What is the Chrominance of “Gray”?

Raja Bala
Xerox Digital Imaging Technology Center, Webster, New York.

Abstract

This paper describes a study performed to understand preferences and tolerances for “gray” in reflection prints. A psychophysical experiment was conducted, with observers looking at printed targets containing near-neutral patches in different spatial layouts and under different illuminations. The task was to choose the patch or patches that were perceived as being closest to gray. Spectral measurements were made of these patches, and converted to CIELAB coordinates for each of the chosen illuminations. The observer responses were tallied into preference histograms as a function of the (a*, b*) coordinates of the patches. Results show that regardless of layout or viewing illumination, the peaks of the histograms occurred in the quadrant where both a* and b* are negative. Follow-on experiments were conducted using more complex targets under less controlled viewing conditions. The same trends were observed, indicating that people generally prefer “cooler” grays, and have low tolerance to errors that result in positive values of a* or b*. These observations can be used to derive robust gray-balancing for color output devices.

Introduction

The standard colorimetric definition for a gray stimulus is that a tristimulus (XYZ) measurement of the stimulus exhibits the same relative proportions as the tristimulus values of the reference adapting white. This is equivalent to stating that in CIELAB coordinates, a* = b* = 0. The typical method of gray-balancing an output device constitutes a calibration procedure which determines the device dependent signals required to produce stimuli that yield measurements of a* = b* = 0 under a particular illuminant [1]. For reflection prints, the de-facto standard for the calibration illuminant is D50 (this is especially true in the graphic arts industry).

Due to inherent errors in the calibration process, noise in the system, and other factors that are often beyond the control of the system designer, it is not always possible to achieve this ideal colorimetric definition of gray. In addition, the images are likely to be viewed under an illumination that is different from the one used for calibration. Furthermore, even if the viewing and calibration illuminations are the same, it is not obvious that colorimetric gray is necessarily coincident with what humans perceive as “gray”.

Hunt et al. [2] have explored gray preference in images viewed by projection transparency and television display. They concluded that subjective preference for gray relates strongly to incomplete adaptation, and can vary significantly depending on the chromaticity and luminance level of the adapting illumination. The focus of the present study is to understand preference and tolerance for gray reproduction on reflective media. The main goal is deriving gray-balance printer calibration that is visually preferred and robust across viewing illuminations.

A series of psychophysical experiments were conducted, beginning with simple targets and carefully controlled viewing conditions, and progressing to more complex targets and varied viewing conditions. These are described next.

Experiment 1: constant-color patches

In this experiment, a CIELAB target was designed with 16 patches whose a*-b* coordinates fell on a 4 x 4 grid in the a*-b* plane. The grid levels for both a* and b* were [-4, -2, 0, 2]. The asymmetry about the (0,0) point was the result of a pilot study which showed that samples for which both a* and b* were significantly positive were seldom selected as gray. The L* value of all the patches was 70. The 16 patches were printed as 0.5”× 6.5” horizontal strips on standard 8.5”× 11” paper. Given the fact that the human visual system is most sensitive to gray variations at very low spatial frequencies [3], each patch was designed to cover a fairly large visual field. Also, it was anticipated that the visual perception of each patch would be strongly affected by surrounding patches; hence 3 targets comprising different layouts of the same patches were generated.

The three CIELAB targets were converted to CMY coordinates using a characterization profile for a Xerox DocuColor 12 xerographic printer. Although this device also employs a black (K) colorant, the profile was derived to put out only CMY, in order to avoid metamerism issues that can arise when viewing 4-colorant prints under different light sources. The CMY targets were then printed on the DocuColor12, and spectral reflectance measurements of the patches were made by a Gretag Spectrolino spectrophotometer. The spectral measurements from the 3 targets were averaged together, then converted to CIELAB coordinates. Three light sources were used in this study: D50, A, and cool white fluorescence (CWF). The reference white for the CIELAB calculation was the tristimulus measurement of the paper under each given source. Thus complete adaptation to the paper was assumed. The (a*, b*) coordinates of the measurements of the 16 patches, computed under each of the three light sources, are shown in Fig. 1.

