

# Viewing Environments for Cross-Media Image Comparisons

*Karen Braun and Mark D. Fairchild*

*Munsell Color Science Laboratory, Center for Imaging Science  
Rochester Institute of Technology, Rochester, New York*

## Abstract

The reproduction of color images in various media typically involves changes in viewing conditions as well as the more obvious changes in the physical properties of the imaging systems. Basic colorimetry, typified by the CIE 1931 XYZ system, can adequately predict color matches across various physical media as long as viewing conditions are identical. However, once changes are made in factors such as illuminant color (white point), illuminance level, surround, and mode of viewing, color-appearance models are required to predict and produce matching images. Several color-appearance models have been proposed. This work concentrates on those that have been published by Fairchild and Berns (RLAB), Hunt, Nayatani *et al.*, von Kries, and the CIE (CIELAB). The quality of the color appearance matches predicted by each of these models is quantitatively scaled for CRT (D65) reproductions of printed images (A and D50) using a paired-comparison paradigm. The data were analyzed using the law of comparative judgments to determine an interval scale of prediction quality for the various models. The experiments were repeated for 5 different viewing arrangements: memory, successive binocular, simultaneous binocular, haploscopic, and successive-*Ganzfeld* haploscopic. The surround and luminance levels are matched for these preliminary experiments. The results indicate that the memory matching technique is most appropriate for psychophysical evaluation of cross-media image reproductions. They also suggest that RLAB performs best.

## Introduction

The overall objective of this work was to test various models that are capable of predicting the tristimulus values of images in different viewing conditions that might be considered a visual match. It has been determined that five models are most important for consideration in this work: RLAB<sup>1,2</sup>, Hunt '93<sup>3,4</sup>, CIELAB, Nayatani<sup>5</sup>, and von Kries<sup>6</sup>. In designing an experiment to test these models, one of the most important considerations is viewing technique. This preliminary study investigated different viewing techniques to determine which is optimum for cross-media color reproduction. Psychophysical experiments were performed by fifteen observers using five different viewing techniques for comparison of hard-copy originals to soft-copy images which were derived with one of the five models. These techniques were memory matching, successive binocular matching, simultaneous binocular matching, successive haploscopic

matching, and simultaneous haploscopic matching.

For each viewing technique, observers judged how well each model reproduced CRT images of a hard copy original when illuminated in a light booth under sources approximating CIE illuminants D50 and A. The objective of this study was to compare the results of these five experiments to see which techniques correlated well with each other, which had the most observer agreement, and which were easiest and most preferable for the observer.

## Viewing Techniques

Five viewing techniques listed above were compared in this experiment. The white point of a Sony GDM-1950 monitor was adjusted to approximate D65. Two sources were used in the booth, approximating illuminants D50 and A. Thus, comparisons were made under two sets of conditions: monitor D65 to booth D50 and monitor D65 to booth A. For this preliminary study, the maximum luminance levels of the booth and the monitor were equal.

The psychophysical experiments were conducted with the room lights off. Observers sat approximately 1 meter from the original images and the monitor screen in all experiments.

## Memory Matching

In the memory matching technique, the original hard copy image and the soft copy reproduction were placed at ninety degrees from each other with respect to the observer, as shown in Fig. 1. This ensured that the observer could not see both images at the same time. To prevent stray light from the booth from falling on the monitor, a curtain was drawn across the booth while the observer inspected the soft copy reproductions. Similarly, while the observer looked at the hard copy original, the monitor was always blacked out.

Observers adapted to an 18% gray card in the light booth at D50 for 60 seconds. (Fairchild and Reniff<sup>7</sup> showed that chromatic adaptation at constant luminance is 90% complete after approximately 60 seconds.) The observers then studied the hard copy original for at least 60 seconds. Observers then turned toward the monitor and adapted to a neutral gray field at a white point of D65 for 60 seconds. Then observers compared the pairs of soft copy reproductions and chose the reproduction which looked most like the original hard copy image. After making such choices for each of ten comparisons, the observer turned back to the booth and repeated the process for each of five test images. The experiment was repeated for the second light source, Incandescent A.

