MIND-A New Psychophysical Algorithm for Determination of Chromatic Adaptation

Peyma Oskoui, Elizabeth Pirrotta*
Hardcopy Technologies Laboratory
HP Laboratories Palo Alto
HPL-1999-158
December, 1999

CRT white point, adapted white point, chromatic adaptation, viewing condition, monitor characterization, psychophysics, neutral point, mixed mode illumination, incomplete adaptation

An iterative psychophysical algorithm is designed and implemented to determine the adapted white point for cathode ray tubes (CRTs) in a given viewing condition. This algorithm can accurately determine the chromatic adaptation point of mixed mode lighting environments. The accuracy of the algorithm has been empirically verified. A detailed description of the algorithm, type of experiment for which this algorithm has been used for, and the empirical results of its application are presented here. Some further usage of such an algorithm is also discussed.

* E. Pirrotta is an independent consultant based in Mountain View, California

I. Introduction

The impetus for designing MIND (Multi-stimuli Interactive Neutral-point Determination) was the need for a psychophysical algorithm to determine the adapted white point of cathode ray tubes (CRTs) under mixed mode illumination conditions. What is adapted white point and why do we need to determine it for CRTs? When a given observer views images on a computer screen, his visual system is affected by both the lighting illuminating the room and the luminance of the computer screen. In a typical office there are more than just one source of illumination: The CRT, the fluorescent lights, daylight from windows, these all affect how the visual system perceives image colors on the monitor. What "color" or chroma the observer does perceive as white under certain lighting conditions is called the adapted white point or the chromatic adaptation point. The degree of influence of each of the illuminants on the visual system in a mixed mode environment is not known off hand. Moreover, knowing the exact adaptation becomes crucial if we propose to use the CIECAM97 color appearance model. This model proposed by Hunt and Luo[1] takes into account many parameters of the viewing condition such as the surround and the chromatic adaptation point in order to accurately model the color transformation model from one viewing condition to another. If we feed the wrong chromatic adaptation point to the model, the results of the model will no longer be correct. Moroney[2] has investigated the tolerances of input parameters to CIECAM for the sRGB standard. The importance of accurate white point adaptation used for the CIECAM97 model is evident in his results: a change of more than ± 0.1 in either x or y direction of the white point causes the results to deviate more than 3ΔE (in CIECAM space).

We also conducted a similar experiment on tolerance of the CIECAM97 model. We measured the tristimulus values of patches in a Q60 color chart, and used those as the input to the CIECAM97 model. We used the daylight illuminants D55, D60 through D70 at 1-degree increments, and D75 as the input adapted white point. The set of results in Jhc space (CIECAM space) was then compared to the results of the D65-white-point case using CIECAM ΔE. Figure 1 shows a plot of the adapted white points used and their resulting ΔEs from the D65 case. This plot in effect shows that if we deviate from the true adapted white point by +/-250^{o}, we will have an average (CIECAM)ΔE of 1.2 and a maximum ΔE of 3. Hence just a 250^{o}K deviation from the true adapted white point can cause a noticeable color shift. Given that the rest of the process will not be error-free, and that this ΔE is only due to the wrong adapted white point, we can easily end up with an unacceptable result from the CIECAM97 model. This study again shows us how important it is to come up with an accurate algorithm for predicting the adapted white point.

The situation becomes worse under mixed mode illumination where the determination of the correct adapted white point for CRTs can be very difficult. We therefore need an experimental model to empirically determine the adapted white point. Using the empirical results, we can then extrapolate a function to determine the adapted white point given the specifics of any viewing condition.

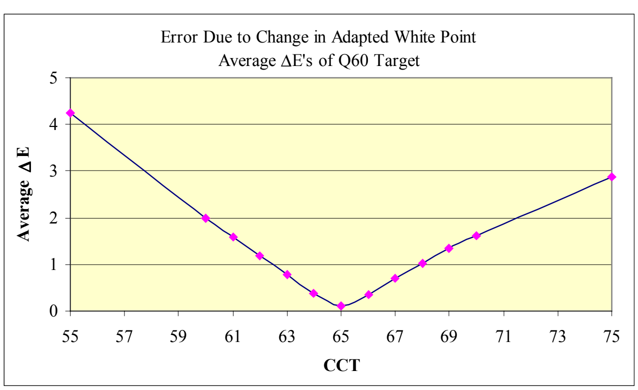


Figure 1. The 224 color patches of a Q60 target are measured under a D65 illuminant. We apply CIECAM97 whose chromatic adaptation point is set at another CCT, from D55 to D75, and measure the difference in output versus input. When the correct adaptation point, D65, is fed to CIECAM97, the ΔE goes to zero, as expected, and the output matches the input. A 250-degree error in adaptation point causes ΔE of 1.2.

