function of horizontal viewing angle.

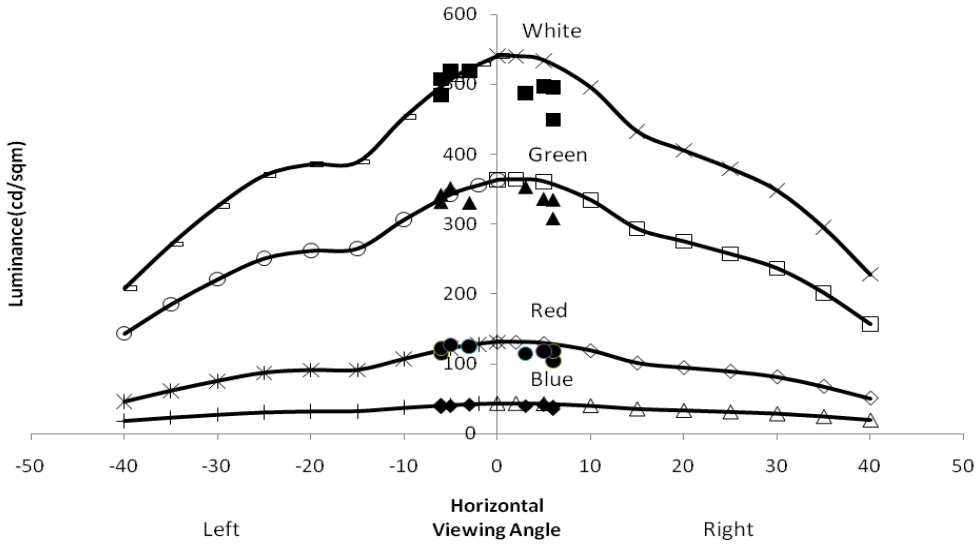
Measurements were taken for each of the three primaries and the white at their maximum settings. The angles varied from 0 degrees to 40 degrees on the left and right, and vertically from 0 to 5 degrees only. This was because of the limitations in the height of the SpecBos tripod.

2.4 Results

2.4.1 Changes in Luminance and Hue at different viewing angles

Figure 2.3a shows the change in luminance as a function of the horizontal viewing angle. All 3 primaries and the white show a decrease in luminance as the viewing angle (in this study, the measuring angle) increases. The changes in the white luminance would be equal to the sum of the three primaries at any given angle. The normalized values in Figure 2.3b shows that the rate of change of luminance measured for the 3 primaries. Note that the change is greatest for the red and essentially identical for the blue and green. Figure 2.3b also shows that the notches in the green and

white luminance profiles that occur at 15^{o} are also present in the red and blue primaries. This indicates that notch seen in the luminance profiles is an inherit characteristic of the display.



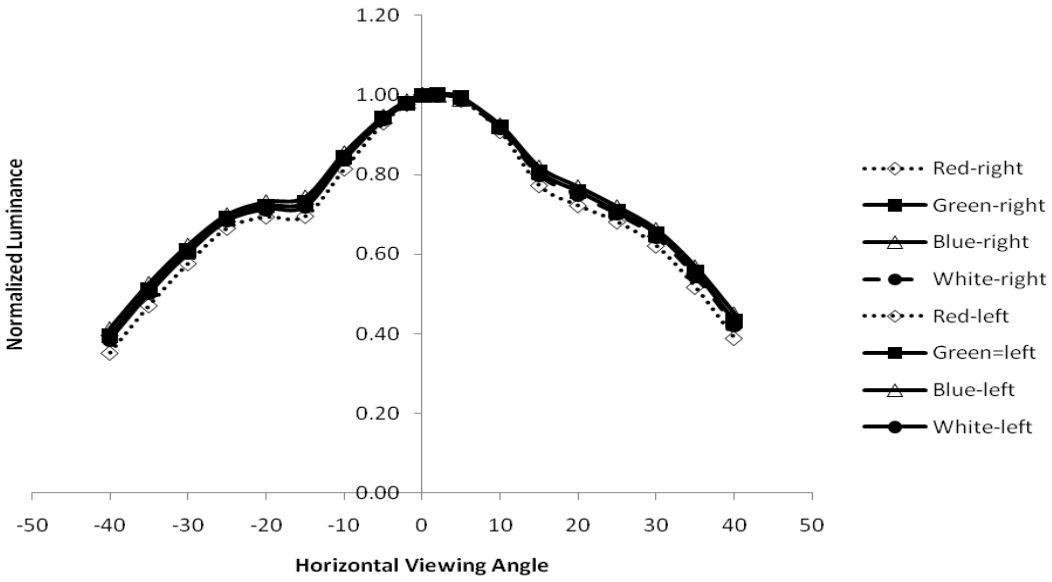


Figure 2.3 a (above) – shows change in luminance as a function of horizontal viewing angle. The shaded markers correspond to the measurements made on a 3x3 array. 2.3b (below) – shows normalized luminance as a function of horizontal viewing angle.

Figure 2.4a shows the results for the vertical direction for a limited range of viewing angles and 2.4b shows the normalized luminance for the same. There is a decrease in luminance as the vertical angle of view changes and there is an asymmetry between the inferior and superior directions. The change at a 5^{o} vertical viewing angle is slightly greater than the change noted with the horizontal meridian.

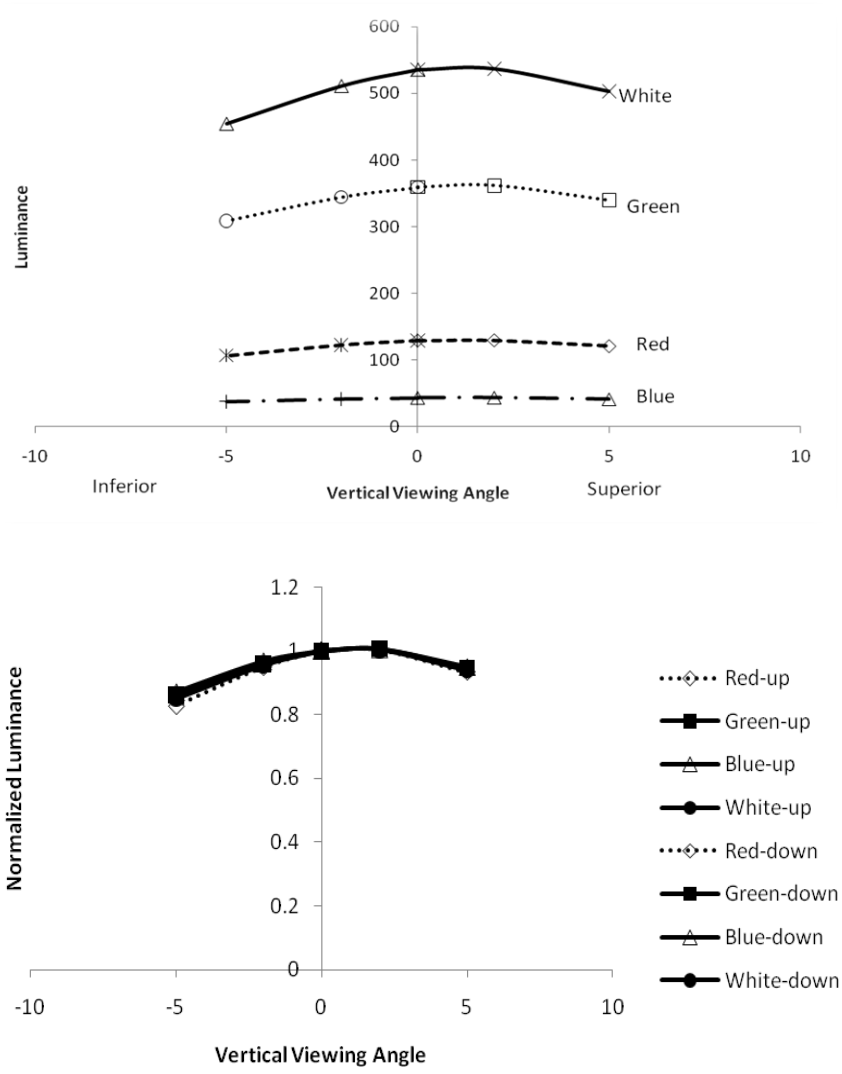


Figure 2.4 a (above) Graph shows the drift in luminance as a function of vertical viewing angles. The drift at 5^{o} is more rapid than that observed with horizontal viewing angles. 2.4 b (below) Shows the normalized luminance for the three primaries at different vertical viewing angles.

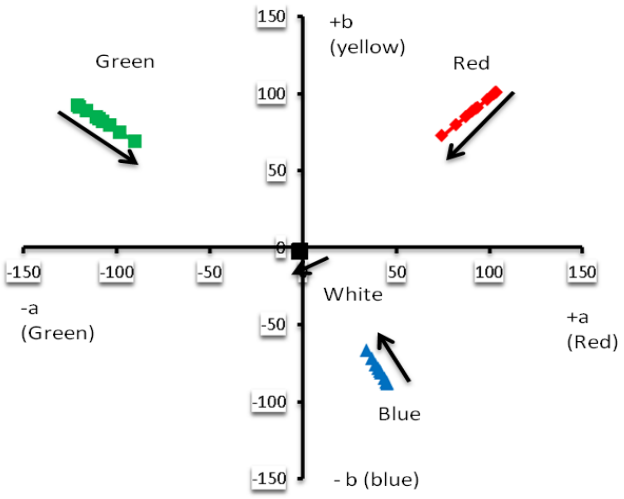


Figure 2.5 Direction of shift on the hue plane. Graph shows the direction of shift in each primary at different viewing angles. The tail of the vector corresponds to the hue at a head on view(0^0) and the head of the vector corresponds to a 40^0 viewing angle.

In addition to the changes in luminance, there were also systematic changes in the chromaticity coordinates of the red, green and blue primaries. Figure 2.5 shows the differences in the hue plane of the CIE1976 L*a*b* space. Figure 2.6 shows Δ hue relative to straight-on viewing as a function of horizontal viewing angle. A ΔE_{Lab} of 1 is the standard industrial calibration precision. The figures show that there are substantial differences in the chromaticity coordinates for the primaries as the viewing angles change; but only a relatively small net effect on the white which is shifted towards the blue-green region of the a*b* space. This small change in the white towards the blue-green region is consistent with approximately equal amounts of desaturation in the 3 primaries and a slightly larger decrease in the luminance of the red primary relative to the blue and green primaries.

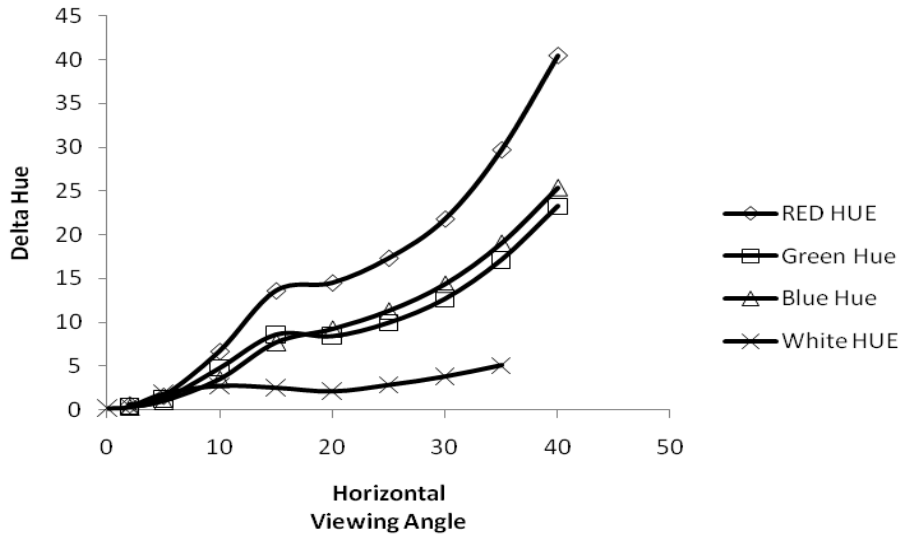


Figure 2.6 Delta Hue (in the a*b* space - $((\Delta a)^2 + (\Delta b)^2)^{1/2}$ as a function of different viewing angles.

2.4.2 Single Microtile™ versus an array

To ensure that the changes in luminance for a single Microtile were consistent with looking at a different Microtile at similar angles, the SpecBos was rotated to the center of each of the display in the 3x3 array to measure the luminance and chromaticity shifts. Figure 2.7 shows a schematic of the set up. The viewing angle (or the measuring angle) is the angle subtended by the center-to-center distance from the center display to each Microtile in the 3x3 array. The horizontally adjacent displays subtend a viewing angle of $5^o\ 4'$ with the center tile; vertically adjacent displays subtend a viewing angle of $3^o\ 48'$; and the diagonal displays subtend a viewing angle of $6^o 19'$. The results are plotted in Figure 2.3a (plotted as filled markers) and are similar to the shifts noticed within a similar range of viewing angles for a single Microtile. These results confirm that the behaviour of an array of displays that are well calibrated and colour matched can be characterized by a single display at different viewing angles.

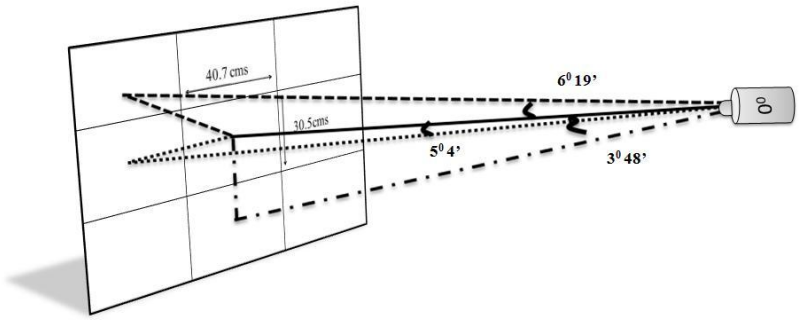


Figure 2.7 Schematic of measurements on a 3x3 array. The Spectrophotometer is placed at a distance three times the diagonal of the array and is fixed perpendicular to the center of the center tile (0^o position). The shifts in Luminance and Hue at different tiles are measured by rotating the spectrophotometer to the center of each display.

2.5 Conclusion

The variations in luminance and hue of the Microtile display can be summarized as a decrease in luminance for all three primaries with the red primary decreasing at a faster rate compared to the blue and green primaries. This results in a shift of the white and other hues within the colour gamut towards blue-green region of the colour space along with a decrease in brightness as the viewing angle increases. These shifts are probably a consequence of maximizing the intensity for a head-on view. Given the nature of the display, the industry is concerned with the acceptability/unacceptability of these colour shifts at different viewing eccentricities. Since, Microtile arrays have a thin black border between each display, any subtle colour difference becomes easier to perceive. Therefore, it is in question if the industrial convention of using $1\Delta E_{Lab}$ as the tolerance limit is applicable to these displays. Given the variability in the perceptibility thresholds (0.5 to 3 ΔE_{Lab} units) and the fact that the industrial convention of unit ΔE_{Lab} for calibration precision was picked based on consensus

between display industries, there is a need to check the precision limits for calibrating the Microtile displays. There is also limited data on how these colour differences relate to the acceptability/unacceptability of a display; therefore, there is still a need for data that is relevant to the Microtile technology.

It is evident from the review in Chapter 1 that the newly developed CIE2000 colour difference equation appears to be among the best in terms of uniformity in the colour space. Nevertheless, we will use the 1976 CIE L*a*b* space and its colour difference equations for a 10 degree observer, as it is currently in practise in almost all display industries. Despite the vast literature on colour discrimination experiments, the threshold values differ based on different field sizes and configurations, which may not be similar to the Microtile viewing conditions. All these emphasize the need to design a perceptibility and acceptability experiment pertaining to the Microtile viewing scenario.

Chapter 3
Perceptibility Thresholds in the Blue-Green region of the L*a*b* - colour space

3.1 Objective

The aim of this study is (i) to design a computer adaptive colour discrimination experiment and (ii) to sample the discrimination stimuli to a specific region of colour space pertaining to the Microtile behaviour discussed in Chapter 2.

3.2 Methods

3.2.1 Experimental Set up

3.2.1.1 Description of monitor settings and graphic system

A Sony Trinitron GDMF520 CRT monitor was used to display the stimulus. The CRT monitor had an aspect ratio of 40 X 30 cms, with a default monitor refresh rate of 100Hz and a spatial resolution of 1024 X 764. The monitor contrast was set at 80 and brightness at 40. The brightness and contrast of the monitor were set so that the monitor black point was less than 1cd/m^2 and any readable text displayed had sharp edges viewed from a distance three times the diagonal of the CRT monitor. The stimulus was generated using the Visual Stimulus Generator (ViSaGe, Cambridge Research System, Version- 8) controlled by a custom written program in Matlab™ (version 2009b). The ViSaGe provides a resolution of 14-bit colour and luminance look up table (LUT) along with integrated support for gamma correction and colour calibration.

3.2.1.2 Stimulus and surround conditions

The stimulus used was a 2x2 array of coloured rectangles displayed on the CRT monitor. The size and viewing distance of the stimulus was set to be equivalent to viewing a section of the 5x5

Microtile array from 7.6 m (three times the diagonal of the array). The rectangles were separated by black horizontal and vertical lines of 2 pixels width (1mm) in order to simulate the seams between the Microtiles. Table 3.1 lists the complete stimulus profile. The three reference test colours selected were a green, white and flesh-tone (skin-tone). The chromaticity co-ordinates of the reference colours were based on the average measurements from maple leaves for the green reference colour; the arm colour from a set of Caucasians for the skin-tone; and the media white point of the CRT monitor was selected for the white reference colour. The reference colours were also based on the commonly used colours in natural scenes suggested by Christie Digital.

Table 3.1 Stimulus parameters

Viewing distance	3.75m
Stimulus presentation duration Inter-stimulus duration	3 seconds maximum† 5 seconds minimum‡
Inter-stimulus screen chromaticity co-ordinates and luminance / Adaptation screen	x_{10}: 0.38, y_{10}: 0.38, L_{10}: 38.5
Media white point	x_{10}: 0.3132; y_{10}: 0.3250; L_{10}: 109.2102 (10^0 observer)
Surround field	x_{10}: 0.421; y_{10}: 0.38; L_{10}= 42.32 (+/- 10% variability)
Reference colours:	White x_{10}: 0.3132; y_{10}: 0.3250; L_{10}: 75 (70% of the luminance of media white point)

	L^*_{10}= 86.3428, a^*_{10}= 0, b^*_{10}= 0. Skin-tone x_{10}: 0.399; y_{10}: 0.378; L_{10}: 33 (30% of the luminance of the media white point) L^*_{10}= 61.841, a^*_{10}= 10.339, b^*_{10}= 25.602. Green x_{10}: 0.367; y_{10}: 0.482; L_{10}: 16 (15% of the luminance of the media white point) L^*_{10}=45.1519, a^*_{10}= -19.908, b^*_{10}= 36.342.