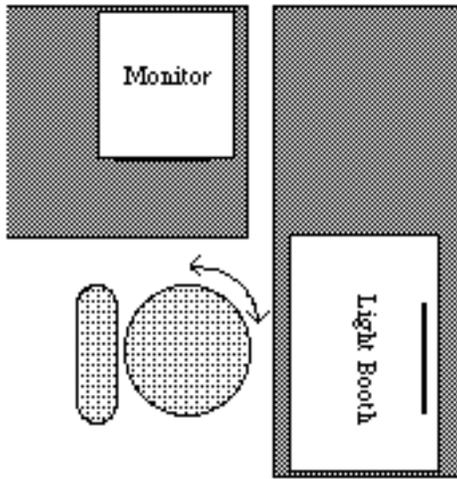


Figure 1. Set-up used for memory matching and successive binocular matching experiments.

### Successive Binocular Matching

Successive binocular matching was similar to memory matching except that the observers could look back at the original at any time. When they looked back, they adapted to the original for 60 seconds before looking at the original, and 60 seconds to the monitor before continuing. The observers could never see both the original and the reproduction at the same time because, again, the monitor went blank when the observers were examining the print original and a curtain covered the booth when the observers looked at the monitor. The set-up is the same as in Fig. 1.

### Simultaneous Binocular Matching

In simultaneous binocular matching, the original and reproductions were side by side and coplanar. The observers adapted to the “environment” for 60 seconds by viewing the gray card in the booth and the gray field on the monitor simultaneously. Then the images appeared on the monitor and the observers began making decisions. The observers could see both the original and the reproduction with both eyes at all times. This set-up is shown in Fig. 2.

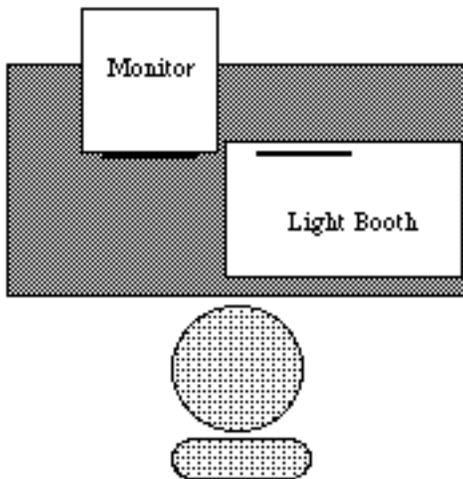


Figure 2. Set-up used for simultaneous binocular matching experiment.

### Simultaneous Haploscopic Matching

In simultaneous haploscopic matching, the observers examined the original with one eye (the right eye in the set-up used here) and the reproduction with the other. This set-up allows each eye to be adapted to a different white point while allowing direct comparison of the images. For haploscopic matching techniques, it is assumed that each eye reaches its adaptation state independently of the other eye. Figures 3 and 4 show the simultaneous haploscopic set-up.

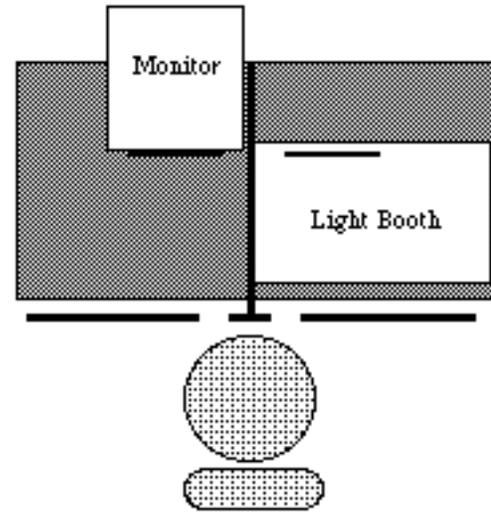
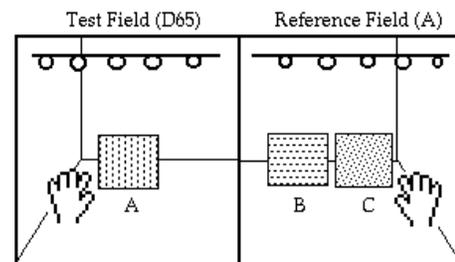


Figure 3. Top view of set-up used for haploscopic matching experiments.



Which sample (B or C) is the better match to the sample A?