In design of a psychophysical experiment we had two main objectives: that the observer's adapted white point be determined accurately and that the results be robust and have low variance. Repeatability was essential to achieve low error, and accuracy was needed due to sensitivity of the CIECAM97 model – as discussed above.

We first decided to use the method of adjustment[3]. This is a proven psychophysics method where the observer is asked to adjust the color of a patch on screen until it looks achromatic (neutral) to him[4]. CIELAB space is used, and the patch lightness is kept constant, while the observer uses the arrow keys to move left/right in a* and up/down in b* directions. Hence by adjusting the amount of red, green, yellow, and blue in the patch, the observer attempts to achieve achromaticity in the patch. This method is proven to be highly accurate; unfortunately it is prone to be noisy and many observers and repeats are needed to achieve the high level of accuracy. One reason for the noise in the results is that it is cumbersome for observers to perform the experiment since it can be tiring and time intensive. We used 8 observers, each adjusting a total of 13 patches. This set of experiments was conducted in a totally dark room where the only source of illumination was the D65 CRT. Hence the expectation was that the observers would completely, or near completely, adapt to the one source of illumination, and that the results would measure close to 6500 degrees Kelvin. Using analysis of variances on the results we concluded that the results were extremely noisy and of high variance. The average color temperature of the resulting achromatic patches was about 1000 degrees Kelvin higher than the CRT temperature of D65 (6500^{o}K). This fact alone was proof that the experiment was inaccurate as well as noisy. To achieve better accuracy in method of adjustment experiments we need to either have a much larger pool of observers or to have each observer repeat the experiment many times. Not having the luxury of making these modifications, we decided to change the method and base our experiment on the method of constant stimuli[5].

In method of constant stimuli the observer is presented with N patches on the screen. He is instructed to choose the one patch that is the most achromatic of the available choices. Based on his selection, the algorithm presents another set of N patches, and the process repeats until a cease criterion is reached. We went through many modifications in the algorithm, but from the very beginning it was obvious that this method is simpler for the observers to perform, and that the accuracy of the algorithm would be comparable to the criteria we had set. Repeatability was the main issue and the reason for our modification of the algorithm.

During the process we also noted that the background image on the screen, over which the patches are superimposed, plays an important role in the determination of adapted white point. In most studies of this type the background is set to be a uniform gray background. We did use this type of background as our null case in deciding a suitable background. However, we made a choice not to use a uniform gray background for the experiment in order to avoid the possibility of color matching by the observers. Color matching is when the observers simply try to choose the patch which matches closest in color to the gray background, instead of purely adapting to the viewing environment. In the following section we will detail the algorithm and its evolution, and discuss some of the studies we conducted in order to select the most suitable background. We will then describe in detail the experiment setup and procedure, and then present the experimental results of the final (current) version of the algorithm. We will then discuss the accuracy and reliability of the algorithm based on these results. We will also briefly discuss further possible uses for the algorithm.

II. Methodology

The method of constant stimuli is the basis for our algorithm. Early on, we decided that achromatic colors provide a simple, reliable estimate for chromatic adaptation of observers in various viewing conditions. The basics of the method are very similar to the experimental method presented by Gorzynski in his thesis [6].

Square patches of 2X2cm are displayed on the CRT screen, overlaying a background image (a random pattern) which is centered on the screen. The 2-cm dimension approximates the observer 2^{o} range of view. Since the monitor luminance is highly non-uniform at the edges of the screen, we limit the algorithm to only use the central 25cm by 22cm of the screen (when using a 20-inch screen). The observer is instructed not to fixate on any particular patch or portion of the screen. He/she is further instructed to scan all the patches and choose the one patch that is the most achromatic of all