Twenty-eight observers, all reporting normal color vision, participated in the psychophysical experiment. The targets were placed one at a time in a GretagMacbeth Spectralight III light booth with neutral surround, and viewed from a distance of 6 feet. One of the 3 aforementioned light sources was used to illuminate the print. All other ambient illumination was turned off. The task was to identify the patches perceived as being gray, or closest to gray. Since no other reference gray was made explicitly available for comparison, observers were making judgments based on preference and/or memory under the given state of adaptation. Observers viewed each of the three target layouts under each of the three illuminations, thus recording responses under a total of 9 experimental conditions. Each time the illumination was changed, observers adapted to a blank sheet of paper for approximately 30 seconds. The ordering of targets and illuminations was randomized from observer to observer.

The observer responses were tallied to generate preference histograms, described with the following notation. Let i and j denote the levels along the a* and b* dimensions on the original 4 x 4 grid. Hence i, j = [-4, -2, 0, 2]. Let patch[i, j] denote the patch whose original CIELAB target values are a* = i and b* = j. Also, let k = [1, 2, ..., 9] denote an experimental condition comprising a particular combination of target layout and illumination. Then $H_k[i, j]$ is defined as the number of times observers chose patch[i, j] under experimental condition k.

Figure 2 shows plots of $H_k[i, j]$ for two of the target layouts under D50. Clearly, observers' responses depended on the physical layout, thus re-iterating the fact that color perception is strongly affected by the immediate surround. However, a common trend can be seen in that the histogram peaks generally occur at patches whose original input target values for a* and b* are both negative. Figure 3 shows results under CWF and A. Again, the peaks occur at those patches measuring at the negative extrema of a* and b*.

Inherent errors in the calibration and measurement process result in the D50 data deviating from the original input grid. However, these errors are not a major concern, since the main idea was to select a set of points whose measurements would cover a region around the colorimetric neutral point (a* = b* = 0). As expected, the measurements change as a function of the light source. Note also that for each of the three sources, there is a patch that lies very close to the colorimetric neutral point. (It is a different patch in each case.)

It must be re-iterated that the actual CIELAB measurements for patch[i, j] do not coincide with the input CIELAB target values [i, j], and furthermore, vary as a function of illumination, as shown in Fig 1. Hence the histogram plots of Figures 2 and 3 should really be correlated with the measurements in Fig 1.