Figure 4. Front view of set-up used for simultaneous haploscopic matching experiment.

### Successive-Ganzfeld Haploscopic Matching

The successive-Ganzfeld haploscopic viewing method, described by Fairchild, Pirotta, and Kim<sup>8</sup>, was similar to simultaneous haploscopic viewing shown in Fig. 3, except that observers were restricted from seeing both images at the same time. A neutral diffuse filter, as known as a Ganzfeld, covered one eye while the other inspected an image. This technique assumes that the eye which is covered by the diffuse filter remains adapted to the appropriate white point. The observers controlled which eye was covered using a foot pedal. Figure 5 shows the successive haploscopic technique.

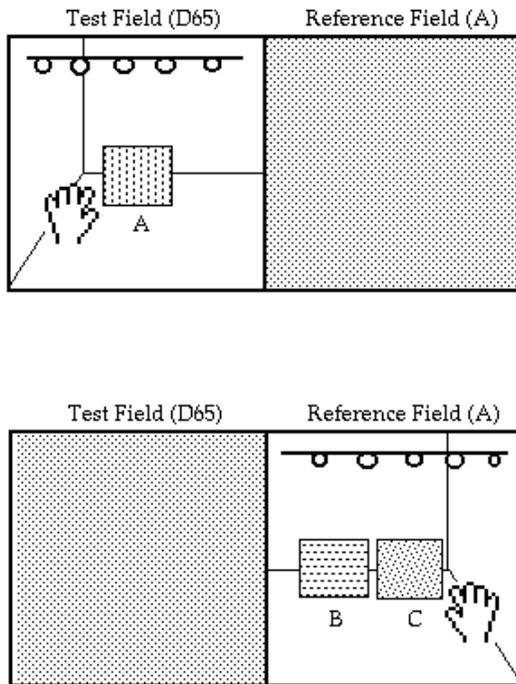


Figure 5. Front view of set-up used for successive haploscopic matching experiment.

## Generation of Images

### Hard-Copy Originals

Five digital images were used in this study: two pictorial, two graphic, and one hybrid. The images, printed on a Kodak XLT 7720 continuous-tone digital printer at 203 dpi, were 6" × 8". A half-inch border of white was included in each image, and this border was adjusted with the rest of the image.

These originals were digitized using a Howtek D4000 drum scanner at the 99 dpi, the monitor resolution. The 6" × 8" hard copy images were then mounted on 8" × 10" spectrally flat, 20% gray cards.

### Scanner Calibration

It was necessary to perform a scanner calibration so that scanner RGB tristimulus values could be accurately converted to CIE XYZ tristimulus values for the two spectral power distributions used in the booth, D50 and A. A calibration target consisting of 140 color patches was printed on the Kodak XLT 7720 printer and the spectral reflectance data of the color patches were measured using a Gretag SPM60 spectrophotometer. The CIE XYZ tristimulus values were calculated using these spectral reflectances, the CIE 1931 2° standard observer, and the spectral power distributions of the two sources, measured using a PhotoResearch PR 703.

Next, the calibration target was scanned with the Howtek scanner in order to obtain RGB scanner values for each of the 140 patches. A 3 × 10 matrix was used to characterize the relationship between scanner RGB tristimulus values and CIE XYZ tristimulus values of the printed target patches. This matrix was used to convert scanner RGB values of the print

originals to CIE XYZ tristimulus values for use in the models.

### Soft-Copy Reproductions

The CIE XYZ tristimulus values of the print originals served as input to the five color appearance models that were used to predict matching images for the D65-balanced CRT.

### Gamut Considerations

The first set of CRT images contained some colors that were out of the monitor's gamut, mostly very high and very low luminance colors. To eliminate gamut mapping considerations, the lightness scale of the original image data was compressed by 10%. This produced images that remained in gamut for all of the appearance transforms. The study of gamut mapping problems is an important, but largely separate issue, that was eliminated as a variable in this research.

## Psychophysics

Each soft-copy reproduction, produced with one of the five models, was compared to every other reproduction, giving a total of ten paired comparisons. The observers chose which of the two soft copy reproductions looked most like the original, using a classic forced-choice paired-comparison paradigm. Observers were given the following instructions:

In this experiment, you will be comparing a printed original image in the light booth to two reproductions on the CRT screen. Your task is to choose which of the reproductions looks most like the original (not necessarily which image looks better.) You will be judging the color of the images only. You can toggle between the two reproductions by using the 1 and 2 keys. To select an image, go to that image by toggling and then press the space bar.