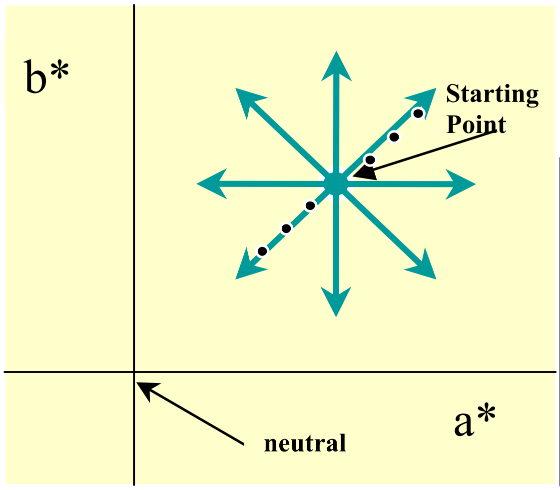


Figure 2. Schematic of how the patches are chosen. The starting point is a random selection from a sample population. Four vectors at 0,35,90,135 degrees are chosen, and 16 sample points (the black dots) on each vector are chosen as the 16 patches to be presented on each screen. The order of display for these 4 vectors (screens) is randomized. The step size between two adjacent sample points is no bigger than 2 ΔE and no less than 0.25. As the observer gets closer to the neutral point the step size is decreased.

where each set consists of four screens presented in random order. Figure 2 depicts the scheme used: The patches on each screen lie along one axis of the asterisk pattern in CIELAB (a*, b*) plane, where the lightness (L* value) is held constant during each run of the experiment. Each of the four consecutive screens comprises one vector (axis) in the asterisk pattern of Figure 2. Each screen holds 16 patches (samples along one vector) plus the central point. The mean of the observer's patch selections from each four consecutive screens becomes the new center point about which the next four screen-full of patches are constructed. The process is repeated until the standard deviation of the four observer selections is less than 0.25 in CIELAB space. The sampling distance between adjacent patches along an axis is between 2 and 0.25 units. When the experiment first starts, and we are far from the actual neutral point, the step size used to traverse along these vectors is 2. After a new central point is computed, the step size is modified to be the same as the standard deviation of the observer's four selections from the previous round. This will never get below 0.25 however, since the termination criterion of the algorithm is set at 0.25 standard deviation. The basis for this choice is that human visual system cannot distinguish between neutral patches closer than 0.25 ΔE to each other. Since the observer is instructed to choose the most achromatic-looking patch from each screen-full of patches, the iterative process does converge towards the observer's perceived neutral point in (a*, b*) space – this point is the chromatic adaptation point we are looking for.

The advantage of the MIND algorithm is its ability to hold the observers accountable for consistency. If the observer consistently chooses bluish gray patches, then the algorithm will converge around a bluish gray and that will be the AWP. But if the observer chooses a bluish gray patch from one screen, a reddish gray patch from another screen, a yellowish gray from the third screen, and so forth, the algorithm will in effect throw him further back and away from the real neutral point. Moreover, the step size (or distance) between adjacent patches will increase since the standard deviation of the observer choices will be high. In effect MIND forces the observer to stay consistent in his/her selections. The more consistent the observer is, the faster the algorithm will converge and meet the termination threshold of 0.25 standard deviation between choices. Figure 3 is a schematic of this process.

During the development of the algorithm the following areas were studied, modified, and improved upon:
a- The measurement process,
b- The background image used,
c- The selection of the new central point,
d- The threshold criterion used.

In the rest of this section we will provide a short description of the evolution of each of the above areas.

II.a Measurement Process

We expect the adapted white point to be a function of the available measurements. These measurements are the CRT white point, the ambient white point, and the colorimetric values for the background image. Our first task was to develop a measurement scheme that would measure correctly and compensate for monitor non-uniformity, and the fluorescent lighting instability.