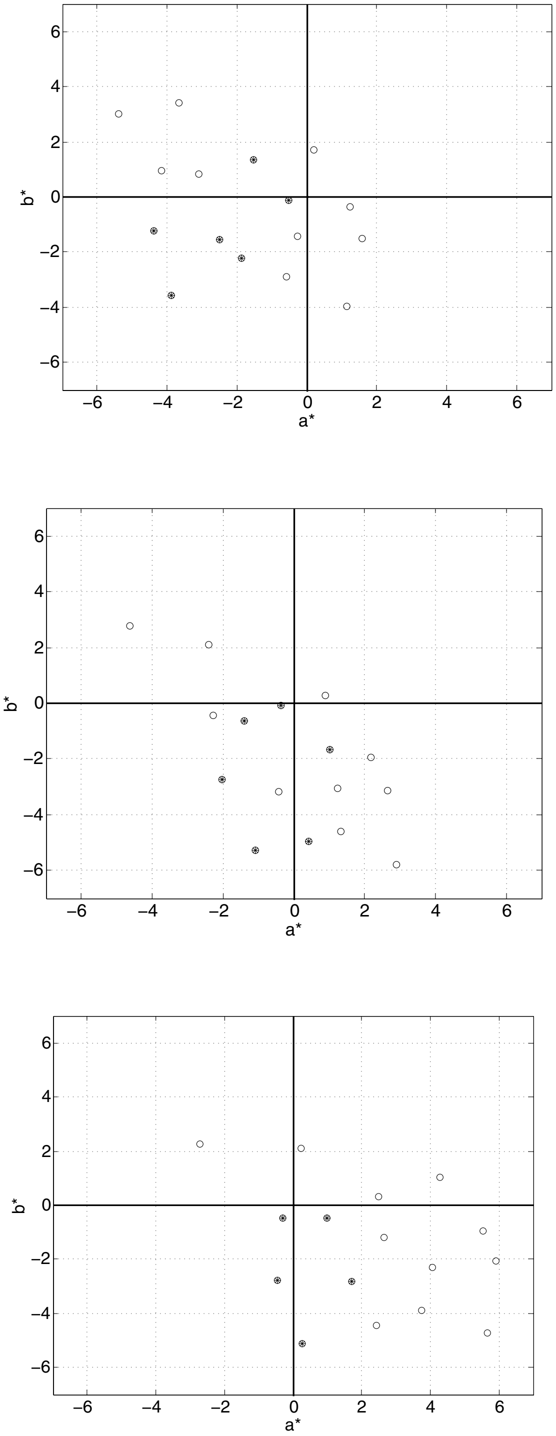

Figure 1: CIELAB measurement for the 16 patches in Experiment 1 under D50 (top), CWF (middle), and A (bottom). The marked points fell within a 75% confidence region of "preferred gray" in the psychophysical experiments.

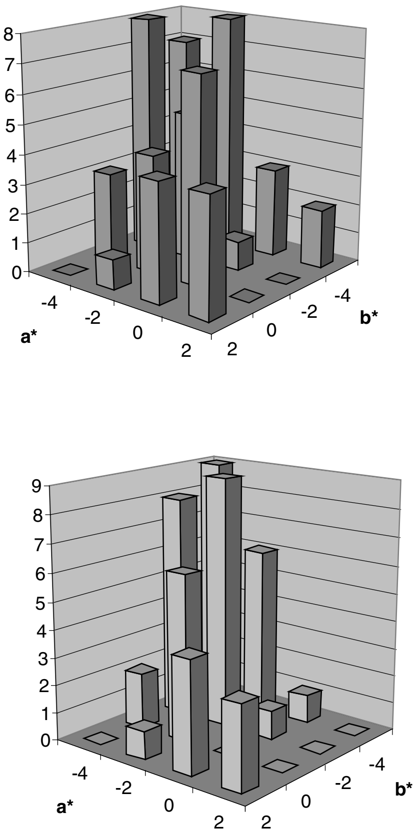

Figure 2: Preference histograms $H_k[i, j]$ for 2 different target layouts under D50.

To gain some insight into how the histograms relate to the measurements, the following analysis was done. The histograms corresponding to the 3 target layouts for D50 were summed to obtain an aggregate histogram for D50. The histogram entries were then sorted in decreasing order. Beginning with the measurement corresponding to the largest histogram entry, each measurement was marked until the cumulative histogram count reached 75% of the total histogram count. The marked points are shown in Fig. 1(a). Similar analyses were carried out for CWF and A sources, shown in Fig. 1(b) and 1(c). The marked points provide an indication of the 75% confidence region about each histogram peak. Clearly, the majority of them lie in the quadrant where both a* and b* are negative. The point that lies closest to a*=b*=0 is always marked; while points with significantly positive a* or b* are never marked.

While it is instructive to examine results for each illumination separately, in reality the same print is often viewed under multiple illuminants. Thus one must also consider preference of a given patch across illuminants. To this end, two aggregate histograms were generated from the 9 individual histograms. The first function, $H_t[i, j]$ is defined as the total number of times that patch [i, j] was chosen over all 9 experimental conditions. The second quantity, H_m [i, j] is the least number of times any given patch was chosen across all 9 conditions.
Mathematically:

$$H_t[i,\ j] = \sum_{k=1}^{9} H_k[i,\ j]; \qquad (1)$$

$$H_m[i,\ j] = \min_k H_k[i,\ j] \qquad (2)$$

Plots of these histograms are shown in Fig 4. Once again, the peaks occur for patches corresponding to measurements of negative a* and b*, or colloquially, the “cooler” grays.