Individual instructions were given for each of the five experiments, explaining the particular method that would be used.

The same observers were used whenever possible to reduce differences in the results of the viewing techniques caused by observer variability. Most observers were in the field of imaging or color science and a few had previous experience in color matching or color reproduction. They ranged in age from 21 to 40 years old. A total of 15 observers were used in each experiment.

## Viewing Technique Selection

Each of the five experiments resulted in a logistic score for each model reproduction of the five images under two illuminants. This score was proportional to how well the model accurately reproduced the original. Logistic scores for each experiment were averaged over the five images to give an average score for each model under the two illuminants.

The viewing techniques were compared to determine which viewing technique should be used for further study of cross-media color reproduction. Criteria used included which technique is most natural, or most like real viewing conditions; which technique has the least inter-observer variation, or most observer agreement; and which technique is preferred by observers, easiest to perform and most comfortable.

### Most Realistic Technique

Hard copy images are often directly compared to CRT images by holding the print next to the CRT. Thus the simultaneous binocular method is probably most typical. However, the observer's state of adaptation under these conditions is questionable, since he can not be adapted to both white points at the same time.

Memory matching is the next most natural method for comparing images. Hard copy images often have a final use where the observer is not comparing the print directly to the monitor original, and is therefore completely adapted to each light source independently. The difficulty with this experimental procedure is that an observer's memory of the original is not perfect, producing some noise in the results. Successive binocular viewing is a potential solution to this problem, however due to the relatively long adaptation period required between comparing the hard copy and soft copy (60 seconds), this method offers little improvement.

A possible solution to the difficulties of adaptation and memory discrepancies is the use of a haploscopic viewing technique. In such a technique, one eye is adapted to one illuminant and the other eye to the other illuminant. In this way, only a very short memory of the images is required (less than a second) but each image is viewed in an adapted environment. Difficulties with this technique are that it is an unnatural way to view and compare images, and it has not been proven that each eye is fully adapted or that the observer can sufficiently discount the illuminants.

### Expansion of Scale

Expansion of scale (sensitivity) was studied by comparing the amount of spread in the logistic scale. If all decisions were random guessing or if observer preference was divided equally among all the models, the frequency of selection would be 7.5 (on average for 15 observers) for each model and the corresponding logistic score would be 0.0. There would be no spread in the scale and all models would be considered equal. If most observers agreed in their choices, frequencies would be close to 0 and 15, giving large positive and negative logistic values.

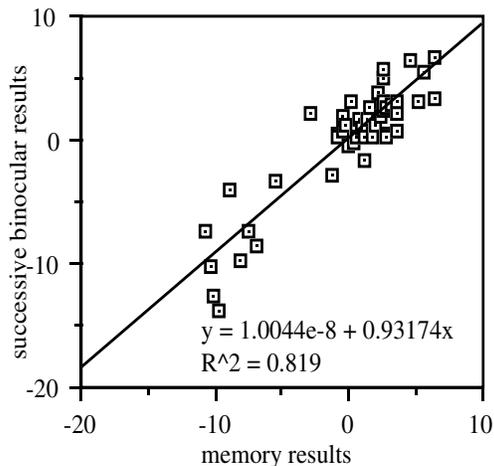


Figure 6. Correlation between memory and successive binocular matching techniques.

The amount of spread between two viewing techniques was compared by plotting the results of one experiment on

a logistic scale as a function of the other. A linear model was used to fit the data and the slope of the regression line indicated the spread. If the slope was 1.0, the two scales had the same amount of spread. If the slope was greater than 1.0, the experiment plotted on the y-axis had a greater spread than the experiment plotted on the x-axis.

Figure 6 shows results of the successive binocular matching experiment versus results of the memory matching experiment.

Each data point represents the logistic value of a given model for one image resulting from the two experiments. The data were pooled by averaging over the images, giving an indication of how each model did on average under each illuminant. Figure 7 shows the average logistic value of each model.