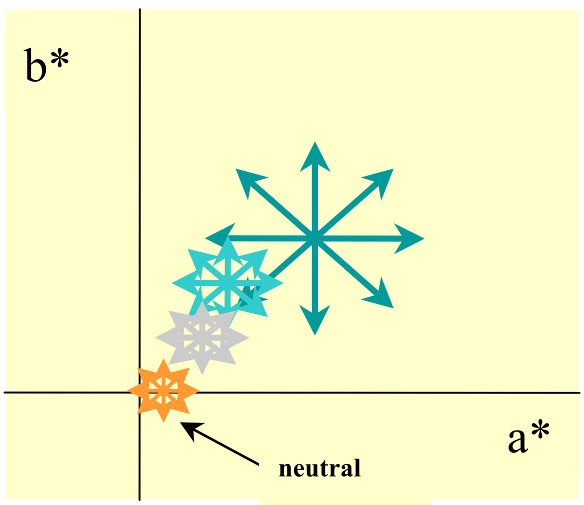


Figure 3. The process of converging to the neutral point. Every four screen-full of patches presented to the observer is one of the asterisk patterns above. The observer's four selections, one from each four screens, are averaged together and yield the center point for the next four screen-full of patches. This process will iterate and take us towards the observer's neutral point in the space. We stop when the standard deviation of the observer's four selections is less than 0.25 rms.

We first ran a study to ascertain the spatial uniformity of our CRT monitor. The monitor was a professional graphics
Sony monitor. Most of the non-uniformity of the monitor was in its luminance channel. When measuring a white patch at 16 locations on the screen (over a 4 by 4 grid), the right edges of the CRT screen dipped significantly (more than 5 cd/m^2) in luminance. Unfortunately the built-in uniformity function of the monitor over-compensated for this luminance non-uniformity, and had to be turned off. The chroma channels were more uniform, although the top left corner of the screen registered a significantly higher chroma for the same white patch as any other spot on the screen.

We also noticed that there is a high temporal non-uniformity in the luminance channel of the CRT, although the chroma channel was stable. There seemed to be more of a fluctuation if the monitor was left on for long periods of time.

To compensate for these non-uniformities, we took the following steps:

1- Used only the middle portion of the screen. We used the middle 25cm out of available 40cm in width, and the middle 22cm out of available 29cm in height. The unused borders are black during the experiment.
2- Turned off the monitor at the end of each day of experiments to preserve it.
3- Calibrated the monitor twice each day, after having warmed it up for at least 3 hours. After it was calibrated as close to the viewing condition's specifications as possible, we would then characterize the monitor.

Note that the calibration/characterization of the monitor was always performed in a totally dark room, where the only source of illumination was the CRT itself.

Characterization of the monitor used the same principles as in the treatise by Berns et al. on CRT metrology and characterization[7]. We measure the Y,x,y values of red, green, blue patches at maximum digital counts of (255,0,0), (0,255,0), and (0,0,255) respectively. Computing CIE tristimulus values XYZ from these measurements gives us the transform matrix from RGB to XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{r,\max} & X_{g,\max} & X_{b,\max} \\ Y_{r,\max} & Y_{g,\max} & Y_{b,\max} \\ Z_{r,\max} & Z_{g,\max} & Z_{b,\max} \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

We also measure the Y,x,y and hence the XYZ tristimulus values for a ramp setup of neutral patches including white (max digital count of 255,255,255) , where the digital counts for the 3 channels are equal for each neutral patch. Using the inverse of above equation and the XYZ values of the neutral ramp, we can compute the "normalized" phosphor tristimulus values, R,G,B $_{\text{normalized}}$ which are related to the digital counts through gain (k_1), offset (k_2), and gamma (γ).

$$R_{\text{normalized}} = \begin{cases} \left(k_{1,r}\left(\frac{d_r}{2^N - 1}\right) + k_{2,r} \right)^{\gamma_r} & ; \left(k_{1,r}\left(\frac{d_r}{2^N - 1}\right) + k_{2,r} \right) \geq 0 \\ 0 & ; \left(k_{1,r}\left(\frac{d_r}{2^N - 1}\right) + k_{2,r} \right) < 0 \end{cases}$$

and same for G $_{\text{normalized}}$ and B $_{\text{normalized}}$. Using nonlinear optimization techniques, we can get estimates for the three gamma, gain, and offset values.

Once we characterize the monitor, we keep the results in a characterization file which is used to convert our CIELAB Lab values to monitor RGB for every color we put up on the screen. This file is hence used to display the background image and the patches correctly, based on the current monitor characterization.

Measurement of the ambient white point was relatively easy. Since we could not turn off the CRT in order to measure the ambient, we instead displayed a solid black screen on the monitor. There was hardly any light emanating from the screen in this mode. After letting the fluorescent lights warm up at their full intensity for at least half an hour, we would measure a halon white disk on the faceplate of the monitor, trying to adjust the illuminance level per our specs. We measured the veiling glare on the black faceplate as well, trying to keep it low by switching off banks of lights most directly affecting the glare problem. Once we were happy with the CRT and ambient white point measurements, we were ready to run the experiment.