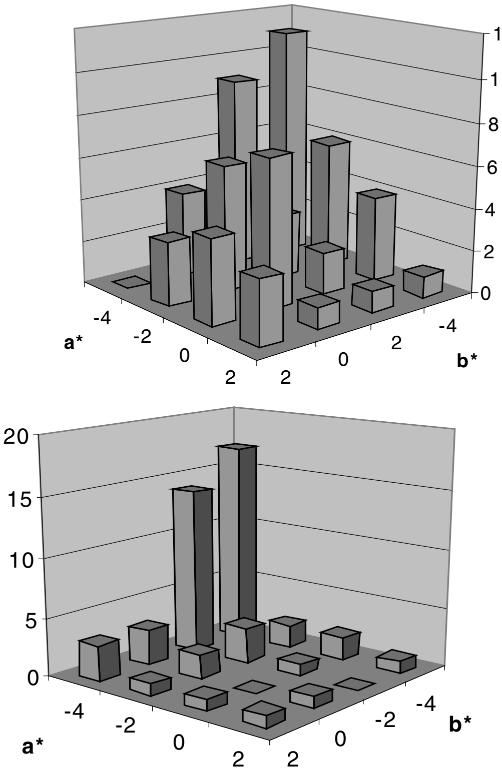

Figure 3: Preference histograms $H_k[i, j]$ for a single target layout under CWF (top) and A (bottom).

It is noteworthy that many observers chose a patch measuring negative values of a* and b* in preference to a patch measuring very close to colorimetric neutral. Furthermore, patches that measured positive a* or b* were seldom chosen as gray. This indicates a low tolerance for near-neutral colors that are on the “warmer” side.

As seen in Fig. 2-4, the histogram peaks often occur at one extreme in the negative quadrant of the chrominance plane. This suggests that the true peak may have been missed, and that it may be necessary to explore patches in this quadrant with even larger chroma. Secondly, all patches were simple constant-color stimuli, with no variation in L*. Thirdly, this experiment was conducted under tightly controlled viewing conditions; hence it was not clear if the results would generalize to more practical scenarios (*e.g.* office environment with mixed lighting). To address these concerns, a follow-on experiment was conducted, described in the next section.

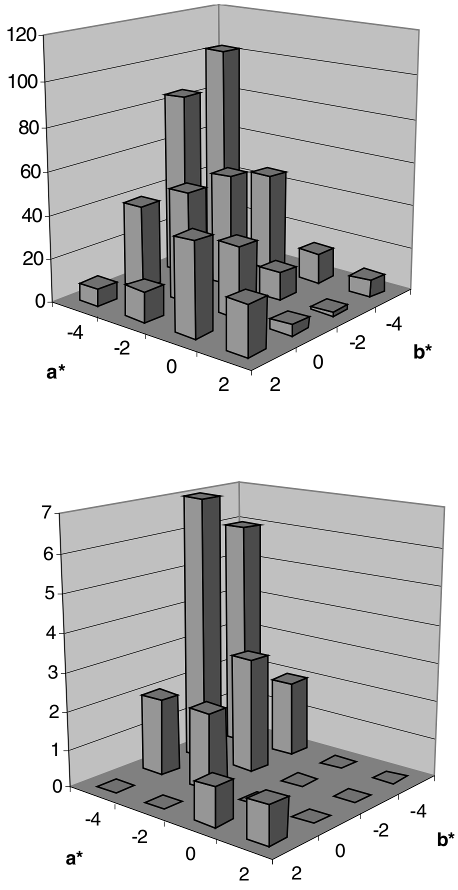

Figure 4: Aggregate preference histograms: $H_t[i,j]$ (top), $H_m[i, j]$(bottom)

Experiment 2: lightness sweeps

In this experiment, a smaller number of [a*, b*] chrominance pairs were considered. The patch that yielded the most favorable results in the first experiment, namely [a*, b*] = [-4, -4], was included, along with the following target values in the vicinity of this point: [a*, b*] = [-2, -4], [-6, -4] and [-4, -6]. Also, the D50 colorimetric neutral was included, since this represents common industry practice.