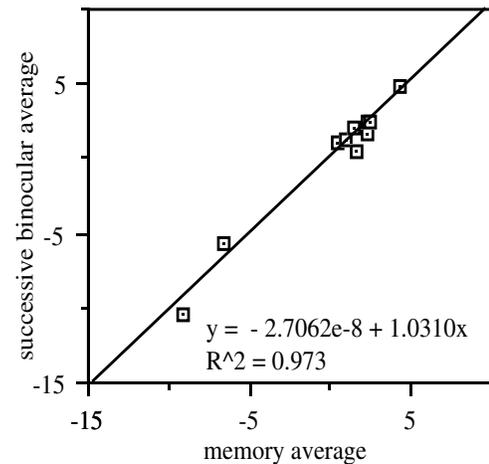


Figure 7. Averaged results of memory and successive binocular matching techniques.

When memory results were plotted on the x-axis and successive binocular on the y-axis as in Fig. 7, the slope of the regression line is 1.03. This appears in Table I, which shows slopes when the average data for each experiment was plotted against the average of each of the other experiments. Memory matching is abbreviated MEM, successive binocular is SCB, simultaneous binocular is SMB, successive haploscopic is SCH, and simultaneous haploscopic is SMH.

The columns represent the experiments plotted on the y-axis and rows give the experiments on the x-axis.

Table I. Slope of regression line for averaged data for each pair of viewing techniques.

	MEM	SCB	SMB	SCH	SMH
MEM	1	1.03	0.40	0.75	0.61
SCB	0.94	1	0.36	0.67	0.55
SMB	1.26	1.24	1	1.09	0.92
SCH	1.08	1.06	0.51	1	0.77
SMH	1.44	1.42	0.69	1.25	1
Avg.	1.145	1.149	0.593	0.950	0.773
Rank	2	1	5	3	4

Table I shows that the successive binocular technique had the most expanded scale of the techniques since, for this technique plotted on the y-axis, the slopes were always greater than one. This technique was followed closely by memory matching. The successive haploscopic scale was more expanded than both simultaneous binocular and simultaneous haploscopic. Simultaneous haploscopic had a slightly more expanded scale than simultaneous binocular. These relationships were true for both the pooled and unpooled data sets.

### Correlation Between Experiments

The correlation coefficient,  $R$ , is a measure of the amount of agreement between experiments when the experiments are plotted as shown in Figs. 6 and 7. When two experiments correlate well and one is a simpler technique, the simpler technique can be considered a representation of the other. When the coefficient of determination,  $R^2$ , is close to one, the results of the two experiments correlate well. Table II gives coefficients of determination for all combinations of experiments when only the averaged data are used.

**Table II. Coefficients of determination for pairs of experiments showing only averaged data.**

$R^2$	MEM	SCB	SMB	SCH	SMH
MEM	1	0.973	0.51	0.807	0.88
SCB		1	0.45	0.703	0.786
SMB			1	0.554	0.639
SCH				1	0.963
SMH					1

The coefficient of determination for the plot of technique B versus technique A was the same as for the plot of A vs. B. Therefore, half the table gives all necessary information.

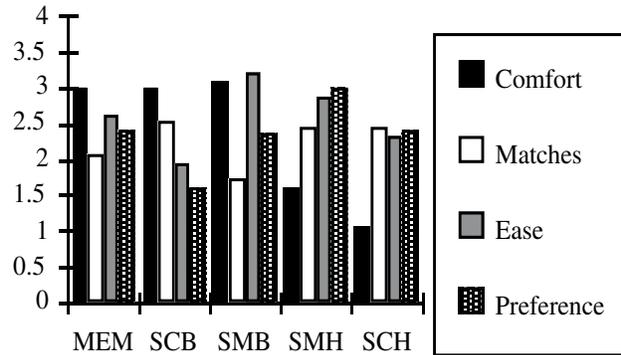
Table II shows that memory matching and successive binocular matching correlate well ( $R^2 = 0.973$ ), as do simultaneous and successive haploscopic viewing techniques ( $R^2 = 0.963$ ). Simultaneous binocular matching does not correlate well with any other technique, with  $R^2$  values less than 0.64. The correlation between binocular (not including simultaneous binocular) and haploscopic techniques varies between 0.703 and 0.880. This does not show strong correlation, implying that there is a fundamental difference between the two methods of viewing. This difference is likely due either to failure of the observer to fully discount the illuminant in the haploscopic techniques, or to adaptation, which may not be as complete for haploscopic viewing.