†the stimulus presentation time was set at 3 seconds based on pilot experiments using both naïve and experienced observers. Three seconds was a sufficient time to make a discrimination decision and to register a response.

‡the inter-stimulus duration is the time period between each stimulus presentation. It was set at 5 seconds to minimize adaptation effects of the stimulus. This time was based on pilot work in which uniform saturated blue and red screens were displayed and the duration of the afterimages on a white screen were determined.

The surround field around the monitor was rectangular and had an outer extent of 12^o X 10^o from the viewing distance. A stage light with daylight filters was used as an indirect source to illuminate the surround. The luminance of the surround was adjusted to be 50% of the luminance of the white test stimulus. The mean luminance of the surround field was 42.3 cd/m^2 with a +/- 10% variability across the field and was visually uniform. A black light baffle of 20 cms length was placed between the monitor and the surround to reduce the stray light onto the monitor. With the baffle, the light on the monitor due to the surround illumination was reduced to < 1.5 cd/m^2.

3.2.1.3 Calibration routine

The simulations on the CRT monitor were aimed to estimate thresholds, therefore precision and uniformity of the display were necessary. Calibration was performed using the ColourCal™ (Minolta, Japan) which was compatible with the vsgDesktop display calibration software. The Gamma correction software of the vsgDesktop was initialized and the options were set to a linear interpolation gamma fit type, with low bias and to a total of 256 measurements for each phosphor in each quadrant. After setting up the interface, the ColourCal was placed on the center of a given quadrant. The room was maintained dark while calibrating. The gamma calibration was run for each of the four quadrants of the CRT monitor and the LUTs generated were stored. The details of calibration assumptions are given in Appendix 1. The phosphor chromaticity co-ordinates and luminances for each quadrant were measured using the SpecBos™ (Jeti, Germany) for a 10^{o} observer and stored separately for each quadrant. After calibration, the monitor was checked to make sure that the white reference colour appeared uniform across each quadrant. Using individual LUT for each quadrant resulted in noticeable differences between quadrants and so various modifications were tried in order to obtain uniformity across quadrants for each reference colour. The modification that showed satisfactory uniformity was when using an average LUT with phosphor luminance and chromaticity coordinates unique for each quadrant.

3.2.1.4 Producing gamma corrected colour on the CRT monitor

The chromaticity co-ordinates and the luminance (x_{10}, y_{10}, L_{10}) of the display colour were entered in the Graphic User Interface. The input co-ordinates were converted to the tristimulus values X_{10} Y_{10} Z_{10}. The tristimulus values of the input colour were then transformed to phosphor luminances (L_R, L_G, L_B) for each quadrant. For each colour, the ratio of the phosphor luminance of the input colour (L_R, L_G, L_B) to the maximum luminance of each phosphor (measured and stored for each quadrant, L_{Rmax},

L_{Gmax}, L_{Bmax}) were calculated to give a normalized luminance value for each phosphor. This value was then multiplied by 16384 (2^{14} – for a 14 bit resolution) to give the index to the LUT. The index value specifies the digital video input to each pixel, which in turn generates the luminance required to produce the desired colour. Figure 3.1 is a schematic of the steps involved to produce the desired colour in each quadrant.

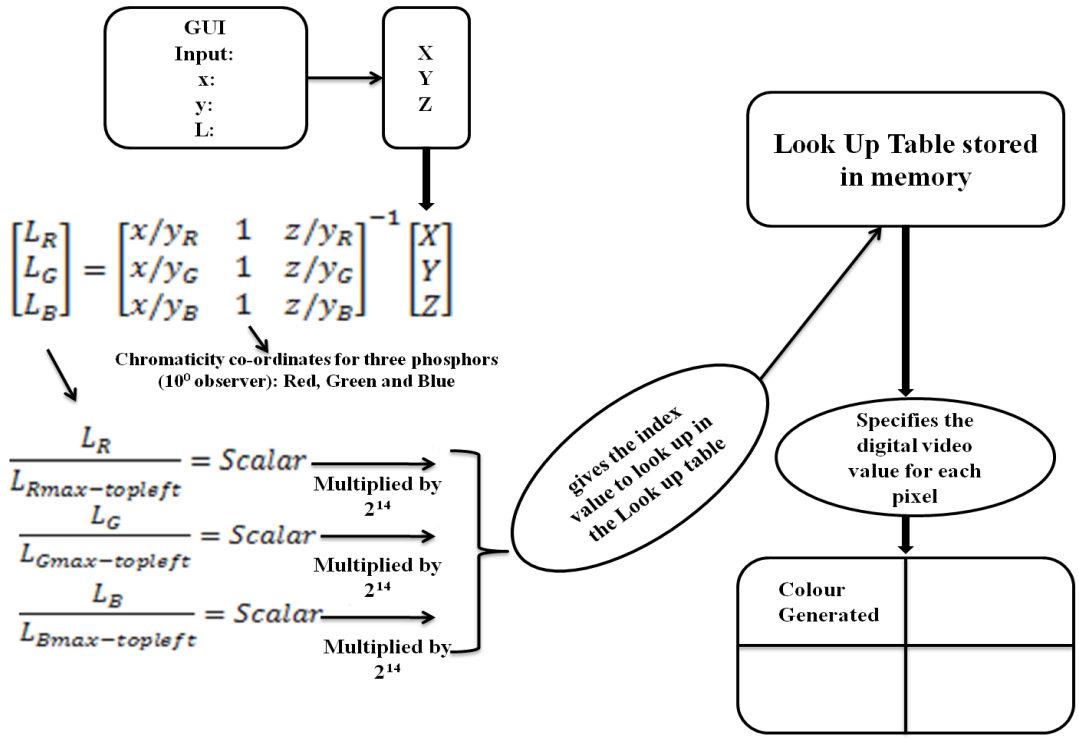


Figure 3.1 Schematic of the steps involved in generating colour on the CRT monitor. Shown above are the essentials of the code built for generating a colour in the top left quadrant. The calculation for a single quadrant and a single input colour is shown; the same steps are involved in producing an output for each quadrant.

3.2.1.5 Sampling stimulus

All the colours were represented in CIE1976 L*a*b* space (10 degree observer) in accordance with the industry convention. A ΔE_{Lab} is defined as the colour difference in an L*a*b* space between the

reference and test colours and is given by $\Delta L^2 + \Delta a^2 + \Delta b^2 \ ^{1/2}$, where ΔL is the difference along luminance dimension, Δa is the difference in hue in the '*a'* direction (red-green), and Δb is the difference in hue along '*b'* (blue-yellow) dimension (Berns, R.S. 2000). Though the L*a*b* space is a non-linear transformation of the CIE1964 xyL space, the colour difference within the space is Euclidean.

The direction of vectors around each reference colour that were sampled in this study was confined to the blue-yellow region (-b to +b axis of the L*a*b* space) of the colour space owing to the observed trends with the Microtile display. Twelve different vectors were sampled, of which five were on the equi luminance plane, five were sampled as projection vectors of the equi-luminance vectors on to the decreasing luminance plane (at 45^o inclination to the equi-luminance plane), and two vectors were sampled along the increasing and decreasing luminance axis. A schematic of the sampling vectors in the L*a*b* space is given in Figure 3.2-A. Assuming the perceptibility thresholds to be equal in all directions, the orientation and the length of the vectors were specified in a spherical polar co-ordinate system (see Appendix 2 for transformations). The length of each vector corresponds to the colour difference (ΔE_{Lab}). The orientation of each vector was specified in terms of two angles namely; the hue angle θ with respect to +a (red) direction, and the Luminance angle ϕ with respect to the increasing luminance direction (+L) as in Figures 3.2-B and C.

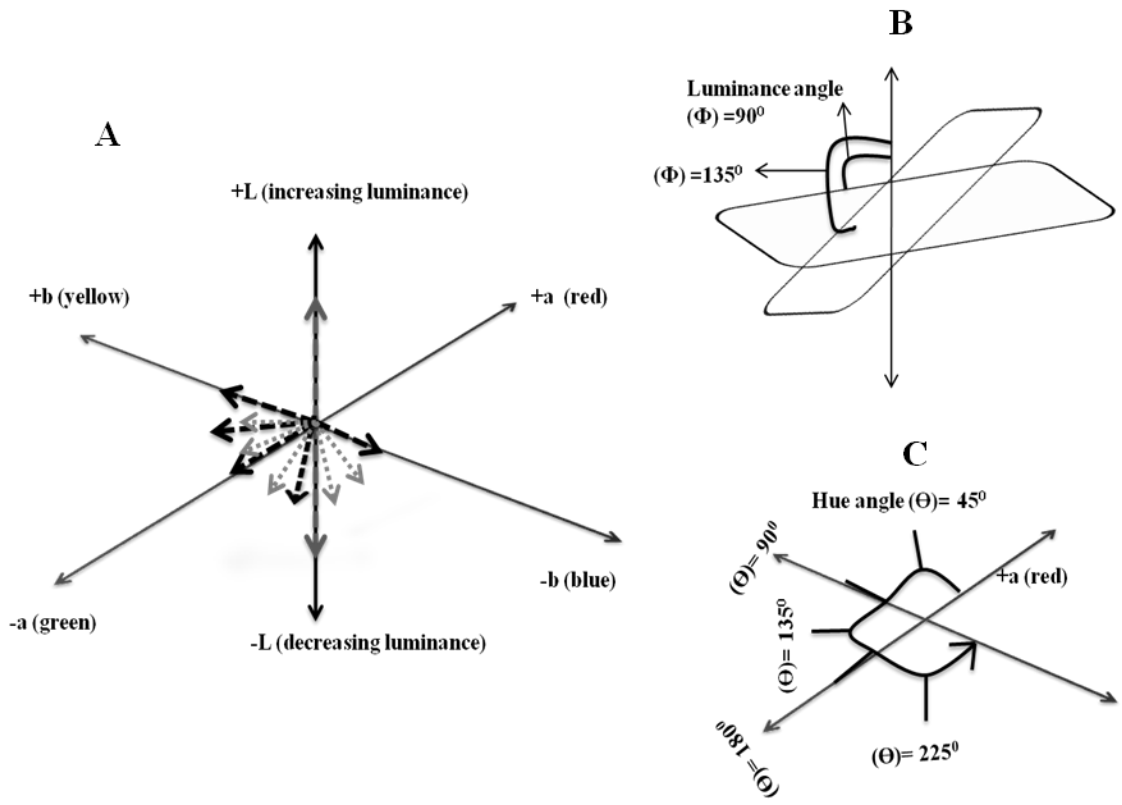


Figure 3.2 Sampling vectors in L*a*b* space. Sampling vectors along the blue-yellow region of the colour space. (A) shows the sampling space with five vectors on the equi-luminance plane (dark hyphenated lines), five projection vectors on to the decreasing luminance plane (light hyphenated lines), and two vectors along the equi-hue plane (B) shows the Luminance angle with respect to the increasing luminance axis (C) shows the hue angle with respect to the +a axis.

3.2.2 Selection of Adaptive psychophysics technique

Adaptive psychophysics has gained popularity, due to its efficiency in estimating thresholds (Kujala, J.V. 2006, Gescheider, G.A. 1997). Non-adaptive techniques use predetermined set of stimulus values that are independent of subject's response, while adaptive techniques are dependent on subject's response. Despite the variety of adaptive psychophysical methods, and the knowledge about each of

the techniques, it is often difficult to choose the best method for a given application (Treutwein, B. 1995).

Commonly used adaptive techniques can be divided in to three broad categories, namely; (1) the adaptive staircase (or up-down) method (2) Parametric Estimation by Sequential Testing method (PEST) and (3) function-constrained methods which are based on the maximum-likelihood parameter estimation (Watson, A.B. 1983). Of these methods, the maximum likelihood technique is widely used for detection or discrimination tasks (Harvey, L.O. 1986; Strasburger, H. 2001, Nachmias, J. 1981) and is also known for its accuracy (Wichmann, F.A. 2001).

A modified Quantitative estimation by sequential testing (QUEST) technique called the ZEST (zippy estimation by sequential testing) was used in this experiment. The difference between the QUEST and the ZEST protocol is in the location of the stimulus for the proceeding trial. The ZEST uses the mean of the posterior probability distribution function (pdf) as the stimulus for the subsequent trial, whereas, the QUEST uses the mode. The ZEST protocol has been shown to outperform the QUEST-mode and the QUEST-median in terms of efficiency and accuracy in estimating thresholds for yes-no type threshold experiment. (King-Smith, P.E. 1994).

The shape of the psychometric function is assumed to be a constant Weibull function in the ZEST paradigm. Fixed parameters are the slope of the function and the guess rate which depends on the psychophysical method used (2AFC has a guess rate of 50% whereas a 4AFC has a guess rate of 25%). The variable parameter is the stimulus value corresponding to the maximum likelihood threshold. This variable parameter is updated for each trial by Baye's rule. The posterior probability distribution function is updated after each trial and depends on the observer's response to the preceding trial (Harvey, L.O. 1986).

The crsPsychMethod ZEST was the subroutine used, which calls for all the fixed and varying parameters, and returns a threshold value for a given direction. Table 3.2 lists the set of parameters entered to initialize the ZEST paradigm.

Table 3.2 ZEST parameters

Experimental paradigm	*'4AFC'*, where 4-AFC is 4-alternative-forced-choice paradigm.
Fixed parameters (Default settings)	
Beta value (slope of the psychometric function)	3.5
Gamma value (guess rate)	0.25
Epsilon value (threshold criterion factor)	1.5 (0.17 log units)
Variable parameter	
Alpha (threshold)	Is updated by Baye's rule after every response.

Modeling the experiment using either QUEST or ZEST showed that over 200 trials would be necessary to obtain a single threshold value if the protocol started at an easily seen supra-threshold value. In order to improve the efficiency, we first started out with a method of descending limits in order to estimate the starting intensity for the ZEST protocol which would be much closer to the actual threshold value. The program started with displaying one of the quadrants with a ΔE_{Lab} of 4 from the other three colours. If the quadrant was correctly identified as being different, then a quadrant with a ΔE_{Lab} of 2 and then 1 were presented. The starting ΔE_{Lab} value for the ZEST was then the smallest ΔE_{Lab} that was correctly identified in the initial set of trials. This initial threshold estimation helped to reduce the number of presentations significantly (by a factor of about 2.25). Once this seed value was fixed, an array of 100 entries was generated for each vector ranging from a

‘0’ to the “seed value’, the first colour difference presented for the actual experiment is then the mean of this array.

The ZEST protocol also calls for a predefined termination rule. The termination rule can be defined by a fixed number of trials or by specifying a cut off standard deviation, so that when the standard deviation of the estimated threshold falls below this specified value, the experiment terminates. Setting a termination rule based on the standard deviation does not make any assumptions on the form of the final probability distribution function (King-Smith, P.E. 1994) and so the threshold value is not influenced by the assumed slope of the psychometric function. The termination rule was set to a standard deviation of < 6% of the mean threshold. This was decided based on the step size calculated from the initial array and the precision of the calibrated CRT monitor. For two vectors the number of entries in the array (from 0 to the seed value) is 200; therefore, each adjacent entry in the array differs by a ΔE of 0.05; a six step sizes yields a difference of 0.30, which is close to twice the smallest difference obtainable using a 14 bit resolution and is the physical limit of spatial uniformity achieved with the calibrated CRT monitor (precision limit for producing a uniform colour).

3.2.3 Colour discrimination task and procedure

The subjects head was positioned using a head and chin rest so that their eye level was approximately in line with the center of the monitor. Subjects viewed the stimulus binocularly and eye movements were allowed during the experiment. Observers were directed to view the four quadrants of the CRT monitor and were instructed to identify the quadrant that appeared different from the rest. Observers were provided with a response box to register their response. A schematic of the entire experiment is given in Figure 3.3

There were 3 reference colours and 12 directions for each colour. A session consists of a reference colour and 12 testing directions. There were six runs within each session with thresholds

determined for two vectors within each run. The vector presented at each trial within a run was randomly determined. The order of the vectors for each reference colour was determined by a random block design as was used for the order of the reference colours. For most subjects, the session consisting of 12 vectors lasted for less than 2 hours. If the session exceeded 2 hrs, then the remaining trials were scheduled for another day/time, in order to avoid fatigue. Because the majority of the subjects were naïve to this kind of task, a short training session for 30 minutes on 2 vectors for the white reference colour were carried out before commencing the actual experiment.

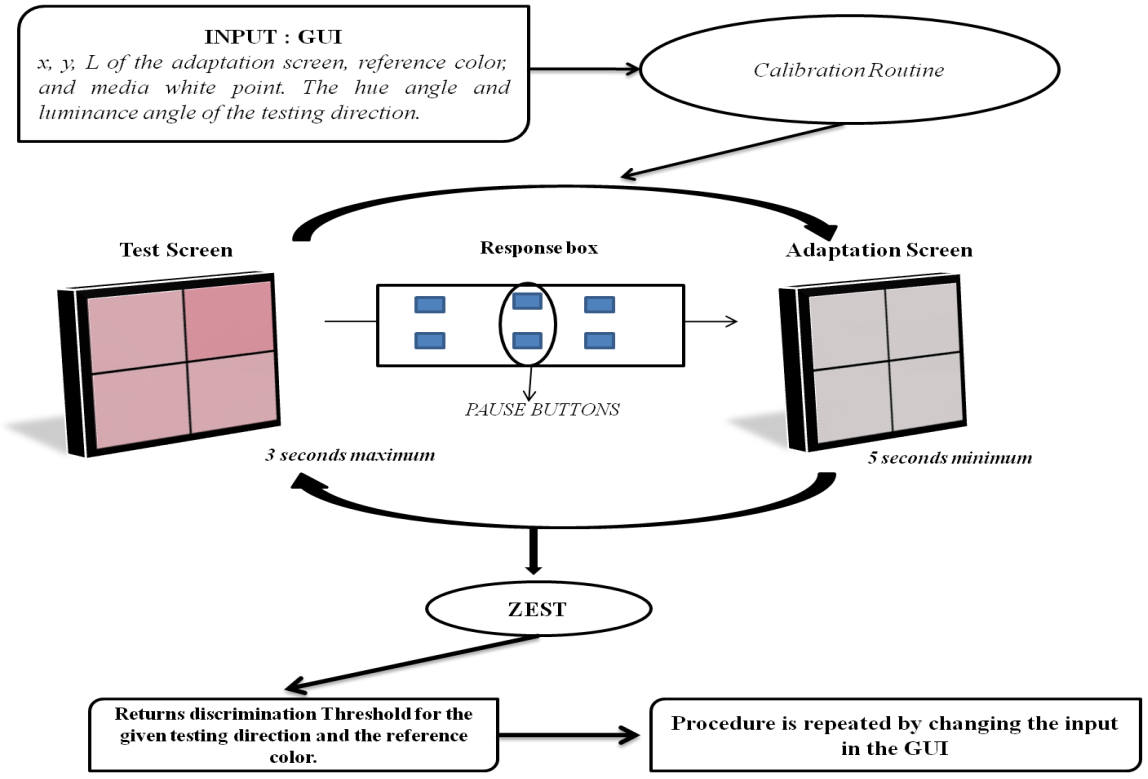


Figure 3.3 Illustration showing the routine in the procedure

3.2.4 Subjects selection and sessions

Twenty subjects were recruited between the age group of 15 to 35 years. All subjects met the inclusion criteria listed in Table 3.3. The experiment was reviewed and approved by the Office of Research Ethics, University of Waterloo.