Sweep targets were generated, fixing the chrominance coordinates to one of the five chosen points, and varying the L* from 80 to 50. Ample white space was included to facilitate complete adaptation. These CIELAB targets were then processed through the DocuColor12 profile and printed, as was done in the first experiment. For each sweep, measurements were made at several locations on the page, and then averaged. The measurements under D50 illumination are shown in Fig 5. The points are labeled so that they can be conveniently correlated with results from the psychophysical experiment.

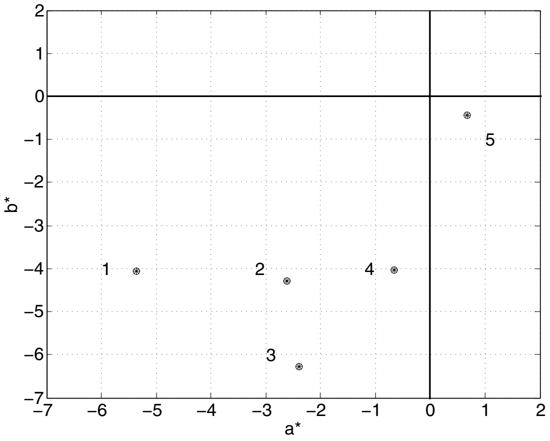

Figure 5: Average chrominance measurements, under D50, of the five sweeps used in Experiment 2.

Twenty-four observers participated in a pair-wise comparison experiment, where the task was to select from each pair of sweeps, the one that was closest to “gray”. To test robustness across different viewing illuminations, the task was performed in each individual’s office. Hence the illumination varied widely, including some mixture of CWF, daylight from windows, and occasionally, incandescent lamps.

The pairwise comparison data can be converted to a preference scale using models such as Thurstone’s law of comparative judgement [4] or the Bradley-Terry model [5]. These preference scales assign a score to a given treatment that indicates its performance relative to the other treatments. In the author’s experience, the outputs of these models are quite similar; however, the Bradley-Terry model provides some additional insight on confidence intervals for each estimated scale. Hence, this model was adopted to generate preference scales for this experiment. The result is shown in Fig 6. Sweeps 1-3 performed the best, with the differences among them being statistically insignificant. From Fig. 5, these sweeps all lie in the negative a*-b* quadrant. Interestingly, the sweep which measured as colorimetric neutral under D50 performed significantly worse than the other sweeps. Recall that the viewing illumination under which the prints were observed was often quite different from D50. Indeed, the combination of media, colorants, and illumination can very well shift the D50 neutral towards a region with positive a* or b*. In these regions, observers readily notice a “warm” hue, and do not accept the stimulus as being “gray”.

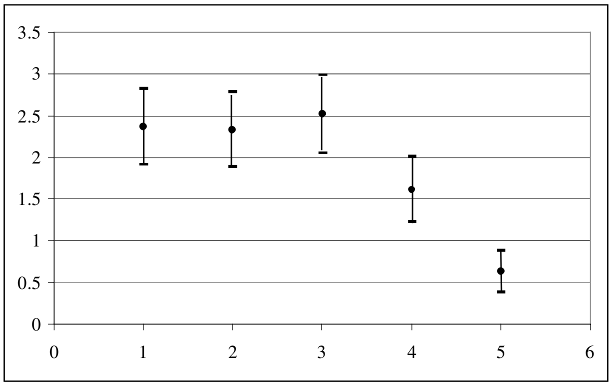

Figure 6: Interval scales for Experiment 2. The stimuli are numbered in accordance with the labels in Fig. 5.