### Observer Preference

Observers were asked to judge the five experiments on physical comfort, goodness of matches, overall ease, and overall preference, on a scale from 0 to 4, where 0 was “worst” and 4 is “best”. Figure 8 shows the results of this survey.

Observers rated simultaneous binocular best for physical comfort. Memory and successive binocular also rated well. Many observers complained of back discomfort in the

haploscopic experiments, from holding their heads in the viewer for long periods of time. Observers found that the matches were not as good for the simultaneous binocular technique as they were for other experiments. This is because the observer is not clearly adapted to either condition and differences in illuminant are obvious to the observer. Observers found the successive binocular technique to be the most difficult overall and to be least preferred due mainly to the long adaptation times required if the observer chose to look back.



*Figure 8. Results of observer survey on a scale of 0 (worst) to 4 (best).*

### Model Performance

These experiments serve as the basis for a future series of experiments that will evaluate the performance of color-appearance models for CRT-to-print color reproduction across a wide range of viewing conditions. The results of these experiments do provide a preliminary view of the various models’ performance. The RLAB model consistently performed better than the other models for the situations studied. The mean scale values (averaged across image content and illuminant change) were: RLAB, 3.19; Hunt, 0.57, von Kries, 0.42, CIELAB, 1.81, and Nayatani, -5.00. Higher scale values indicate better model performance. Scale value differences of 1.07 are statistically significant at 95% confidence. The RLAB model performed significantly better and the Nayatani model performed significantly worse than all of the other models in these experiments.

### Conclusions

Since simultaneous binocular viewing does not allow the observer to adapt to either the print’s or the monitor’s white point, it should not be used. The correlation between the remaining binocular experiments and the haploscopic experiments was not high enough to conclude that those experiments gave the same results. Therefore, since the binocular experiments are a more natural way of viewing and comparing images, the haploscopic techniques should not be used.

Two factors make memory matching the recommended technique, even though the successive binocular technique showed a slight expansion of scale. First, as stated earlier, observers looked back an average of only 4.1 times during one hundred comparisons, making the two techniques nearly the same. Second, observers rated the memory technique

significantly easier and more preferable than the successive binocular technique. Since Tables I and II shows that successive binocular matching only slightly increases the range over memory matching, memory matching is chosen as the recommended viewing technique.

### Acknowledgements

This research was supported by the Eastman Kodak Company. Additional funding was provided through the NSF-NYS/IUCRC and NYSSTF-CAT Center for Electronic Imaging Systems. The authors thank Paula Alessi, of Eastman Kodak, for her hours of assistance to this project, including six hours of observations. The time and patience of each of the observers performing the five experiments is greatly appreciated.

### References

1. M.D. Fairchild, Formulation and testing of an incomplete-chromatic adaptation model, *Color Res. Appl.* **16**, 243-250 (1991).

2. M.D. Fairchild, R.S. Berns, Image color appearance specification through extension of CIELAB, *Color Res. Appl.* **18**, 178-190 (1993).
3. R.W.G. Hunt, Revised colour-appearance model for related and unrelated colours, *Color Res. Appl.* **16**, 146-165 (1991).
4. R.W.G. Hunt, An improved predictor of colourfulness in a model of colour vision, *Color Res. Appl.* **19**(1) (1994).
5. Y. Nayatani, K. Takahama, and H. Sobagaki, Field trials on color appearance of chromatic colors under various light sources, *Color Res. Appl.* **15**, 210-221 (1990).
6. J. von Kries, Chromatic adaptation, *Festschrift der Albrecht-Ludwig-Universität*, (1902).
7. M.D. Fairchild and L. Reniff, Time-course of chromatic adaptation for color-appearance judgements, *J. Opt. Soc. Am. A* **12**, 824-833 (1995).
8. M.D. Fairchild, E. Pirotta, and T. Kim, Successive-Ganzfield haploscopic viewing technique for color-appearance research, *Color Res. Appl.* **19**, 214-221 (1994).

published previously in the IS&T 1994 Annual Conference Proceedings, page 391