Table 3.3 Inclusion criteria

Clinical test	Inclusion Criteria
Visual Acuity	Binocular visual acuity 6/7.5 or better acuity with the high contrast Bailey-Lovie chart
Ishihara Colour Vision Screening plates	<4 incorrect responses
Hardy Rand Rittler (HRR) plates 4th edition	No errors on plates 5 to 10
Farnsworth-Munsell 100 Hue discrimination test	Total error scores less than or equal to the average for their age group (Kinnear, PR. 2002).

3.3 Results and Discussion

Figures 3.4 through 3.6 shows the mean discrimination threshold for each colour plotted as a function of hue angle. Twelve directions are labelled from 1 to 12 in an anti-clockwise direction, starting from +b (yellow direction) axis. One to five represents vectors on equi-luminance plane and six to ten represents the projection vectors of one to five on to the decreasing luminance plane (luminance angle: 135^0), and eleven and twelve represent vectors on the equi-hue plane. The labels are colour coded to indicate the hue in the direction of the vector from the origin (reference colour). The bumpy

nature of the thresholds at different hue angles show that the contours of thresholds in a three dimensional space are not spherical, which is consistent with previous studies cited in Chapter 1.

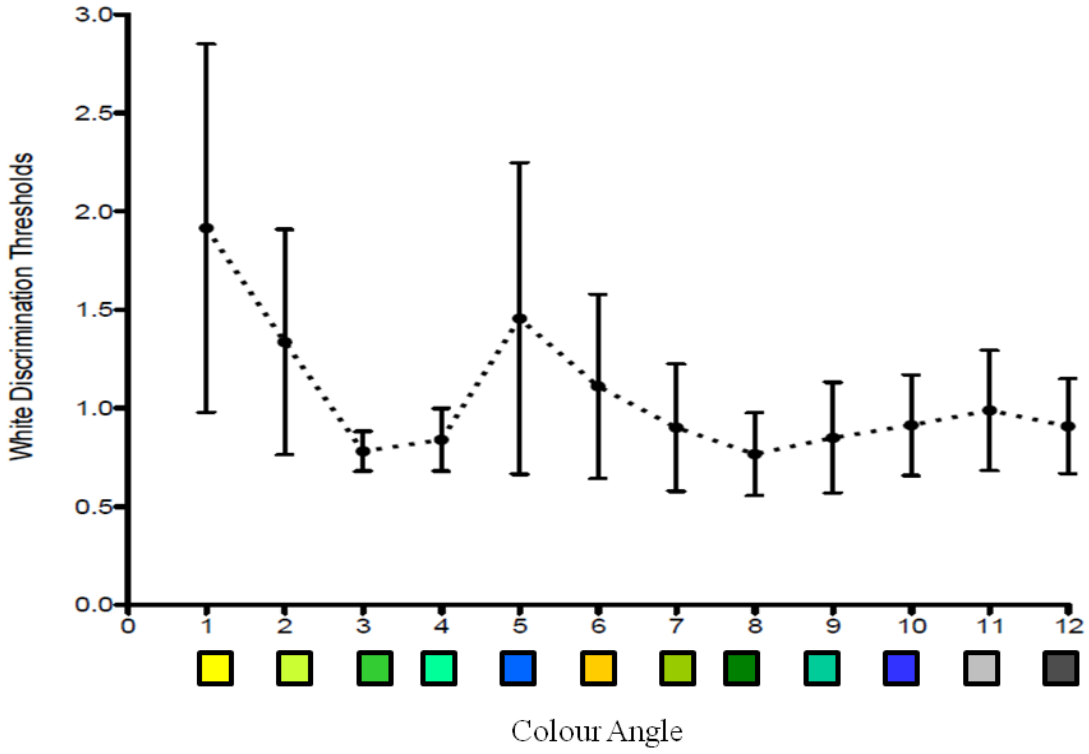


Figure 3.4 Colour difference thresholds from a white reference as a function of colour angle. One through five are vectors on the equi-luminance plane, 1- $\theta=90^0$, 2- $\theta=135^0$, 3- $\theta=180^0$, 4- $\theta=225^0$, 5- $\theta=270^0$; six through ten are projection vectors with luminance angle $\phi=135^0$, 6- $\theta=90^0$, 7- $\theta=135^0$, 8- $\theta=180^0$, 9- $\theta=225^0$, 10- $\theta=270^0$; Luminance vectors 11- $\phi=0^0$ and 12- $\phi=180^0$.

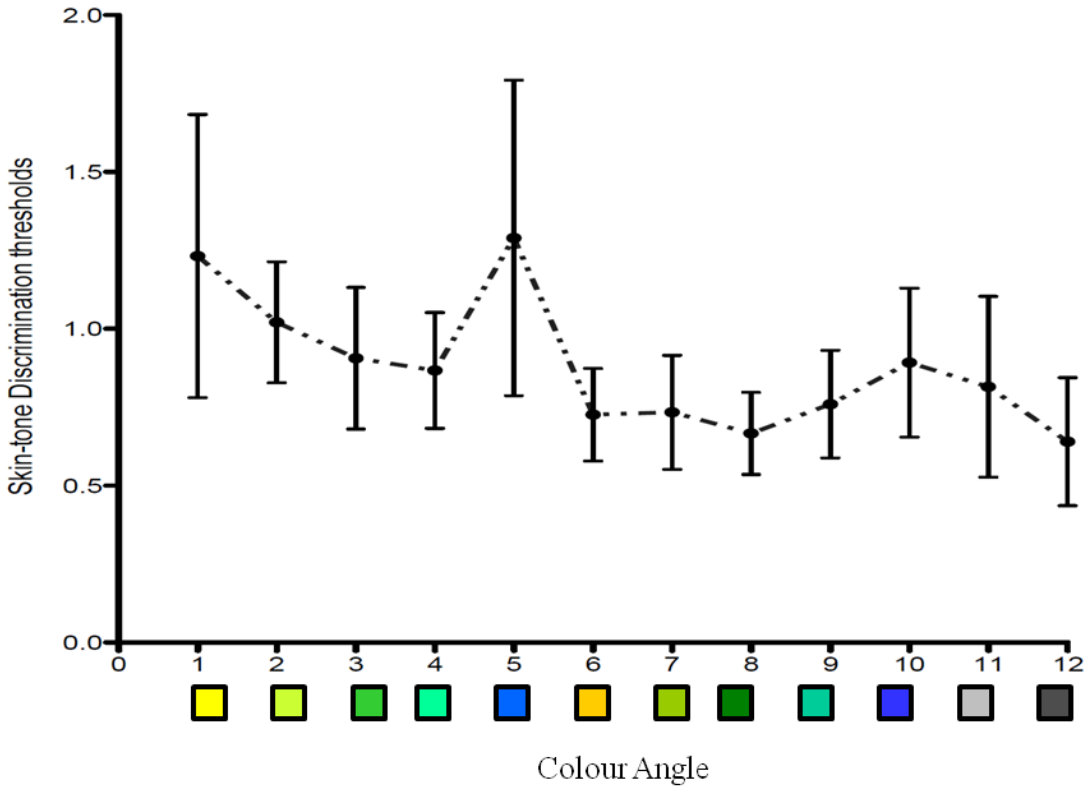


Figure 3.5 Colour difference thresholds from a skin-tone reference as a function of colour angle. One through five are vectors on the equi-luminance plane, 1- Θ=90^0, 2- Θ =135^0, 3- Θ =180^0, 4- Θ =225^0, 5- Θ =270^0; six through ten are projection vectors with luminance angle ϕ =135^0, 6- Θ =90^0, 7- Θ =135^0, 8- Θ =180^0, 9- Θ =225^0, 10- Θ =270^0; Luminance vectors 11- ϕ =0^0 and 12- ϕ =180^0.

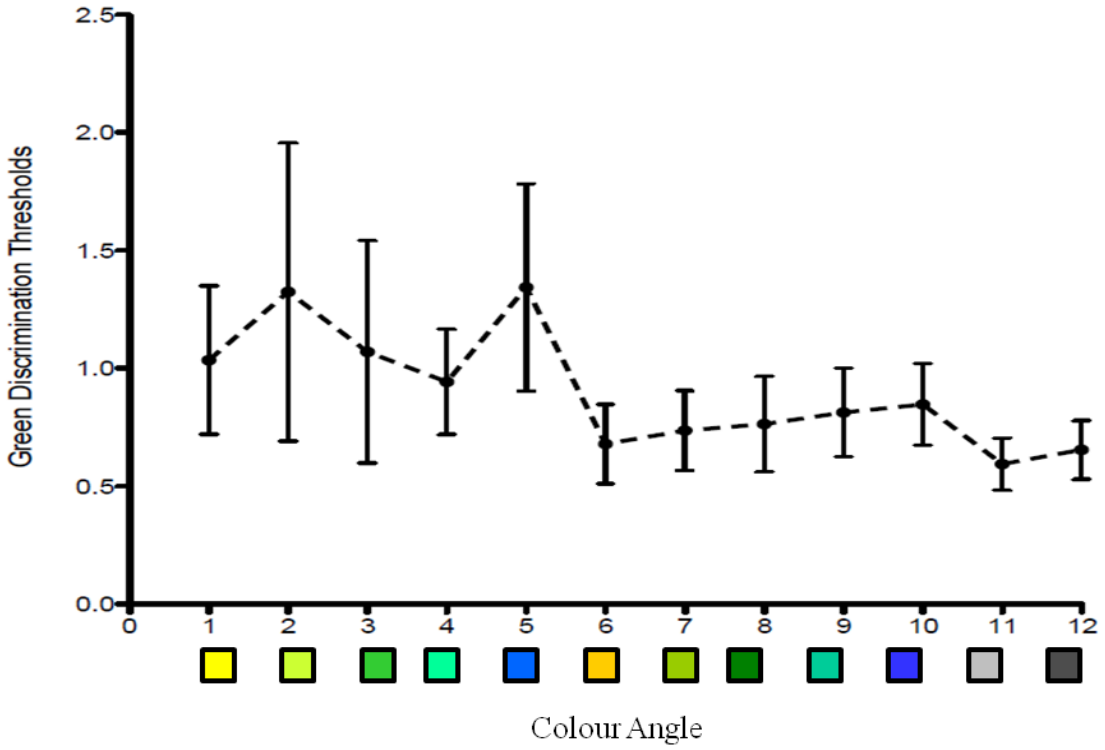


Figure 3.6 Colour difference thresholds from a green reference as a function of colour angle. One through five are vectors on the equi-luminance plane, 1- θ=90^0, 2- θ =135^0, 3- θ =180^0, 4- θ =225^0, 5- θ =270^0; six through ten are projection vectors with luminance angle ϕ =135^0, 6- θ =90^0, 7- θ =135^0, 8- θ =180^0, 9- θ =225^0, 10- θ =270^0; Luminance vectors 11- ϕ =0^0 and 12- ϕ =180^0.

All statistical analysis was done using the Statistical Package for Social Sciences (SPSS^{TM}, version 19). A Repeated Measures Analysis of Variance (RMANOVA) was carried out, with colour and angles as with-in subject factors. Results showed significant main effects of reference colours {F (2, 38) =11.797, $p \leq 0.05$} and angles {F (11, 209) =32.226, $p \leq 0.05$}; along with significant interactions between reference colours and angles {F (22, 418) = 5.661, $p \leq 0.0.5$}. A summary of all mean values calculated for different reference colours and for vectors on different reference planes are given in Table 3.4.

Table 3.4 Mean and standard deviation of the thresholds on the equi-luminance plane, equi-hue plane, and decreasing luminance plane.

	Average Vectors on Equi-Luminance plane		Average Vectors on Decreasing luminance plane		Average Vectors on Equi-Hue plane		Average of all Vectors	
Reference colours	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
White	1.2656	0.4793	0.9085	0.3235	0.9487	0.3118	1.064	0.2465
Skin-tone	1.0633	0.2293	0.7366	0.1324	0.7278	0.2245	0.8792	0.1513
Green	1.1434	0.3048	0.7493	0.1213	0.6245	0.0895	0.9006	0.1686

Further, pair wise comparisons of the main effects were carried out with Bonferroni correction ($p \leq 0.017$ for the reference colours and $p \leq 0.0007$ for the different angles) to interpret the significance of the main effects and its interactions. Between the reference colours, the mean threshold for the white reference colour was significantly higher ($p \leq 0.017$) than the mean thresholds of the skin-tone and green, whereas there was no significant difference for the mean thresholds between the skin-tone and green reference colours. It is evident from the results in Table 3.4 and Figures 3.4 to 3.6 that a significant angle effect in the RMANOVA was due to higher mean thresholds for the vectors on the equi-luminance plane compared to the average thresholds for vectors that vary in luminance and/or hue (non equi-luminance vectors). Also the mean thresholds between the equi-luminance and the non equi-luminance planes within each reference colour were statistically significant ($p \leq 0.0007$) except for the white. The interactions between the two main effects namely, the reference colours and the angles can be summarized by the percentage change in the mean

thresholds between different planes within each reference colours. The mean threshold for the white reference colour on the equi-luminance plane was 39% higher than the mean thresholds on the decreasing luminance plane (vectors that change in both chromaticity and luminance) whereas, the thresholds on the equi-luminance plane for the skin-tone and green reference colours were 44% and 52% higher than the respective average thresholds on the decreasing luminance plane. Relative to the mean thresholds on the equi-hue plane (vectors that decrease and increase in only luminance), the equi-luminance thresholds were 3**3**% higher for the white reference, 46% higher for the skin-tone and 84% higher for the green reference. The percentage change between the mean thresholds on the equi-hue and decreasing luminance planes were minimal. The significant difference between mean thresholds on different planes suggests that the discrimination thresholds have asymmetries three dimensionally and is also suggestive of asymmetries at different regions of the chromaticity space.

Although the vector directions were limited to only half the hue circle, thresholds along the blue-yellow directions were relatively higher than in the red-green direction (particularly noted with the white and skin-tone reference colours). However, comparison of mean thresholds for the different vectors within a given plane showed no significant difference (except for complexities within the equi-luminance plane) in the thresholds measured within a given plane ($p>0.0007$). Nevertheless, the trend shows that the discrimination ellipsoids are oriented with their major axis along the blue-yellow dimension, which is expected for the sampled age group.

In terms of the Microtile colour shifts shown in Chapter 2, the mean thresholds for the vectors heading towards the blue green region of the colour diagram and decreasing in luminance (vectors labelled 8, 9 in figures 3.6, 3.7, and 3.8) were less than ΔE_{Lab} of 1 for all three reference colours. This shows that using a tolerance of 1 ΔE_{Lab} for the Microtile display may be too liberal particularly for non-white colours and discerning viewers.

3.4 Conclusion

The industrial standard limit for displays has traditionally been 1 unit of ΔE in the CIE 1976 L*a*b* space. This rule is applicable, when considering the mean of all possible directions in colour space; however, this study focuses on thresholds from vectors sampled in a direction specific to the Microtile displays. It is found that the discrimination thresholds (average of non equi-Luminance vectors) are less than 1 ΔE_{Lab} for the reference colours - 0.9286 (white), 0.7322 (skin-tone) and 0.6869 (green). Therefore, the company may want to use a precision value close to threshold in order to cut down the noticeable colour shifts on the Microtile displays.

In more general terms, the results confirm that the L*a*b* space is not a perceptually uniform space and that thresholds in the equal luminance plane are slightly higher than colour difference thresholds that include a change in luminance along with a change in chromaticity. The sampled vectors in this experiment were along a given region in the space reflecting the characteristics of one particular Microtile model and may not generalize to other directions in colour space. Given the rapid developments in display technologies, it would be important to determine whether the conclusions of this study carry over to a wider selection of colour directions. This question is addressed in Chapter 4.

Chapter 4

Perceptibility Thresholds – Expanded

4.1 Objective

The objective of this study is to determine the colour difference thresholds in eighteen different directions for four reference colours in the L*a*b* space.

4.2 Methods

4.2.1 Experimental Setup

The experimental set up was exactly the same as adapted for the first experiment and is discussed in detail in Chapter 3. Prior to the commencement of the experiment, the media white point of the CRT monitor was verified. The chromaticity co-ordinates and luminance of the surround field were also measured to ensure stability of the reference viewing conditions. The stimulus and the program used were the same as used in the first part of the experiment (for a detailed explanation see Chapter 3).

4.2.2 Stimulus Sampling

The discrimination task is nearly the same procedure presented in Chapter 3. The exceptions were the number of reference colours and the sampling directions. The reference colours picked for this experiment are the same set of colours used for the previous study in addition to a new colour sampled from the blue region of the colour space. The goal was to sample at least one colour from each region of the colour space. Because some of the other reference colours were based on natural scenes, the blue chromaticity co-ordinates were based on the chromaticity coordinates of the sky measured on a clear sunny day. The chromaticity co-ordinates and the luminance of the reference colours used are tabulated in Table 4.1.

Table 4.1 Chromaticity co-ordinates of the reference colours used in this experiment

Reference colours	Chromaticity co-ordinates
White	x_{10}: 0.3132, y_{10}: 0.3250, L_{10}:75; L^*_{10}= 86.3428, a^*_{10}= 0, b^*_{10}= 0.
Skin-tone	x_{10}: 0.399; y_{10}:0.378; L_{10}:33; L^*_{10}= 61.841, a^*_{10}= 10.339, b^*_{10}= 25.602.
Green	x_{10}: 0.367; y_{10}:0.482; L_{10}:16; L^*_{10}=45.1519, a^*_{10}= -19.908, b^*_{10}= 36.342.
Blue	x_{10}: 0.253; y_{10}:0.266; L_{10}:10; L^*_{10}= 36.284, a^*_{10}= -0.984, b^*_{10}= -15.821.

A total of eighteen directions were sampled with the vectors equally spaced across all directions in the CIE1974 L*a*b* space. Out of the eighteen vectors, eight vectors were sampled on the equi-luminance plane, two vectors on the equi-hue plane, four vectors were projection vectors onto the increasing luminance plane and four vectors were projection vectors onto the decreasing luminance plane. The vectors on the equi-luminance plane head from a hue angle of zero corresponding to +a (red) direction and are sampled every 45^o in the anti-clockwise direction. The 45^o vectors on each quadrant of the equi-luminance plane were projected onto the increasing and decreasing luminance plane. The sampling directions with the hue and luminance angles are illustrated in Figure 4.1.