Experiment 3: complex images

The simple stimuli used in the previous experiments are useful for quantitative analysis; however they are not representative of images typically used by consumers. This section describes an initial experiment to study gray preference in realistic images, and under nominal office viewing conditions. A set of 5 images was used, shown in Fig. 7. One image was a business graphic, generated in Microsoft PowerPoint, with a gray (R=G=B) sweep as the background. This image was converted to CIELAB assuming sRGB for the original image colorimetry. The remaining four were pictorial CIELAB images. Two of these, Bridge and Tailor, were color images with some neutral content, and the remaining two, Lady and Lighthouse, were grayscale versions of color images. The latter were obtained by setting the a* and b* channels of the color images to zero.

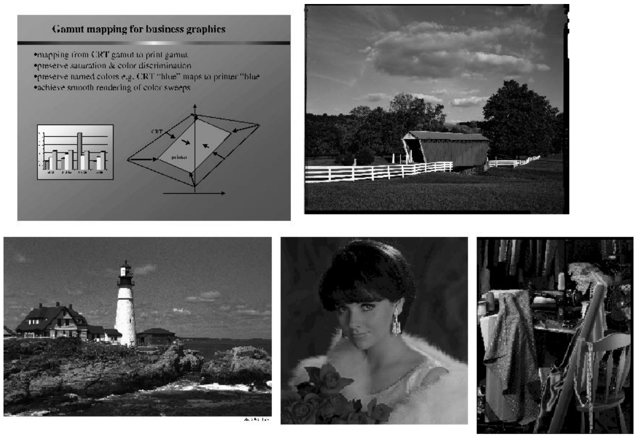

Figure 7: Grayscale thumbnails of the images used in Expt. 3. Clockwise from top left: Graphics, Bridge[1], Tailor[1], Lady, Lighthouse[2]. (Sources: [1]GATF; [2]Eastman Kodak Co.)

Two hardcopy renditions of each image were made. In one case, the CIELAB image was mapped through the colorimetric profile for the DocuColor 12 and printed. The profile was characterized using media-relative colorimetry under D50 illumination. In the second case, the image first underwent a local warping applied to the neutral axis to shift colorimetric neutrals towards preferred gray. The warping function was chosen to map the neutral axis (i.e. $0 \leq L^* \leq 100$, a*=b*=0) smoothly to a curved locus that passes through the CIELAB point [70, -2.75, -3.5]. This point was chosen as an average among the highly ranked data in the previous experiments. The warping preserved L* and can be described by two 1-D functions, $f_a(L^*)$ and $f_b(L^*)$ that warp a* and b*. Plots of these functions are shown in Fig. 8.

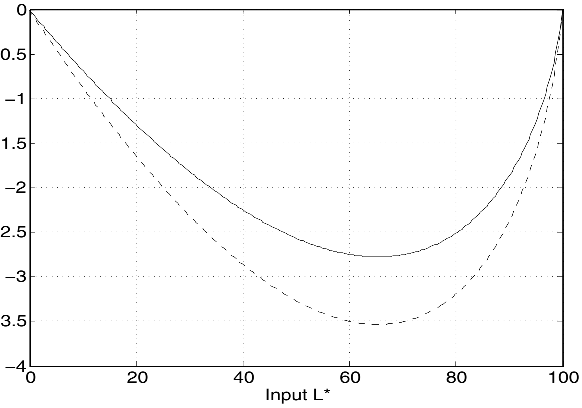

Figure 8: Warping functions f_a (solid) and f_b (dashed) used to warp a and b*coordinates towards preferred gray.*

The effect of the warping decayed smoothly as one moved away from the neutral axis. Specifically, given an input CIELAB value, $[L^*_{in}, a^*_{in}, b^*_{in}]$, the a* shift for the given input color was given by:

$$a^*_{out} = a^*_{in} + f_a(L^*_{in}) e^{-\alpha (C^*_{in})^2}, \qquad (3)$$

where C^*_{in} is the chroma of the input color, and α is a parameter that controls the rate of decay. A corresponding shift was applied to b*. It was desired that the warping function exhibits a significant effect around the neutral axis, yet have little effect on other memory colors such as fleshtones. To this end, a decay constant of $\alpha = 0.004$ was heuristically chosen for the experiments.