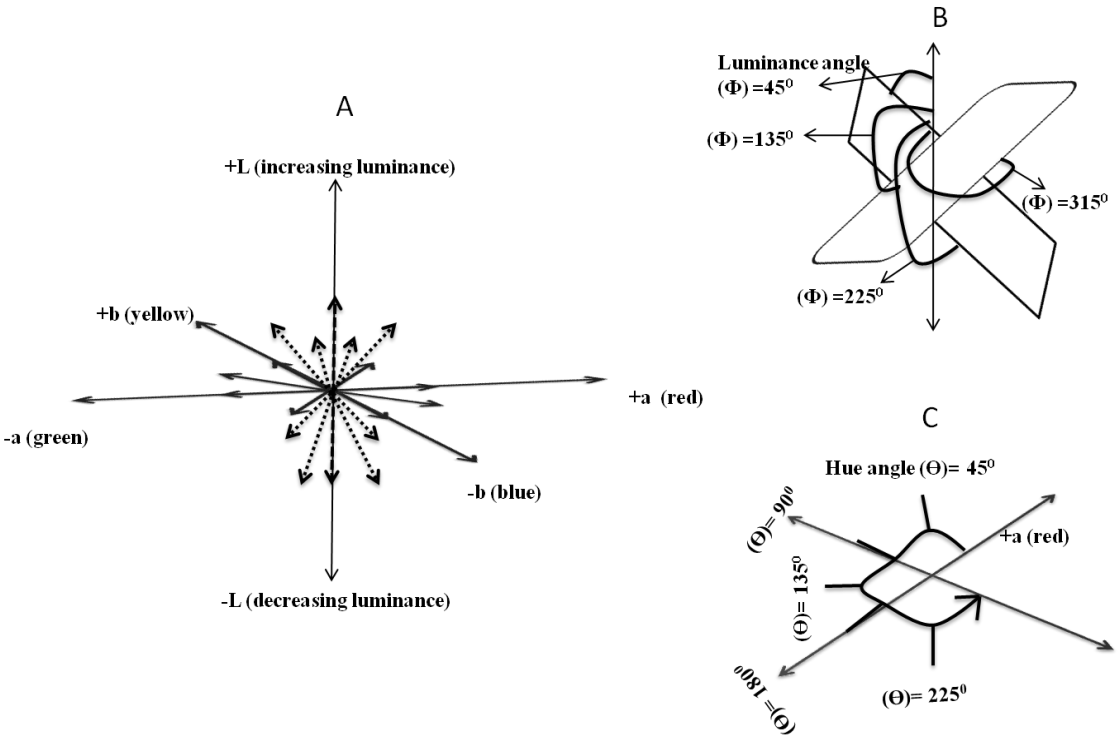


Figure 4.1 (A) Sampling vectors in the L*a*b* space. The dotted vectors represent the projection vectors. (B) Luminance angles 45^{o}, 315^{o}- corresponds to an increasing luminance direction and 135^{o}, 225^{o} – represents the decreasing luminance direction. (C) Hue angles are measured with respect to the +a axis in an anti-clockwise direction.

4.2.3 Subjects

Twenty subjects were recruited for this experiment. Eighteen were undergraduate student volunteers, who had no prior experience and were naïve to this task. The other two subjects were graduate students who participated in the first part of the experiment and therefore they were well trained. All subjects were between the age group of 20 to 30 years of age. The inclusion criteria were the same as listed in Table 3.3. Although the majority of subjects were naïve, their colour discrimination was better than average based on the Farnsworth-Munsell 100 Hue Scores. With this selection criterion, a

subject pool with normal colour vision and who were, presumably, able to make more discerning judgments regarding colour differences were recruited for the study.

4.2.4 Sessions

The order of presentation was determined using a random block design. There were 4 reference colours and 18 directions for each colour. A session consisted of a reference colour and 18 testing directions. There were nine runs within each session and 2 vectors within each run. An initial demonstration session (for 5 minutes) was carried out prior to the actual experiment. The demonstration was to familiarize subjects with the task and to ensure that the subjects understand the procedure. The entire experiment was split into eight visits, each visit consisting of half a session and lasted for a maximum of 2 hours in order to avoid fatigue. The number of trials per session varied from subject to subject depending on the speed of the performance.

4.3 Results

Figures 4.2 to 4.5 shows discrimination thresholds plotted as a function of colour angles. The colour angles are labelled from 1 to 18 indicating eighteen directions. Colour angles marked 1 to 8 are vectors on the equi-luminance plane; angles 9 and 10 are vectors on the equi-hue plane (as increasing and decreasing luminance vectors); angles labelled 11 to 14 are projection vectors on the increasing luminance plane and angles marked 15 to 18 are projection vectors on the decreasing luminance plane. From the figures it is evident that the bumpy pattern of thresholds is retained for all reference colours, with the double bumps being more evident for the white and blue thresholds and is suggestive of an ellipsoidal configuration of thresholds in a three dimensional space. A relatively higher threshold for the vectors labelled 3 and 7 in the figures is suggestive of an orientation towards the blue-yellow region of the colour space.

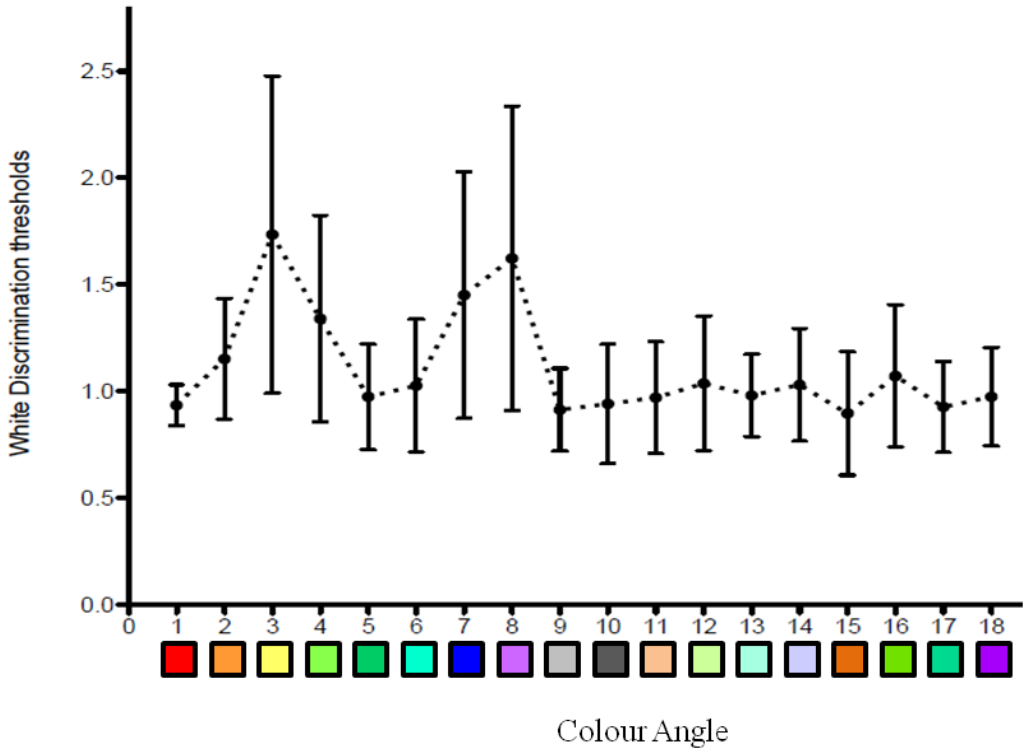


Figure 4.2 Colour difference thresholds from a white reference as a function of colour angle. One through eight are vectors on the equi-luminance plane, 1- θ=0°, 2- θ =45°, 3- θ =90°, 4- θ =135°, 5- θ =180°, 6- θ =225°, 7- θ =270°, 8- θ =315°; 9- ϕ=0° and 10- ϕ =180° are Luminance vectors; eleven through fourteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the increasing luminance plane (ϕ = 45°, ϕ = 315°); fifteen through eighteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the decreasing luminance plane (ϕ = 135°, ϕ = 225°).

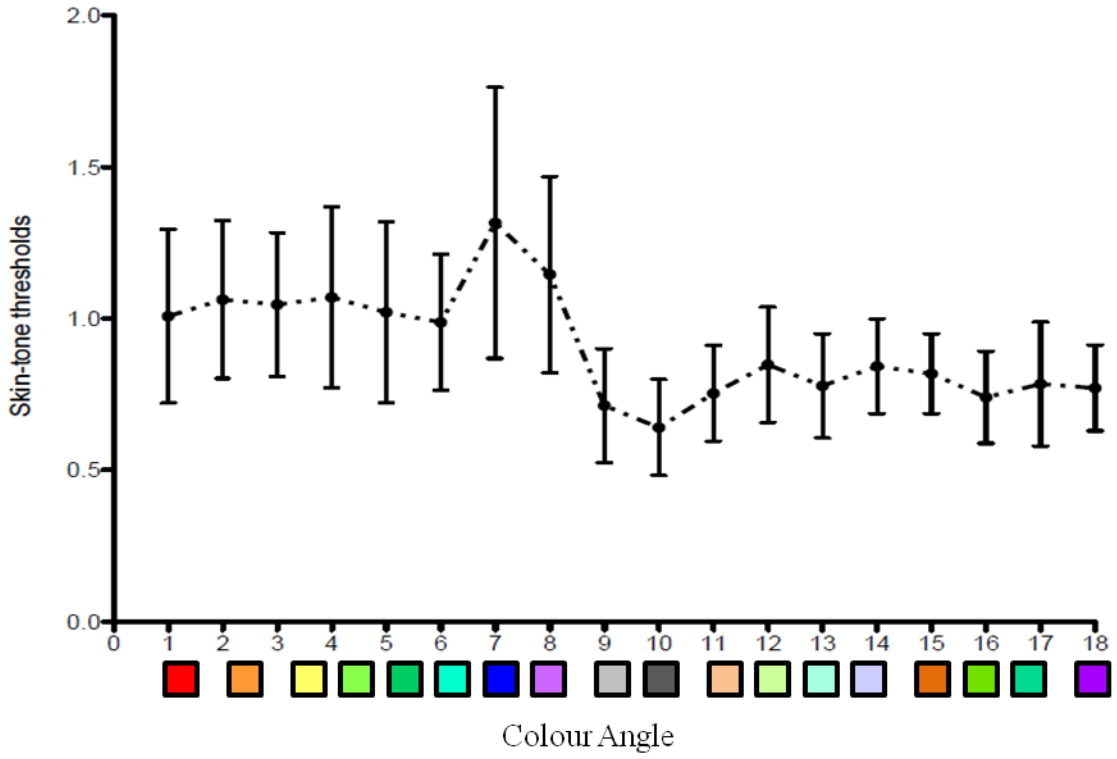


Figure 4.3 Colour difference thresholds from a skin tone reference as a function of colour angle. One through eight are vectors on the equi-luminance plane, 1- θ=0°, 2- θ =45°, 3- θ =90°, 4- θ =135°, 5- θ =180°, 6- θ =225°, 7- θ =270°, 8- θ =315°; 9- ϕ=0° and 10- ϕ =180° are Luminance vectors; eleven through fourteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the increasing luminance plane (ϕ = 45°, ϕ = 315°); fifteen through eighteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the decreasing luminance plane (ϕ = 135°, ϕ = 225°).

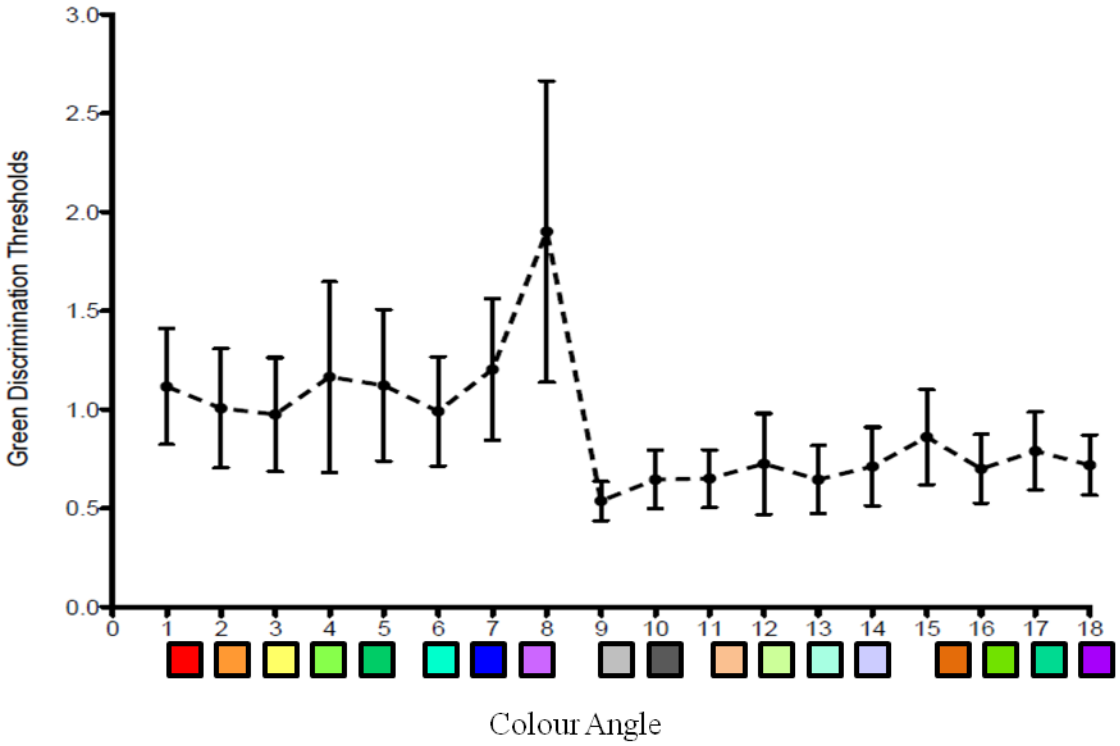


Figure 4.4 Colour difference thresholds from a green reference as a function of colour angle. One through eight are vectors on the equi-luminance plane, 1- θ=0°, 2- θ =45°, 3- θ =90°, 4- θ =135°, 5- θ =180°, 6- θ =225°, 7- θ =270°, 8- θ =315°; 9- ϕ=0° and 10- ϕ =180° are Luminance vectors; eleven through fourteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the increasing luminance plane (ϕ = 45°, ϕ = 315°); fifteen through eighteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the decreasing luminance plane (ϕ = 135°, ϕ = 225°).

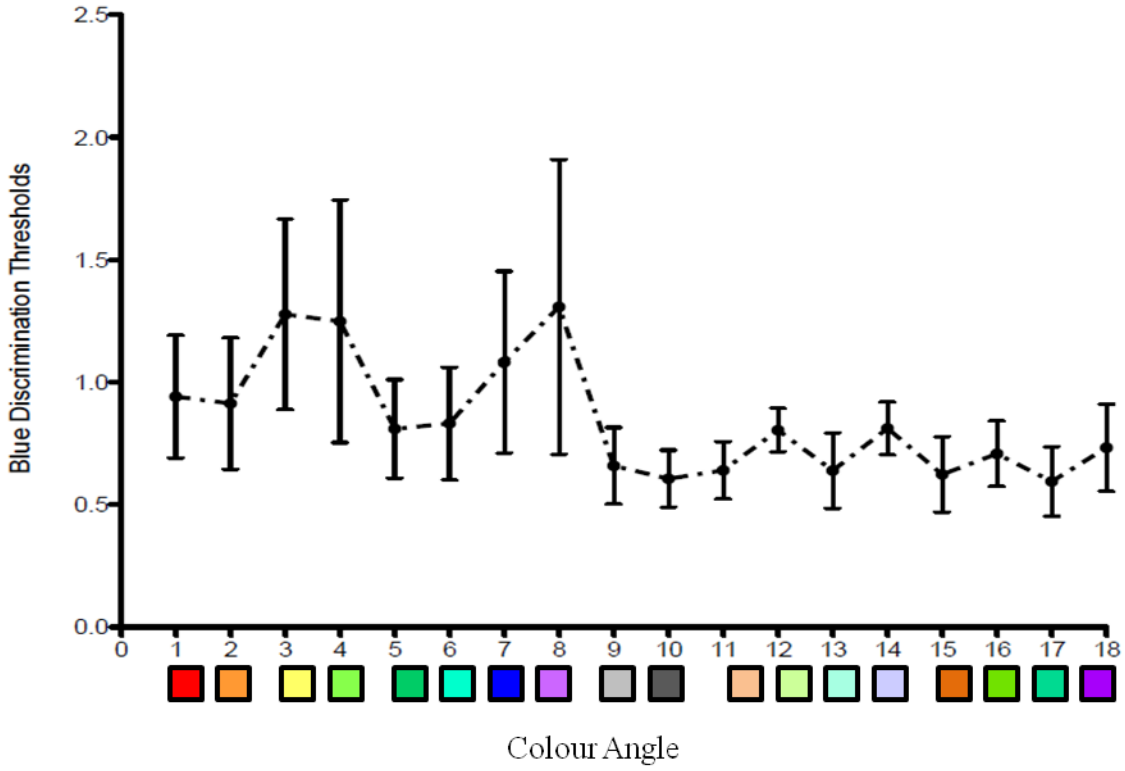


Figure 4.5 Colour difference thresholds from a blue reference as a function of colour angle. One through eight are vectors on the equi-luminance plane, 1- θ=0°, 2- θ =45°, 3- θ =90°, 4- θ =135°, 5- θ =180°, 6- θ =225°, 7- θ =270°, 8- θ =315°; 9- ϕ=0° and 10- ϕ =180° are Luminance vectors; eleven through fourteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the increasing luminance plane (ϕ = 45°, ϕ = 315°); fifteen through eighteen are projection vectors of θ =45°, θ =135°, θ =225°, θ =315° on to the decreasing luminance plane (ϕ = 135°, ϕ = 225°).

A Repeated Measures Analysis of Variance (RMANOVA) was used for the analysis and showed significant effects of reference colours {F (3, 57) = 29.507}, angles {F (17, 323) = 42.191}, and interactions between the reference colours and angles {F (51, 969) = 4.653} at a $p \leq 0.05$. A summary of the mean and standard deviations of the thresholds on different vector grouping and for all four reference colours are tabulated in Table 4.2.

Table 4.2 Average thresholds in different planes. In particular, the average of vectors on the non equi-luminance plane is of interest.

	Average of Vectors on equi-luminance plane		**Average of Vectors on equi-hue plane**		**Average of Vectors on increasing luminance plane**		**Average of Vectors on decreasing luminance plane**	
Reference colours	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
White	1.2795	0.2055	0.9270	0.1805	1.0046	0.1555	0.9669	0.1725
Skin-tone	1.082	0.1612	0.6769	0.1314	0.8055	0.1367	0.7785	0.1048
Green	1.1849	0.2705	0.5912	0.1051	0.6830	0.1590	0.7672	0.1544
Blue	1.0514	0.1842	0.6525	0.1111	0.7240	0.0823	0.6647	0.1199

Pair wise comparisons of the main effects were carried out with Bonferroni correction ($p \leq 0.008$ for reference colours and $p \leq 0.0003$) to interpret the significance of the main effects and its interactions. Comparison of mean thresholds between the reference colours showed that the thresholds for the white references colour was significantly ($p \leq 0.008$) higher compared to the skin-tone, blue and green reference colours. These observations were similar to that noted in the first part of the perceptibility experiment (Chapter 3). A higher threshold for the white reference colour can be attributed to the fact that the surround colour was close to the white reference colour used and that the subjects were adapted to this surround throughout the experiment. It is evident from table 4.2 and

figures 4.2 to 4.5 that a significant angle effect was due to the higher mean threshold on the equi-luminance plane compared to the mean thresholds on the other planes.

The interactions between the two main effects can be summarized by the percentage change in mean thresholds between different planes for a reference colour. The mean thresholds for the white on the equi luminance plane are 30% higher than the thresholds on the non equi-luminance plane (average of vectors on the increasing and decreasing luminance plane), whereas the mean thresholds on the equi-luminance plane for the skin-tone, green and blue were 37%, 63% and 51% higher than the respective averages on the non equi-luminance plane. Relative to the mean thresholds on the equi-hue plane (vectors that decrease and increase in only luminance), the equi-luminance thresholds were 38% higher for the white reference, 60% higher for the skin-tone, 100% higher the green, and 61% higher for the blue reference colour. The percentage change between the mean thresholds on the equi-hue and non equi-luminance planes were minimal with thresholds on the non equi-luminance plane being higher than the thresholds on equi-hue plane. The significant difference between mean thresholds on different planes is suggestive of asymmetries at different regions of the chromaticity space as observed with the previous results in Chapter 3. These results also shows that the discrimination in the equi-luminance plane is lower compared to the discrimination in the equi-hue and non equi-luminance directions. It is also noted that the percentage change of thresholds across different planes is less for white compared to different colours. Also, comparison of mean thresholds between the equi- luminance and the non equi luminance planes within each reference colour, showed statistically significant difference ($p \leq 0.0003$) except for the white. This is suggestive of a more spherical contour for the white thresholds compared to other colours which is expected for a colour close to the origin in an L*a*b* space and with higher luminance.

4.4 Discussion

Colour difference thresholds estimated from a plethora of visual data has shown wide variability. Observer experience accounts for one of the most important sources of variability. It is possible to select either highly trained observers or naïve subjects. The highly trained subject group would be valuable in understanding the discriminability limits of human colour vision, whereas results from naïve subjects would be more representative of the population from an end user or consumer's point of view. The latter approach was used in this study, as we were ultimately concerned with image quality of a display viewed by the general public.

One of the issues raised in this approach was the validity of the results obtained from a group of naïve observers, in comparison with previous results, particularly ones that used trained subjects. A direct comparison is by the threshold values found in the first and second experiments. Although subjects in both experiments were unfamiliar with colour difference threshold measurements, the first group of subjects had a longer training session and had more practice with visual psychophysical tasks in general. Nevertheless, the thresholds in the first experiment were marginally higher than the thresholds in the second experiment, which used subjects who were less familiar with threshold experiments. Thus, it appears that naïve subjects with good colour discrimination can provide data that is comparable to more experienced subjects.

We were also able to examine the data from the two subjects who participated in both experiments to determine their repeatability. This information is useful because the time consuming nature of threshold experiments have always posed problems in acquiring repeatability measures. There were nine vectors (of which 5 were equi-luminance vectors) and 3 reference colours that were in common to both parts of the perceptibility experiment. A RMANOVA was performed using these common colours and vectors to see the effect of repeating threshold measurements in the two sessions

(session here refers to the first and second parts of the perceptibility experiment). There was no significant main effect of session {F (1, 1) = 1.028, $p > 0.05$} or any significant interaction involving the session term. This shows that these two subjects were consistent in their performance and there was no effect of practice on the thresholds.

Comparisons of the threshold data with previous results are difficult because of differences in stimulus parameters and procedures. Nevertheless, four studies were selected based on either the stimulus size or the use of CRT displays. The characteristics of each of these study data sets are tabulated in Table 4.3. Figure 4.6 shows discrimination ellipses in the CIE 1931 x, y diagram for three data sets. The sizes of the current ellipses are approximately 2 times larger than that reported by Brown (1957), Wyszecki and Fielder (1971) in their data based on the variability of the colour matches. The larger size of the current ellipses is equivalent to about 4 standard deviations from mean colour, instead of 2 standard deviations that was reported as thresholds in Brown's and Fielder's data. Although the colours picked from Brown's and Wyszecki's data were not identical to the reference colours used in the current study, the orientation of the ellipses were similar for almost all reference colours as shown in figure 4.6 (especially the ellipses from Brown's data). The differences in size can be attributed to several factors, including the procedure used in deriving thresholds, the stimulus size, the surround conditions and the number and experience of the observers. These factors are tabulated for each study in Table 4.3.

Table 4.3 Characteristics of comparison data sets (idea adapted from Pridmore, R.W. 2005)

Study	Stimulus	Task	Luminance	Definition of discrimination thresholds	Number of observers	Number of colour centers (surround colour)	Mean threshold range
Baribeau and Robertson[3]	Colours simulated on CRT monitor (~ 10^o)	Colour discrimination (four-alternative forced choice procedure)	Fixed luminance at L_{ab}*= 51	Euclidean distance in the L*a*b* space given by ΔE_{ab}	3	18 (40^o background field at 50% neutral grey)	0.3 to 0.9 (ΔE_{ab} units in L*a*b* space)
Brown[1]	Binocular wide field colorimeter (subtending 10^o at viewing position)	Colour matching of the test field to the comparison field	On an average-17.052 cd/m^2	Distribution of matched results	12	22 (white surround at 8.9cd/m^2)	(+/-) 0.00070 to 0.00605 (of the reference color in CIE1931 xy space)

Study	Stimulus	Task	Luminance	Definition of discrimination thresholds	Number of observers	Number of colour centers (surround colour)	Mean threshold range
Wyszecki and Fielder[2]	Binocular 7- field colorimeter with two hexagon fields subtending 3^o each.	Colour matching	12 cd/m^2	Distribution of matched results	3	28 (40^o white surround at 6 cd/m^2)	(+/-) 0.0011 to 0.0093 (of the reference color in CIE1931 xy space)
Xu and Yaguchi[4]	Colours simulated on CRT monitor (2^o stimulus)	Colour discrimination task (stimulus presentation - interleaved staircase method, forced choice procedure)	Sampled across the L*a*b* space (L* varying from 35 to 87)	Euclidean distance in the L*a*b* space given by ΔE_{ab}	8	5 (6^o background with luminance same as the reference colour)	2 to 3.19 (ΔE_{ab} units in L*a*b* space)

[1]Brown, W.R.J. 1957; [2]Wyszecki, G. 1971; [3]Baribeau, R. 2005; [5]Xu, H. 2005.

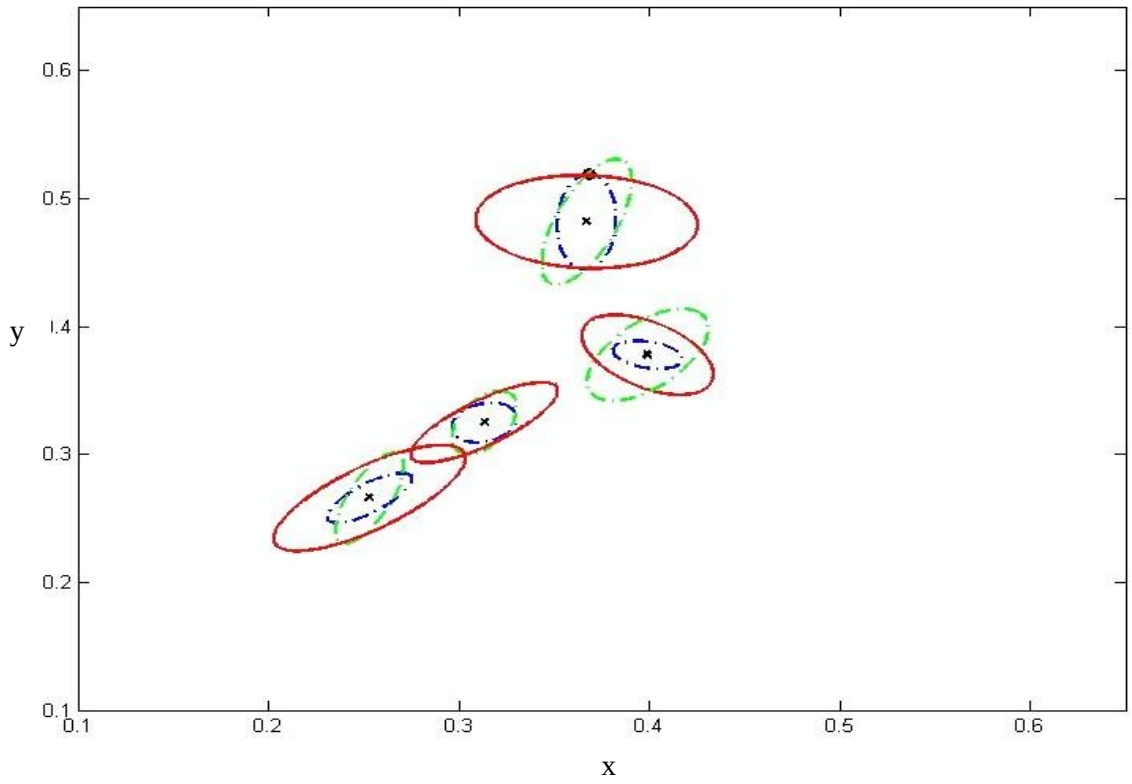


Figure 4.6 Discrimination ellipses in CIE 1931 xy space. Blue represents Brown's data set; Green represents Wyszecki and Fielder's data set; Red represents the current study's data set. All ellipses are magnified 10 times the original size. The colour centers from Brown's and Wyszecki's data are close to the reference colours but not exact as shown.

Baribeau and Robertson's stimulus conditions and procedures were very close to the one used in the current study, but their subjects were highly experienced. However, their study reported thresholds from only one direction in the colour space for 18 hues at a fixed chroma and luminance level. For comparison with their results, the position of the reference colours in the current study were translated in terms of hue angles and the average threshold for the equi-luminance plane was compared. The mean threshold at hue angles of 60^{o} (corresponded to the skin-tone reference colour), 120^{o} (corresponded to a green reference colour) and 260^{o} (corresponded to the blue reference colour)

were 0.5, 0.6 and 0.6 respectively. These values when compared to our study (skin-tone: 1.08, green: 1.18, blue: 1.05) were lower by a factor of 2.

Xu and Yaguchi also used a CRT monitor and a forced choice procedure to measure colour difference thresholds. Although the reference colours used in this study were not identical to Xu's study (differed by a ΔE of 30) the area of their ellipses in general was larger for Xu's data. Their ellipses were larger by a factor of 2.0 than the ellipses reported in this study. This difference could be due to the fact that the stimulus used in Xu's study was a 2^o field which is smaller than the stimulus used in the current study.

Given the range of threshold values from 0.5 to 3 ΔE_{Lab} units reported in the literature, it is not too surprising that our mean results fell within this range. Another useful comparison of our findings was to determine the frequency of individuals who obtained results near the two boundaries of this range. Xu and Yaguchi's results were used as an upper limit for threshold for naïve subjects and Brown's results were used as a lower threshold limit. For this comparison, we used only the major axis values of the discrimination ellipses in the equi-luminance plane. There were 3 observers out of 20 (15%) who had major axis greater than the mean threshold of ΔE_{Lab} as reported by Xu and Yaguchi, but only one naïve observer (5%) whose threshold was equal to the mean value reported by Brown. Thus, the mean thresholds found in this study are not as good as the thresholds which were based on the variability in colour matches but are within the ranges reported for thresholds based on a force-choice procedure. Our results, combined with the other experiments which used a forced choice procedure demonstrated that thresholds based on colour matching variability are probably too conservative for quality control purposes.

The results in this chapter and in the previous one show that as the mean threshold increases along the yellow-blue direction there is an increase in variability. The source of variability was an

increase in the inter-observer variability for the yellow-blue thresholds. To analyze this trend, the ratio of the yellow (vector along the +b direction) to red (vector along the +a direction) thresholds were calculated for the white and blue reference colours for each subject. Eighty five percent of the subjects had yellow threshold higher than the red threshold by a factor varying from greater than 1 to 3.13 and 15% of the subjects had yellow thresholds that ranged from a factor 0.83 to 1 relative to the red. Also, subjects who showed higher sensitivity in the yellow direction for the white reference colour did not show similar trends with the blue reference colour, therefore apart from the inter-observer variability, intra-observer variability also accounts for the larger variability noticed with higher means. However, the reason for such behaviour is not obvious.

One of the issues raised was in the selection of the CIE 1974 L*a*b* colour difference formula over the CIE2000 equation. Figure 4.7 shows the mean colour difference thresholds for the CIE 1974 L*a*b* equation and the CIE2000 transformations for average data for each reference colour. Note that thresholds in CIE2000 are more periodic as a function of the colour angle, which confirms that the CIE 2000 colour difference space is not spherical. In general, transformations to the CIE 2000 space increases the area of the ellipse as the reference colours are moved away from the center in a hue plane. However, in the present data it was observed that there is an increase in area of the ellipse for the skin-tone reference colour, but a decrease in the area for the green reference colour (which is even farther than the skin-tone reference colour from the origin). This discrepancy could be accounted by the fact that the sampled reference colours were at different levels of luminance and were not on an equi-luminance plane. It was noted that the orientation of our ellipses were relatively more tilted towards the center of the space, with the CIE 2000 transformations. This shows that transforming from a CIE L*a*b* space to the CIE2000 space only changes the orientation and area of the ellipse to present a more uniform elliptical space and thus does not offer any significant advantage for the sampled reference colours.

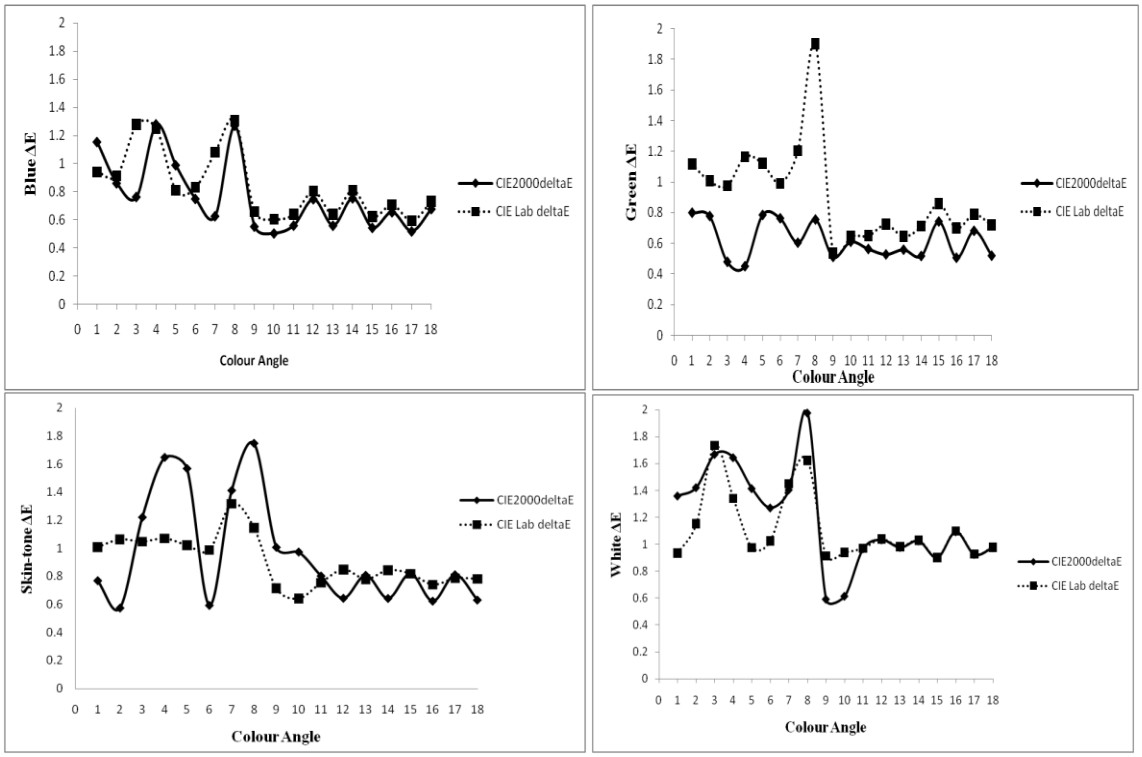


Figure 4.7 CIE2000 Delta E as a function of Hue angle. Delta E in L*a*b* space is also plotted for comparison.

4.5 Conclusion

The average of thresholds that vary in both hue and luminance (non equi-luminance vectors) are tabulated in Table 4.4 and is used to scale measured colour differences in terms of perceptible threshold steps. It is evident from Table 4.4 that the perceptibility thresholds are lower than the conventional industrial standards of using $1\Delta E_{Lab}$ as a precision limit for calibration. This means that the traditional limit is a lenient criterion for calibration.

Table 4.4 Values presented represent a perceptible threshold unit for each reference colour.

Colours	**Average threshold of vectors that vary in both hue and luminance (vectors on the increasing and decreasing luminance plane)**
White	0.9858
Skin-tone	0.7919
Green	0.7251
Blue	0.6943

Having set up a metric to quantify colour difference for CRT colours, the next step is to simulate the changes in hue and luminance as a function of viewing angle for complex scenes and subjectively determine the acceptability and image degradation as a function of ΔE.

Chapter 5

Rating of supra-threshold degradation simulated on static complex scenes

5.1 Background

Subjective ratings often play an important role in colour rendering and colour reproduction industries. One of the first experiments on the correlation between colour differences and acceptability was carried out by Davidson et al. in the year 1953. The purpose was to help establish a quality control metric for colour production of yarn in carpet industry. Their results showed that a colour difference within 2.5 times ellipsoids reported by MacAdam's were acceptable 50% of the time (~ $4\Delta E_{Lab}$). More recent acceptability type experiments involve esthetic restorative materials research, where the acceptability index helps in selection of tooth ceramic implants and restorative materials like tooth fillings etc (Douglas, R.D. 1998; Lindsey, D.T. 2007). There results showed that the 50% acceptability of crowns is between the ΔE ranges of 1.2 to 2.7, i.e. approximately 4 MacAdam's units. A more recent study using LCD generated colours was carried out by Urban et al (Urban, P. 2011). Their study focused on obtaining tolerance levels at 50% rejection probability between the anchor and test pairs, in comparison to the surface colour data. Results showed that tolerance 50 corresponded to about a ΔE_{Lab} of 2.2 on an average from the anchor to the test point. These results are well below the range specified by Davidson for carpet industry. Nevertheless acceptability data has always been related to establishing tolerance levels for observers/customers and is specific for a given industry/sample. This chapter focuses on establishing tolerance levels for the degradation produced at different viewing angles in comparison to an undistorted image in a Microtile™ display viewing scenario.