Six observers participated in a forced-choice experiment. The task was to choose the preferred image among the two reproductions. The experiments were conducted in the observers' offices; hence viewing conditions were uncontrolled, but realistic. For the graphics image alone, a softcopy version of the image was displayed on the computer in the given observer's office, to serve as a reference original. Observers were allowed to directly compare the hardcopy and softcopy versions of this image in typical office lighting.

Table 1 shows the results of the experiment. For the graphics image, all 6 observers picked the warped gray reproduction. In addition to supporting the previous findings, this result may also be indicative of the fact that the white point of typical displays is often bluer than the adapting white in typical hardcopy viewing conditions. Hence the warped gray reproduction, which moves neutral colors towards a bluish hue may actually result in a better representation of the softcopy version when both are viewed simultaneously under the same ambient lighting. For the Bridge image, the results were tied. Most observers commented that they could not notice a significant difference among the two reproductions; there was only a small amount of gray content in the clouds. The Tailor image contains significant neutral content, hence as expected the warped gray reproduction was strongly preferred. Results for the grayscale pictorials were strongly dependent on image content. For the scene containing fleshtones, observers disliked the bluish caste produced by the warped gray algorithm. On the other hand, for the scene containing sky and water, observers preferred the bluish cast.

Table 1: Results of Expt 3, showing the number of times the "warped gray" reproduction was selected as the preferred image.

Image	# of "warped gray" selections (out of a total of 6)
Graphics	6
Bridge	3
Tailor	5
Lady	1
Lighthouse	4

Conclusions

The experiments described in this paper indicate that preferred gray in reflection prints occurs in a region in the chrominance plane which lies predominantly in the quadrant where a* and b* are both negative. Observer preference is broadly peaked in this quadrant. Visual tolerance for gray appears to be asymmetric about $a^*=b^*=0$, and reduces rapidly for colors in quadrants where a* or b* is significantly positive (i.e. greater than 2).

The second experiment strongly suggests that gray-balancing for D50 neutral may not be the most desirable goal in a practical color management system. Calibration errors and changes in viewing illumination could shift the colorimetrically gray input in a direction that results in a visually unacceptable reproduction (namely positive a*, b*). If, on the other hand, a device is gray-balanced for chrominance values that lie substantially in the negative quadrant, this allows for more robustness to system errors while remaining in a safety zone of colors that are visually perceived as gray.

The experiments on complex images indicate that a shift towards "cooler" grays applied locally in color space is desirable for graphics and color pictorials. For grayscale pictorials, preference depends strongly on image content. In practical applications, grayscale images are often rendered with a black (K) colorant, which allows no freedom to adjust the chrominance of gray; hence the issue is moot.

The assumption that observers adapt completely to the medium under a viewing illumination must be more carefully tested. For example, in the case of illumination A, it is likely that adaptation is incomplete; and observers select the bluer neutrals in part to compensate for the yellowish light source [2]. Gray perception is thus likely affected by a combination of adaptation and preference.

References

1. H. R. Kang, *Color Technology for Electronic Imaging Devices*, SPIE, Bellingham, WA., 1997, Chapter 4.
2. R. W. G. Hunt and L. M. Winter, "Colour adaptation in picture-viewing situations", *J. Photogr. Sci.*, **23**, pp. 112-115, 1975.
3. P. K. Kaiser and R. M. Boynton, *Human Color Vision*, OSA, Washington, DC., 1996, pp. 399-409.
4. L. L. Thurstone, "A law of comparative judgment", *Psych Re*v., **34**, pp. 273-286, 1927.
5. Ralph A. Bradley, "Paired comparison: some basic procedures and examples", *Handbook of Statistics*, **4**, pp. 299-326, 1984.

Biography

Raja Bala received the B.S. degree from the University of Texas at Arlington in 1987, and the M.S. and Ph.D. degrees from Purdue University in 1988 and 1992, respectively, all in Electrical Engineering. Since then, he has been employed at the Xerox Digital Imaging Technology Center, where he is a Principal Scientist leading a color science project. He holds 13 patents and over 30 publications in the field of color imaging. He is a member of IS&T.