5.2 Objective

The objective of this study is to (i) simulate the shift in luminance and hue at different viewing angles measured on the Microtile™ display, on complex static scenes (ii) and to have observers rate the acceptability and degradation of these images.

5.3 Methods

5.3.1 Equipment

The simulations were carried out using the Visual Stimulus Generator (ViSaGe, Cambridge research system) controlled by Matlab on a Sony Trinitron CRT monitor used for the perceptibility experiment. Since the rating experiment was carried out soon after the perceptibility thresholds task, only a calibration check was performed to see the stability of the CRT monitor. The surround conditions were maintained the same as in the perceptibility experiments (see Chapter 3 for details) and was checked before the commencement of the experiment. The stimulus parameters were the same as defined in Chapter 3 except for the stimulus itself, which were three static complex images.

5.3.2 Simulation

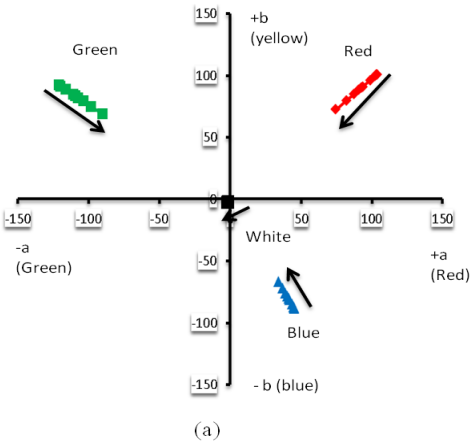
5.3.2.1 Selection of static images

Three complex static images, a white car (BMW- Lumma CLR 600 on a uniform grey background), a green landscape and a portrait (unfamiliar face) were picked from the web and the chromaticity co-ordinates and luminance of the images were altered using PhotoShop™ to match the reference colours used in the threshold study (the images are not printed due to restricted copy right privilege). These images were selected in consultation with the sponsors (Christie Digital, Ontario). Selections were made to show three general categories of images and to select images so that the colour co-ordinates of the total image (calculated by averaging across each pixel) were approximately the same

as the reference colours used in the perceptibility experiments. The size of each image was maintained at 1024 X 768. The simulations were programmed using the image processing tool box in Matlab.

5.3.2.2 Programming

The Microtile display characteristics as a function of viewing angle were discussed in Chapter 2. From the observed trends, it was noted that the luminance decreases and the hue shifts towards the blue-green region of the colour space as viewing angle changes as shown in Figure 5.1a (same as Figure 2.6). The goal of simulation was to match these observed shifts on the three complex images.



(b)

viewing angle	Car Delta E	Portrait delta E	Landscape Delta E
0	0	0	0
2	0.3604	0.3385	0.2959
5	1.1377	1.0474	0.8948
10	4.0876	3.7141	3.3903
15	8.2861	7.4782	6.9261
20	9.3029	8.3194	7.8127
30	13.6582	12.1025	11.5416
40	24.4244	21.657	20.6623

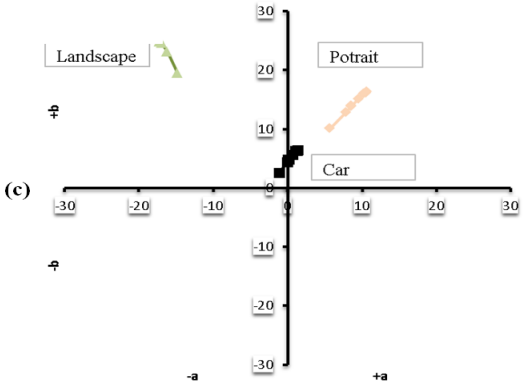


Figure 5.1(a) Hue shifts noticed in the microtile primaries (plotted with in the Microtile colour gamut), (b) Delta E of the complex image at different viewing angles, (c) Hue shifts simulated with complex images (plotted with in the CRT colour gamut).

The normalized luminance calculated for the three primaries of the Microtile™ display at different eccentricities were used in simulating the trend on the CRT monitor. The stimulus represented a 2x2 assembly of Microtiles™ displaying the image. Each quadrant is considered as a huge matrix with each pixel of the monitor representing an element of the matrix. The information to each pixel is in turn an additive combination of the red, green and the blue phosphors output. Therefore the image matrix is the sum of three equal size phosphor matrices R, G, and B. To simulate the measured attenuations of the Microtile primaries at different viewing angles, each phosphor matrix is multiplied by a scalar - the normalized luminance of the primaries (from the figure 2.3b) at

each viewing angle. A schematic of the simulation code is given in Figure 5.2. Figure 5.1c shows the trend in shift simulated for the three complex static images. Note that the shifts follow the same trend observed in the Microtile primaries as in Figure 5.1a, but in addition there is a systematic decrease in luminance as the hue becomes more desaturated. The Graphic User Interface was programmed to read in attenuations of the primaries for eight different viewing angles from any display.

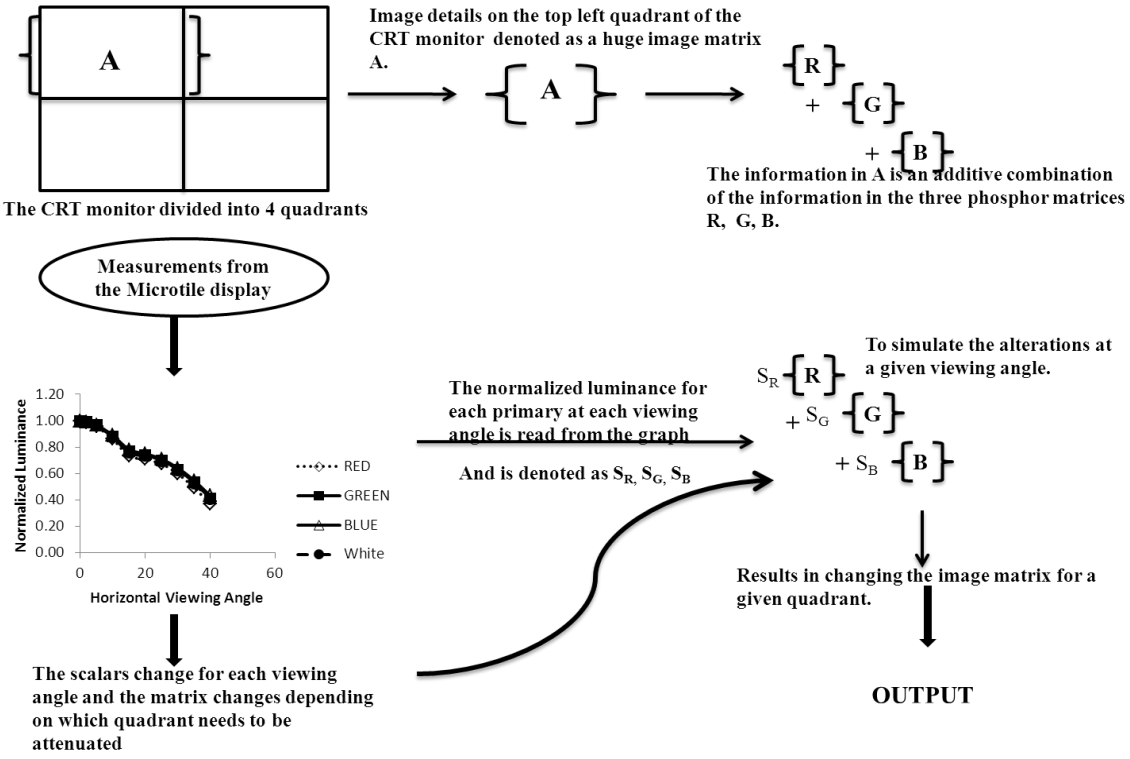


Figure 5.2 Schematic of the steps involved in simulating image distortions at different viewing angles.

5.3.3 Rating scales

The aim of the rating experiment is to provide the manufacturer of Microtile displays an observer/customer point of view on the alterations at different viewing angles. A pilot study was carried out with a set of three naïve observers and three trained observers (who either participated in a

rating task before or who designed rating experiments). The goal was to see which scale was best utilized and least confusing to use. Three scales were evaluated. One was a continuous scale ranging from a 0 to 100 in which subjects were asked to rate the percentage degradation of the total image. The second scale was a value scale from $0 to $1000 where subjects were instructed to imagine themselves as customers and to rate each presentation as too how much they would spend if they were to purchase it. The third scale was a categorical scale with eight categories corresponding to eight alterations ranging from the best image (acceptable) to the worst image (unacceptable). Subjects were instructed that each alteration corresponds to a different viewing angle, and were asked to rate each presentation based on the overall appearance of the image. Each time the anchors were provided (the best and worst case) and the instructions were scripted. After each rating observers were questioned on the ease with which they were able to use the scale and the strategy they followed to classify observations.

In general, observers preferred the continuous scale, because it was easy to understand and the rating data was fairly continuous corresponding to different alterations presented. The value scale showed clumping of data at either ends. And subjects always almost rated presentations as either worth $0 or $1000; thus this rating system was equivalent to an acceptable/unacceptable type judgment. The categorical scale based on different levels of acceptability/unacceptability, was easy to understand but while rating, subjects found it especially confusing to categorize alterations that they thought were in the transition from acceptable to unacceptable. Based on the observations from the pilot study, a continuous scale to rate image degradation and a 2 – category, acceptable/unacceptable scale was picked for the rating study.

5.3.4 Subjects

Twenty subjects participated in the rating study. Ten were naïve observers and ten were observers who participated in perceptibility thresholds experiment but not necessarily trained in rating tasks. An

initial colour vision testing was performed on all new candidates (as in Chapter 3). Subjects were divided into two groups, with each group having ten members of which five were from the group of subjects participated in the perceptibility experiments and five were from the set of new participants. Subjects coded with even numbers started with the image degradation rating task, while those assigned odd numbers started with the acceptability rating. Within each rating task, there were 3 iterations before subjects could swap scales. Each trial comprised of rating all three images using a given scale, and the order of presentation of the three images were pre-assigned in such a way that at every visit the rating begins with a different image. After completing 3 iterations/trials (1 session) for a given rating scale, subjects swapped to the other rating scale. The three trials within a session were scheduled in such a way that there was at least one day in between every visit, to reduce the role of memory in rating.

For each image there were eight alteration levels and 4 repetitions of each alteration (two on the left and two on the right halves of the CRT monitor). Therefore, for a given image there were 32 presentations and a total of 96 presentations for a given trial for all three images. There was no time constraint on making each judgement; however, subjects on an average did not exceed 1 minute to respond to a given presentation. The stimulus was presented until the response was registered and between each presentation a white adaptation screen was presented for 5 seconds to reduce any particular adaptation effect. Subjects were positioned using a head and chin rest and eye movements were allowed during the experiment.

5.3.5 Procedure

Image degradation was rated on a continuous scale with the anchors representing best image (0% degraded) and worst image (100% degraded). The categorical scale was based on two categories, acceptable/unacceptable. Subjects were asked to pen down their response and to verbally spell out the response each time, to the experimenter. The instructions for the image degradation rating and

acceptability rating are shown in Figures 5.3 and 5.4. Instructions were scripted out to avoid excess/less information.

Image degradation:
Instructions:
→ For the next 10 minutes what you will be seeing is a set of alterations to this image (unaltered image is shown).
→ These alterations characterize the appearance of the image as the viewer's angle of view change from a straight ahead view to an extreme angle.
→ Your task is to rate each image as to how "degraded" the total image appears to you; given that you are viewing these scenes from different viewing angles. The maximum degradation on this scene is ___ (show them the worst image).
→ You will score on a continuous scale from 0 to 100 (0 – being no image degradation and 100 being maximum image degradation).

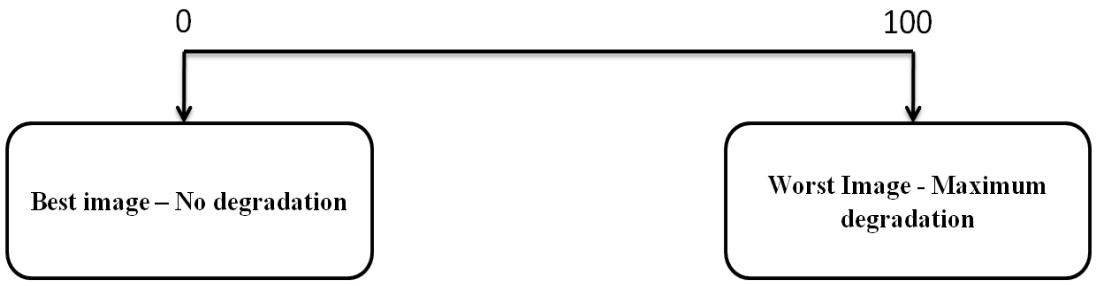


Figure 5.3 Instructions for Image degradation rating.

Acceptability Rating- Instructions:

→ For the next 10 minutes what you will be seeing is a set of alterations to a given image (unaltered image is presented).

→These alterations characterize the appearance of the image as the viewer's angle of view changes from a straight ahead view to an extreme angle.

→ The best and the worst images are presented (anchors)

→ Your task is to categorize each of the scenes as being "acceptable or not acceptable", based on the overall appearance of the image.

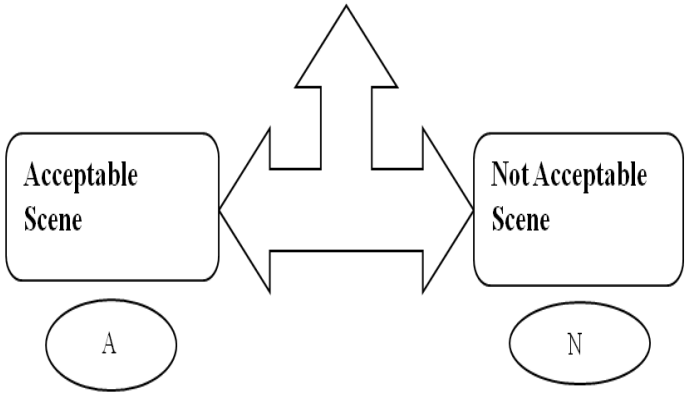


Figure 5.4 Instructions for acceptability rating.

5.4 Results and Discussion

5.4.1 Acceptability rating results

Figures 5.5 to 5.7 show percentage acceptability results for the three images as function of both the simulated average ΔE and the corresponding viewing angle for the image. A RMANOVA was carried out to determine if there was a practice effect. There was no statistically significant trial effect ($p > 0.05$), and so all three results from rating iterations were included in the analysis. The percentage acceptability was calculated based on the ratio of the number of times a simulation level was rated acceptable (from all three trials and across all subjects) to the total number of presentations for that given simulation level. As tabulated in Figure 5.1b, the ΔE at each viewing angle is different for

different images and the simulated ΔE's are labeled below each viewing angle for the ease of comparison.

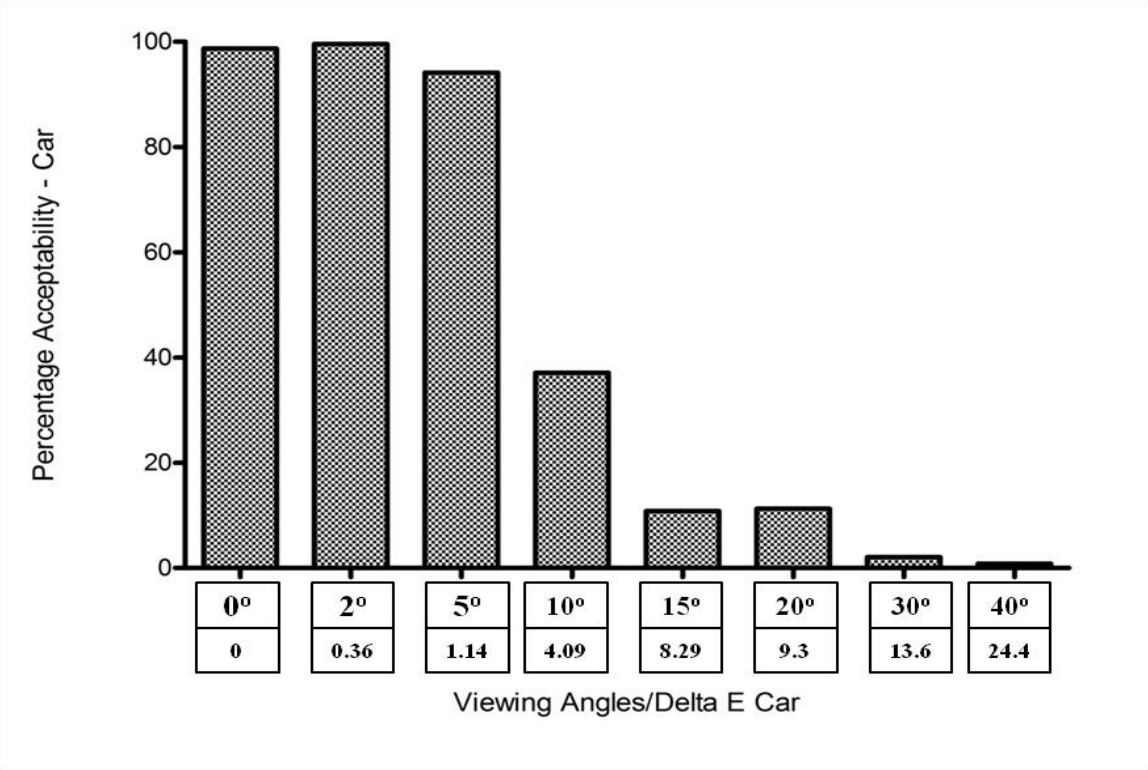


Figure 5.5 Acceptability rating for white car at different levels of distortion.

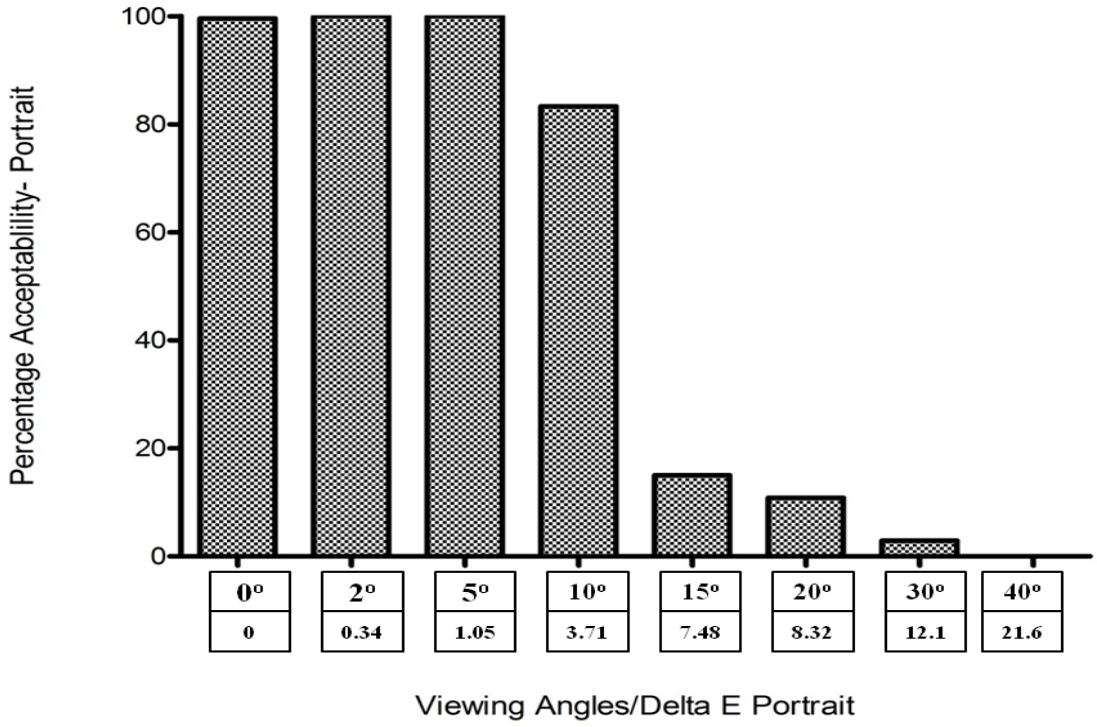


Figure 5.6 Acceptability Rating results for Portrait at different levels of distortion.

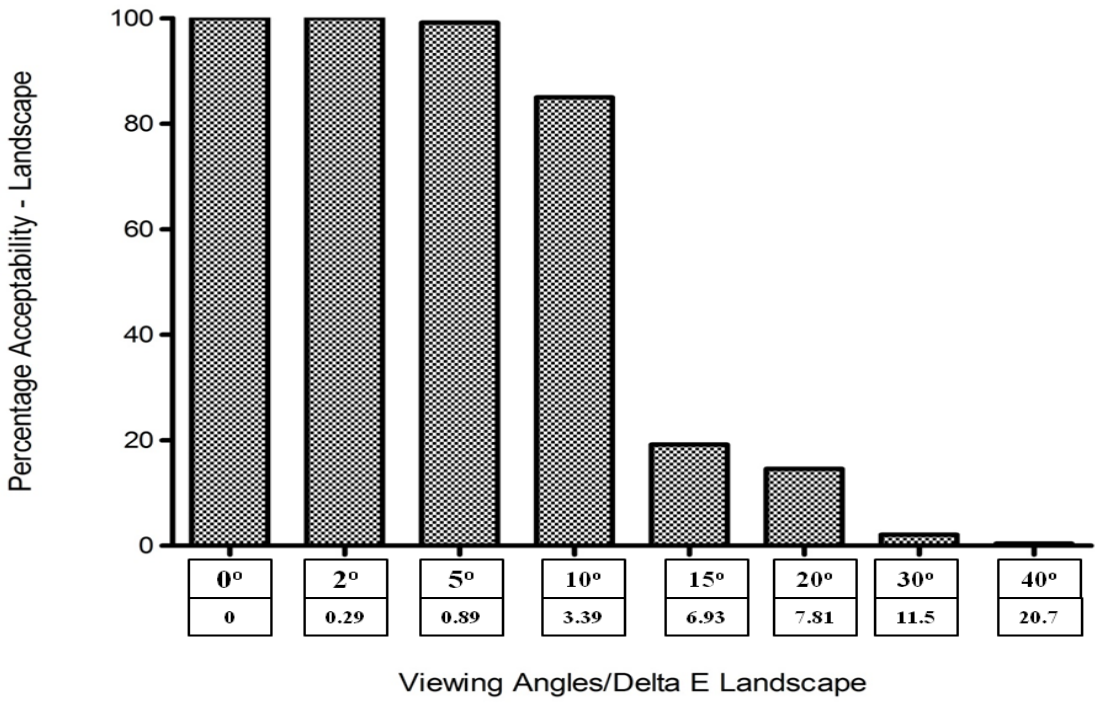


Figure 5.7 Acceptability Rating results for Landscape at different levels of distortion.

Note that the percentage of acceptability rating quickly falls below 20% beyond a viewing angle of 10^o for all three scenes. The difference between the scenes is that the acceptability ratings begin to decrease at 5^o for the car and 10^o for the other two scenes. This difference is not simply due to the fact that the average ΔE for the car is greater than the ΔE for other images for a given angle. The unacceptability rating begins for the car at a viewing angle of 5^o corresponding to a ΔE of just over 1, whereas the acceptability ratings for the other two scenes does not decrease until the ΔE is greater than 3. If the ΔEs are expressed in steps of perceptibility thresholds (of the uniform reference colours discussed in Chapters 3 & 4) the ΔE at a viewing angle of 10^o for the car is about 4.3 times the threshold of the white uniform reference colour, whereas the ΔEs at a viewing angle of 10^o for the other two images were about 5 times the threshold for their respective uniform reference colours. One

might predict that the acceptability ratings at this viewing angle would be similar for the three images, or the car would have a slightly higher acceptability based on multiples of threshold scale. However, it was noted that the acceptability ratings for the car was less by a factor of 2 compared to the acceptability ratings for the other two images at a viewing angle of 10^o. This finding suggests that the acceptability judgments were influenced by the differences in contours and textures between the images. Furthermore, the actual objects in the image may also interact with the judgments with the possibility that individuals are more discerning if the image is a manufactured object such as a car. Nevertheless, it can be inferred that at a viewing angle of 5^o (around a ΔE of 1.5 times thresholds) the acceptability begins to decrease and beyond a viewing angle of 10^o (around a ΔE of 5 times thresholds) most images would be unacceptable.

Comparing the present results to some of the previous acceptability results discussed in the introduction of this Chapter, the 50% acceptability for the car occurs between ΔEs of 1.5 to 4, whereas, the 50% acceptability occurs beyond a ΔE value of 3.3 for the other two images. The results for the car are consistent with the previous results, but the ΔE for 50% acceptability for the other 2 images was higher. This could be the difference between uniform colours or different complex images (like teeth or carpet) used in previous studies versus complex natural scenes used in this study. Since the stimuli used were natural scenes, the acceptability judgements depend on several factors (like contours, details, luminance differences across the scene, etc) as opposed to judging uniform colours.

5.4.2 Results from Image degradation rating

Figure 5.8 to 5.10 summarizes the results from the image degradation rating as a function of ΔE. A RMANOVA revealed a significant trial effect $\{F(2, 38) = 5.046, p \leq 0.05\}$, and a pair wise comparison (using Bonferroni correction at $p \leq 0.017$) revealed that between trials, the first trial was significantly different ($p \leq 0.017$) from the second and third trials showing a learning effect, and there

was no significant difference between the second and third trials ($p > 0.017$). Therefore, ratings from first trial were dropped. Recall that the scale was designed as a continuous scale from 0 (no degradation) to 100 (maximum degradation). The lower variability at the extremes (best image and worst image) is probably due to the effect of fixing the anchors. The linear regression results are shown in Figures 5.8 to 5.10. Table 5.1 shows the regression results. There was high correlation between the ΔE and image degradation for all three images with all correlation coefficients above 0.95.

Table 5.1 Regression Results for the Image degradation rating data

Best Fit Line for the rating data for each image	r value (95% CI range)
Car: y=3.748x + 6.925	0.958 (0.943 to 0.969)
Portrait: y=4.615x – 1.167	0.966 (0.954 to 0.975)
Landscape: y=4.857x – 0.3604	0.962 (0.948 to 0.972)

The effect of the image is reflected in the slope and intercept of each linear equation. The linear function for the degradation rating of the car has a higher y-intercept and a lower slope compared with to the portrait and the landscape functions. This difference in the rating for the car shows that the lower ΔEs were rated to be more degraded when compared with the other two images. This trend in the judgments for the car was also seen in acceptability ratings. Distortions corresponding to 5^o viewing angle (a ΔE of over 1.5 times thresholds) were rated to be less acceptable, with a substantial decrease in the acceptability at 10^o viewing angle (a ΔE of 5 times thresholds) compared to the judgments made with the other two images.

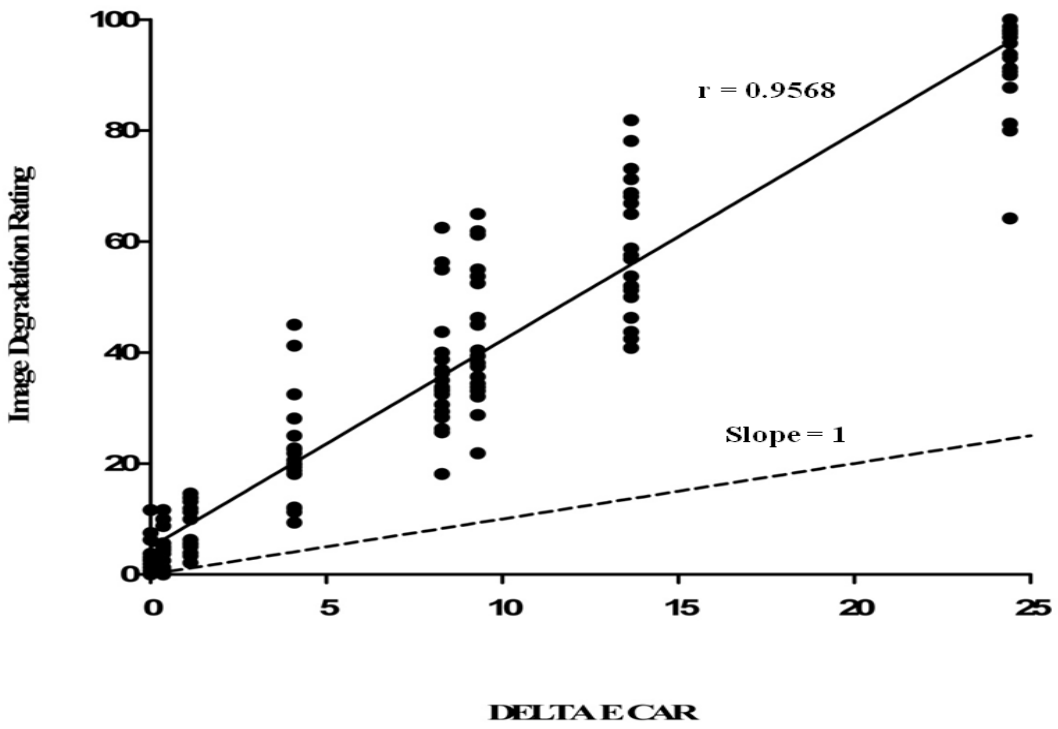


Figure 5.8 Image Degradation rating as a function of the simulated ΔE for the car at different viewing angles. The data is fitted with a linear function given by, y=3.748x + 6.925. The dashed line represents the line with slope 1 for comparison.

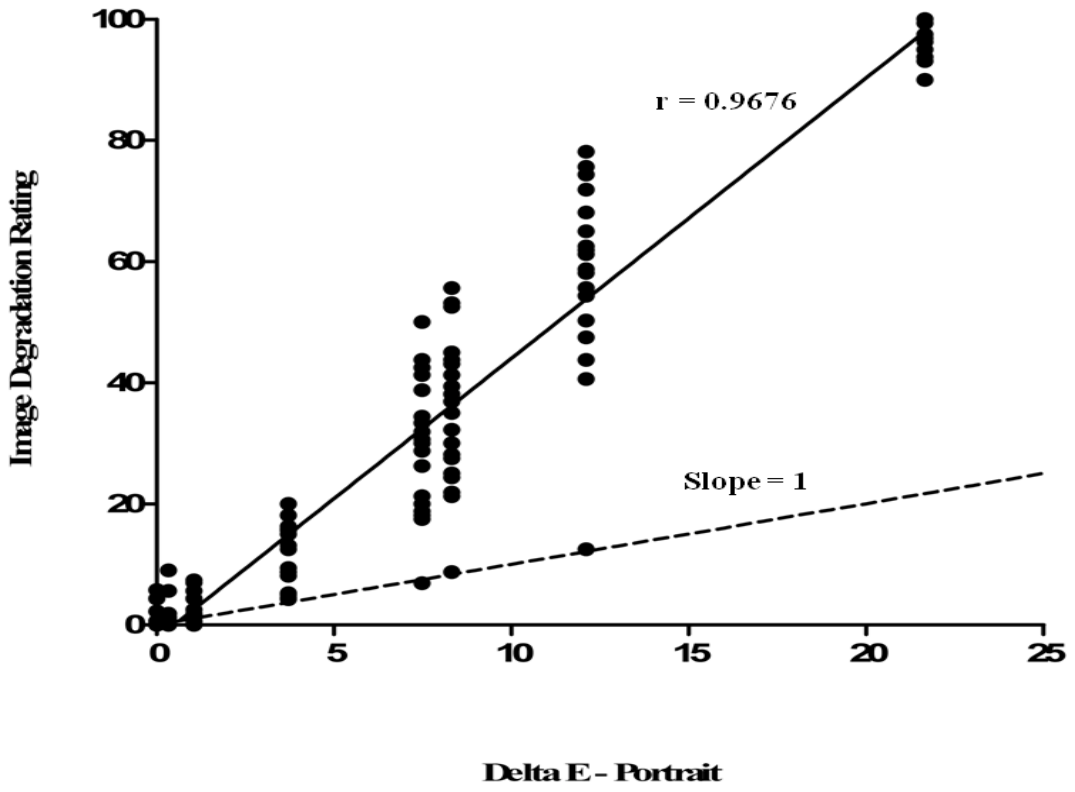


Figure 5.9 Image Degradation rating as a function of the simulated ΔE for the portrait at different viewing angles. The data is fitted with a linear function given by, y=4.615x – 1.167. The dashed line represents the line with slope 1 for comparison.

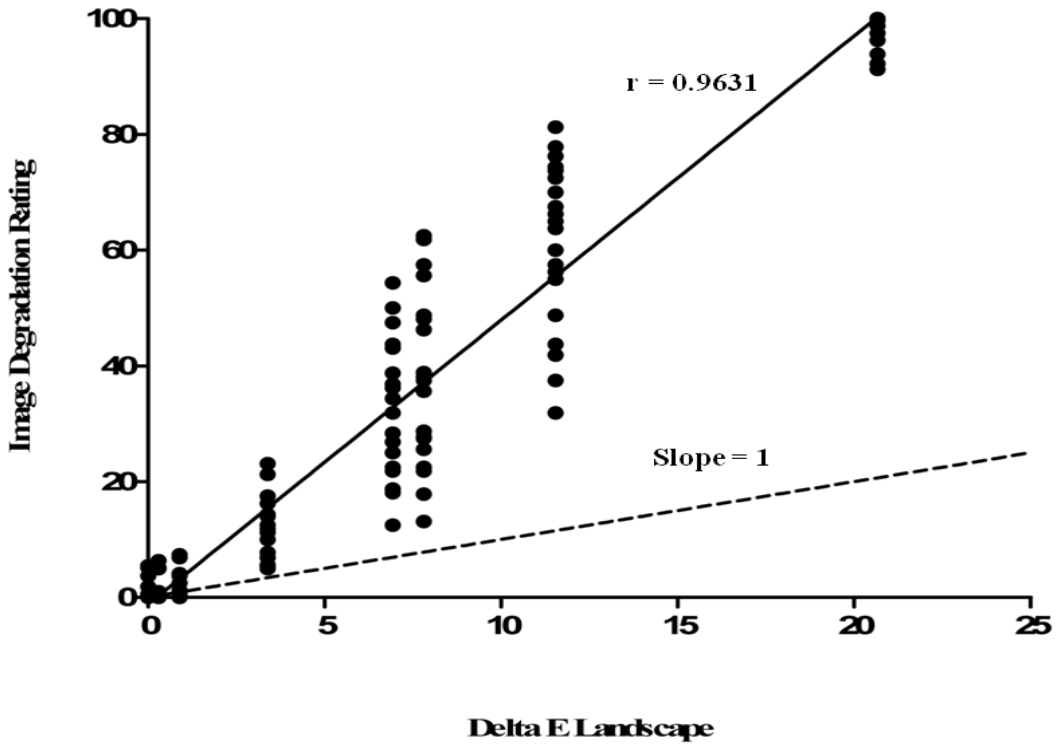


Figure 5.10 Image Degradation rating as a function of the simulated ΔE for the Landscape at different viewing angles. The data is fitted with a linear function given by, y=4.857x – 0.3604. The dashed line represents the line with slope 1 for comparison.

Of interest was whether a power function or logarithmic function provides a better fit to the data. Although both nonlinear functions provided excellent fits (power function with r values of 0.96 and logarithmic function with r value of 0.90) to the data; however, neither function provided a significantly better correlation than the linear regression based on the 95% confidence intervals of the linear correlation coefficient. Thus the results are inconclusive as to whether the image degradation ratings using this paradigm were better explained by Steven's Power Law or Fechner's Law. The inability to distinguish between the two may be due to a limited range of ΔEs sampled along with the fixed anchors or a combination of the two factors. Nevertheless, the results do show that a linear regression is sufficient in explaining the trend.

5.5 Conclusion

The subjective rating results were based on a simulated distortion of one half of the image relative to the other half which was unaltered. Irrespective of the complexities in the image, an image with a degradation rating greater than ~30% is unacceptable on more than 50% of the presentations. In terms of the ΔE values, any average distortion in colour above five supra threshold steps will likely be unacceptable to all discerning viewers. Nevertheless, the image context does have some influence on the ratings, but this effect is not obvious from these experiments. It may depend on whether the image is a natural object or a manufactured object. In terms of the Microtile displays, the tolerance for different viewing angles would be less than 15^{o} eccentricity for natural images and 10^{o} for a manufactured object such as a white car. These tolerance limits are restricted to static images and so it would be interesting if the findings generalize to dynamic scenes which have an added dimension of movement and changes in content over time.

The interactive rating algorithm programmed in this experiment provides flexibility to simulate the observed trends with different displays and is not restricted to Microtile displays.

Chapter 6

General Conclusions

The aim of this study was to evaluate the subjective tolerance to image distortions at different viewing eccentricities for the Microtile Display. The preliminary experiments were based on studying the characteristics of the Microtile displays with different viewing angles. The later experiments aimed at quantifying the changes in hue and luminance as a function of viewing angle in terms of perceptibility thresholds and to rate the acceptability of these shifts when simulated on complex scenes.

6.1 Results from preliminary studies on the Microtile™ display characteristics

The Microtile™ display behaviour as a function of viewing eccentricity was measured using a photometer (SpecBos). It was noted that the luminance decreased with viewing eccentricity for all 3 primaries with the red primary drifting at a faster rate compared to the green and the blue primaries. This was evident with a colour drift of the primaries towards the blue-green region and a decrease in luminance in the L*a*b* colour space. Given the technical limitation of the Microtile displays, the concern was to obtain observer tolerance rating for these shifts at different viewing eccentricities.

The enormous amount of colour difference literature has a general drawback. All reported colour difference thresholds were limited to implicit or explicit viewing conditions and this posed a need to design a perceptibility threshold task pertaining to the Microtile display viewing conditions. Subsequent experiments were planned based on two goals, (i) to obtain new data on perceptibility thresholds in order to define measured delta Es in terms of threshold units specific to the Microtile display; and (ii) to have subjects rate the acceptability and degradation of the measured changes simulated on complex static scenes.

6.2 Results from perceptibility experiments

The perceptibility experiment was conducted in two phases, (i) with vectors sampled across the blue-yellow region of the colour space (region specific to the observed Microtile shifts) for three reference colours, and (ii) with vectors sampled around each of the four reference colours in representative directions throughout the colour space, so that results can be used for different displays. The thresholds from the non equi-luminance directions were averaged for each of the reference colours to provide the discrimination threshold. Thus the measured ΔE values were now interpreted in steps of threshold units. The results from perceptibility experiments are summarized in Table 6.1.

Table 6.1 summarizes the discrimination thresholds for the reference colours from the two phases of the experiment.

Colours	**Average threshold (1st phase of the perceptibility experiment for 3 reference colours) – Specific to the Microtile displays.**	**Average threshold (2nd phase with larger sampling direction for 4 reference colours)**
White	0.9286	0.9858
Skin-tone	0.7322	0.7919
Green	0.6869	0.7251
Blue	--	0.6943

6.3 Results from rating experiment

The rating experiment was designed to provide a measure of observer/customer tolerance to the shifts noticed with the Microtile displays. An acceptable/unacceptable categorical scale was constructed to provide a direct opinion on the appearance of the images. In order to get a sense of how degraded the simulated alterations to the images appeared, an image degradation scale (continuous scale from 0 – no degradation to 100- maximum degradation) was also used.

Results from these experiments show that alteration greater than 5 times thresholds (five supra-threshold steps) were always rated to be acceptable less than 50% of the time and the degradation was rated to be less than 30% degraded. This translates into viewing angles that are greater than 10^{o} (or below 15^{o}), that would be judged as intolerable by most individuals who were presumably discerning observers. However, the ratings are relative to an undistorted, head-on view appearance and therefore are an exaggeration to the actual scenario, but helps in providing judgements for a range of ΔEs. Results from all the experiments are summarized in Table 6.2. The threshold units used to scale ΔEs are specific for the Microtile displays.

Table 6.2 Summary of all results

Visual angle		0	2	5	10	15	20	30	40
	Delta E	0	0.36	1.138	4.088	8.286	9.303	13.658	24.424
Car	Threshold units	0	0.385	1.216	4.368	8.856	9.942	14.597	24.979
	Degradation Rating	3.29	4.67	9.37	23.75	38.35	44.77	61.71	94.87
	Percentage Acceptability	98.75	99.583	94.167	36.667	10.833	11.25	2.083	0.833
	Delta E	0	0.338	1.047	3.714	7.478	8.319	12.102	21.657
Portrait	Threshold units	0	0.462	1.43	5.073	10.213	11.362	16.529	29.578
	Degradation Rating	1.17	1.55	2.28	12.9	30.85	37.32	60.44	97.50
	Acceptability Rating	99.58	100	100	83.33	15	10.83	2.91	0
	Delta E	0	0.296	0.895	3.39	6.926	7.813	11.542	20.662
Landscape	Threshold units	0	0.431	1.303	4.936	10.083	11.374	16.803	30.081
	Degradation Rating	1.7	1.80	2.46	12.85	32.29	41.38	62.049	98.19
	Acceptability Rating	100	100	99.17	85	19.17	14.58	2.08	0.42

Thus results, for a Microtile display show that colour differences beyond 5 times thresholds are unacceptable more than 50% of the time and correspond to a viewing angle greater than 10^{o} (or less than 15^{o}). Besides, the perceptibility thresholds also show that the general calibration precision of 1 ΔE_{Lab} is a lenient cut off and should be decreased by at least 25% for the Microtile displays.

Appendix A

Calibration assumptions

There are certain standard assumptions with the calibration routine. It is essential to check if these assumptions hold correct after the monitor has been calibrated. The output of the CRT monitor after calibration is tested for phosphor independence, spatial independence, monitor temporal stability, black point, and variability of the output.

Monitor stability

The stability of the monitor is the time period required for the monitor to warm up and present a stable output. The Sony Trinitron monitor used in this study was tested using the ColourCal and the LightScan™ software which helps record the output in a given sampling frequency. The light scan was programmed to record a reading at every 5 minutes interval for about an hour. Figure 1.a shows a graph of the luminance of the monitor with time.

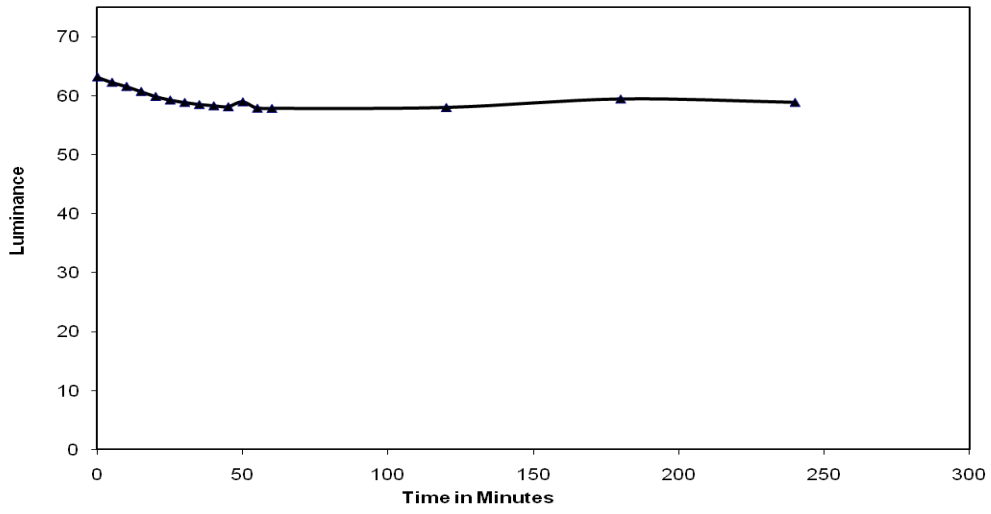


Figure 1.A CRT Monitor stability with time.

As seen in the graph, the luminance begins to stabilize at after about 40 minutes from the time it is switched on. The readings were continued after 60 minutes to about seven hours and the stability sustained. Therefore, the CRT monitor was warmed up for about 1 hour before the onset of the experiment. The output after an hour of warm up within each quadrant of the CRT screen was monitored to see the fluctuation after every one, two, and three hours of warming up. The luminance values were on an average stable over time.

Phosphor independence

In order to check for channel independence with in each of the quadrants with the CRT monitor, the monitor white point is calculated after calibration, and the luminance of the media white point is the maximum output of a given CRT monitor. Luminance at 12.5%, 25%, 50%, 75% and 100% of the media white point displayed on different quadrants of the video monitor and the output was measured using a ColourCal (2^0 observer). The measured co-ordinates are converted to the tristimulus values and used in calculating the Luminance of the individual phosphors for a given quadrant at a given input luminance level. The calculated output has to be close to 25% of the maximum luminance of each of the phosphor.

Given below is an example of the calculation:

Media white point L: 101.02

25% of 101.02 = 25.3

Maximum luminance of the phosphors

R=25.2 (25% = 6.3)

G=61.91 (25% = 15.48)

B=13.91 (25% = 3.47)

Measured output from Colour Cal= X: 25.82, Y: 25.82, Z: 35.59

$$\begin{matrix} L_R \\ L_G \\ L_B \end{matrix} = \begin{matrix} 1.85 & 0.469 & 2.05 \\ 1 & 1 & 1 \\ 0.082 & 0.167 & 10.46 \end{matrix}^{-1} \begin{matrix} 25.82 \\ 25.82 \\ 35.59 \end{matrix}$$

$$\begin{matrix} L_R \\ L_G \\ L_B \end{matrix} = \begin{matrix} 6.387 \\ 16.339 \\ 3.091 \end{matrix}$$

Phosphor luminance calculated are ~25% of the maximum luminance of each phosphor. Sum of phosphor luminance = 25.8 which equals a 25% luminance of the media white point for the given quadrant. This was repeated for each quadrant and all three times the monitor was calibrated. The behaviour was same in all iterations.

Monitor black point

The monitor was warmed up for about an hour and the input was set to zero luminance. The testing ambience was maintained and the SpecBos was used to measure the reflectance from the monitor. The luminance varied between 0.124 to 1 cd/m^2 which was almost negligible. Visually the screen appeared dark with the ambient lighting and therefore a reflectance term in neglected while performing transformations from XYZ to phosphor luminance.

Output variability

A reference colour was set in all four quadrants of the calibrated CRT monitor and the colour co-ordinates were measured using the SpecBos (10^0 observer) in the testing ambience. Measurements showed that the variability in the luminance was 2 cd/m^2 between each quadrant and the chromaticity co-ordinates was < 2%.

Spatial homogeneity

After calibration and setting up the testing ambience. The reference colours were presented on the test screen and two highly trained observers and two naïve observers were asked to look at the screen and report for any visual non-uniformity. The screen was visually homogeneous.

Appendix B

Spherical polar co-ordinate system transformation

The discrimination thresholds around each reference colour is assumed to be equal in all directions in the L*a*b* space. Hence, the length and orientation of the vectors in different directions are represented in spherical polar co-ordinate system. Figure 2.A shows a general form of a spherical polar system translated to L*a*b* space. r is vector length form the origin to any point p whose co-ordinates in space are(x, y, z); s- projection of r on the xy plane; Φ – angle between the vector and zenith axis (inclination angle/Luminance angle); $\boldsymbol{\theta}$- angle between projection vector s and x axis (azimuth angle/Hue angle); 1 and 2 represent the triangle in reference.

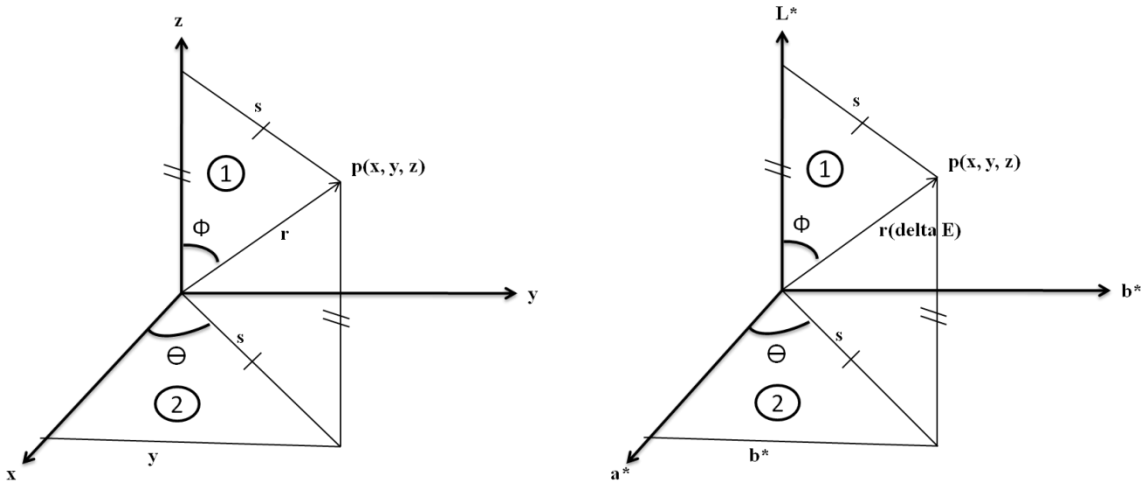


Figure 2.A Left: General form of a spherical co-ordinate system. Right: in terms of L*a*b*space.

This system is adapted in the L*a*b* space, with r representing the delta E (discrimination threshold) for a reference colour (origin) in a given direction. The x, y, z co-ordinates represent the

a*, b*, L* axis respectively; the inclination angle is the luminance angle and the azimuth angle is the hue angle.

Derivation of the Cartesian form;

In Δ2,

$$\cos\theta = \frac{a^*}{s}$$

$a^* = s * \cos\theta$ ---------- (i)

From Δ1,

$$\sin\Phi = \frac{s}{r}$$

$s = r * \sin\Phi$ ---------- (ii)

Substituting (ii) in (i),

$$a^* = r * \sin\Phi * \cos\theta$$

From Δ2,

$$\sin\theta = \frac{b^*}{s}$$

$$b^* = s * \sin\theta$$

Substituting (ii) in the above equation,

$$b^* = r * \sin\Phi * \sin\theta$$

Consider Δ1,

$$\cos\Phi = \frac{L^*}{r}$$

$$L^* = r * \cos\Phi$$

Therefore the Cartesian form of the co-ordinates is given,

$$L^* = r * \cos\Phi$$

$$a^* = r * \sin\Phi * \cos\theta$$

$$b^* = r * sin\,\Phi * sin\,\theta$$

Numerical example:

For instance vector length (delta E) r=4, $\Phi=90^0$ (vector on the equi-luminance plane), $\Theta=180^0$ (along the –a axis green direction).

$$L^* = 4 * cos\,90^0$$

$$L^* = 0$$

$$a^* = 4 * sin 90^0 * cos\,180^0$$

$$a^* = 4 * 1 * -1$$

$$a^* = -4$$

$$b^* = 4 * sin\,90^0 * sin\,180^0$$

$$b^* = 4 * 1 * 0$$

$$b^* = 0$$

References

Baribeau, R., & Robertson, A. R. (2005). Estimation of hue discrimination thresholds using a computer interactive method. *Color Research & Application, 30*(6), 410-415.

Bass, M., & Van Stryland, E. W. (2010). *Handbook of optics, vision and vision optics (3rd edition)* McGraw-Hill Professional.

Berns, R. S. (2000). *Billmeyer and saltzman's principles of color technology* Wiley New York.

Berns, R. S., Gorzynski, M. E., & Motta, R. J. (1993). CRT colorimetry. part II: Metrology. *Color Research & Application, 18*(5), 315-325.

Berns, R. S., Motta, R. J., & Gorzynski, M. E. (1993). CRT colorimetry. part I: Theory and practice. *Color Research & Application, 18*(5), 299-314.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization.

Brown, W. R. J. (1957). Color discrimination of twelve observers. *JOSA, 47*(2), 137-143.

Chickering, K. (1969). Perceptual significance of the differences between CIE tristimulus values. *JOSA, 59*(8), 986-990.

Chickering, K. (1971). FMC color-difference formulas: Clarification concerning usage. *JOSA, 61*(1), 118-122.

Cui, G., Luo, M. R., Rigg, B., & Li, W. (2001). Colour-difference evaluation using CRT colours. part I: Data gathering and testing colour difference formulae. *Color Research & Application, 26*(5), 394-402.

Davidson, H. R., & Friede, E. (1953). The size of acceptable color differences. *JOSA, 43*(7), 581-589.

Douglas, R. D., & Brewer, J. D. (1998). Acceptability of shade differences in metal ceramic crowns. *The Journal of Prosthetic Dentistry, 79*(3), 254-260.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals* Lawrence Erlbaum.

GLASSES, L., McKinney, A., Reilly, C., & Schnelle, P. (1958). Cube-root color coordinate system. *JOSA, 48*(10), 736-740.

Guan, S. S., & Luo, M. R. (1999). A colour-difference formula for assessing large colour differences. *Color Research & Application, 24*(5), 344-355.

Harvey, L. O. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, 18*(6), 623-632.

Kaiser, P. K., Boynton, R. M., & Swanson, W. H. (1996). *Human color vision* OSA.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*(7), 885-912.

Kinnear, P., & Sahraie, A. (2002). New farnsworth-munsell 100 hue test norms of normal observers for each year of age 5–22 and for age decades 30–70. *British Journal of Ophthalmology, 86*(12), 1408.

Kujala, J. V., & Lukka, T. J. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology, 50*(4), 369-389.

Lindsey, D. T., & Wee, A. G. (2007). Perceptibility and acceptability of CIELAB color differences in computer-simulated teeth. *Journal of Dentistry, 35*(7), 593-599.

Luo, M. R., Cui, G., & Li, C. (2006). Uniform colour spaces based on CIECAM02 colour appearance model. *Color Research & Application, 31*(4), 320-330.

Luo, M. R., Cui, G., & Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application, 26*(5), 340-350.

Luo, M., & Rigg, B. (1986). Chromaticity-discrimination ellipses for surface colours. *Color Research & Application, 11*(1), 25-42.

Melgosa, M. (2000). Testing CIELAB-based color-difference formulas. *Color Research & Application, 25*(1), 49-55.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research, 21*(2), 215-223.

Poirson, A. B., & Wandell, B. A. (1990). Task-dependent color discrimination. *JOSA A, 7*(4), 776-782.

Pridmore, R. W., & Melgosa, M. (2005). Effect of luminance of samples on color discrimination ellipses: Analysis and prediction of data. *Color Research & Application, 30*(3), 186-197.

Robertson, A. (1977). The CIE 1976 color-difference formulae. *Color Research and Application, 2*(1), 7-11.

Shevell, S. K. (2003). *The science of color* Elsevier Science.

Strasburger, H. (2001). Invariance of the psychometric function for character recognition across the visual field. *Perception & Psychophysics, 63*(8), 1356.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*(17), 2503-2522.

Urban, P., Fedutina, M., & Lissner, I. (2011). Analyzing small suprathreshold differences of LCD-generated colors. *JOSA A, 28*(7), 1500-1512.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A bayesian adaptive psychometric method. *Attention, Perception, & Psychophysics*, *33*(2), 113-120.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293.

Wright, W. (1941). The sensitivity of the eye to small colour differences. *Proceedings of the Physical Society, 53*, 93.

Wyszecki, G., & Stiles, W. S. (1982). **Color science: Concepts and methods, quantitative data and formulae.** *2nd Ed. New York, NY: John Wiley & Sons.*,

Wyszecki, G., & Fielder, G. (1971). New color-matching ellipses. *JOSA, 61*(9), 1135-1152.

Wyszecki, G., & Wright, H. (1965). Field trial of the 1964 CIE color-difference formula. *JOSA, 55*(9), 1166-1174.

Xu, H., & Yaguchi, H. (2005). Visual evaluation at scale of threshold to suprathreshold color difference. *Color Research & Application, 30*(3), 198-208.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293.

Wright, W. (1941). The sensitivity of the eye to small colour differences. *Proceedings of the Physical Society, 53*, 93.

Wyszecki, G., & Stiles, W. S. (1982). **Color science: Concepts and methods, quantitative data and formulae.** *2nd Ed. New York, NY: John Wiley & Sons.,*

Wyszecki, G., & Fielder, G. (1971). New color-matching ellipses. *JOSA, 61*(9), 1135-1152.

Wyszecki, G., & Wright, H. (1965). Field trial of the 1964 CIE color-difference formula. *JOSA, 55*(9), 1166-1174.

Xu, H., & Yaguchi, H. (2005). Visual evaluation at scale of threshold to suprathreshold color difference. *Color Research & Application, 30*(3), 198-208.