After one observer session was over, we would redisplay his/her results on the monitor, under the exact same viewing conditions that the experiment was run in, and measure the tristimulus values (Y,x,y) for each resultant patch.

We would, at the same session, also measure the background image while ambient illumination is on. In effect, the background chroma measurement gives us a measure for the adapting field. Since the background covers the entire used portion of the screen, we decided to make 9 measurements of the background image, overlaying a grid of 3 by 3 over the used portion of the screen. This way we can discount the spatial non-uniformity of the CRT, and by measuring the background at the same session when the experiment is run, we also discount the temporal non-uniformity of the CRT. One question that we looked into was the contribution of each of these 9 grid areas to the overall background measurement. The used portion of the CRT subtends almost 25 degrees of observer's range of view. So do the areas closer to the center of the screen influence the vision more than those at the top or bottom corners? We used three different models to look at the result of these nine measurements:

1- Averaging scheme: simply take average of the nine measurements and call that the background measurement.
2- Use a weighting pattern, where the measurements take on weightings of :

1	2	1
2	3	2
1	2	1

3- Use the cosine law based on photographic lens range of view, where each area's contribution is proportionate to cosine to power of four of the angle between the observer and the screen area he/she focuses on.

The differences between the above three procedures were so small (within 0.1 ΔE of each other) that we decided they all yield the same results, and hence opted for method 1, the simple averaging scheme.

The most important thing to note in all these measurements is that the background measurement was done under the exact same viewing conditions that the observer was subjected to. In other words, this was a measure of the adapting field under our mixed mode illumination conditions. This is crucial in understanding the results that we obtained and will present in the next section.

II.b Background Image

At the onset of the experiment and its design we were sure of only one thing regarding the background image: We did not want to use a uniform gray background in order to avoid any color matching by observers. We wanted to simulate real images, where there are many different colors and patterns, but the gray-world rule holds and the average CIELAB value of the entire image does resemble an achromatic patch. We hence tried random patterns of chromatic and achromatic patches. In all cases the average value of the image was L* of 60, and zero (a*, b*).

It is well known that the background plays an important role in the adaptation. Our initial trial runs proved this by yielding vastly different results based on the background chosen. We hence decided to make a more rigorous effort of addressing the issue of the background and its factors that affect the outcome. The factors that we changed in our background study were the chroma, the range of lightness, and size of the dots (where a dot is the unit sample of the background and is specified in terms of number of pixels comprising it). Figure 4 shows some of the backgrounds tested and the adaptation results.

Background	Pixel Size	Chroma Range	Lightness Range	% Completely Adapted	AWP CCT
Solid Gray	0	0	L*=60	98.5	6855
4X4(A)	4	0	30-90	98.4	6865
4X4(35)	4	0-35	L*=60	94.1	7177
2X2(35)	2	0-35	L*=60	94.5	7144
2X2(30s)	2	0-30	30-90	94.4	7157

Figure 4. Table of the characteristics of our four finalist background patterns against the control case of the solid gray background. The adaptation results and the correlated color temperatures of the adaptation point are also shown.

For more detailed information, please refer to a paper that we have published on this subject [8].

We made the following important observations and conclusions from our background study:

1- Chromatic random backgrounds are not the same as pictorial images. The visual system uses different mechanisms when the visual clues of an image are not present, and hence the adaptation is very poor using such backgrounds, as seen in the results from the last three backgrounds above.

2- Chromatic random backgrounds (with mean [a*, b*] of zero) are not the same as a uniform gray background. Individual pixels (dots) do not get mixed additively, and hence the results are drastically different from the solid gray background case. Hence we should not assume that the adapting field integrates to gray.

3- An achromatic random background with relatively large dot size (four by four pixels) can perform almost as well as the solid gray background, as seen in the second row of the table above.

After we concluded that the achromatic random backgrounds could work for our adaptation problem, we focused on improving the achromatic background by making it more pleasing and less noisy for the observer. We high-pass filtered the original 4X4 achromatic background of the above table, and added some symmetrical features to it. One of such constructed patterns resulted in better adaptation ratios. We believe that the reason is the non-interfering nature of the image, where the observer's visual system is not distracted or tired out by the background. A picture of our final background used is included as Figure 5.

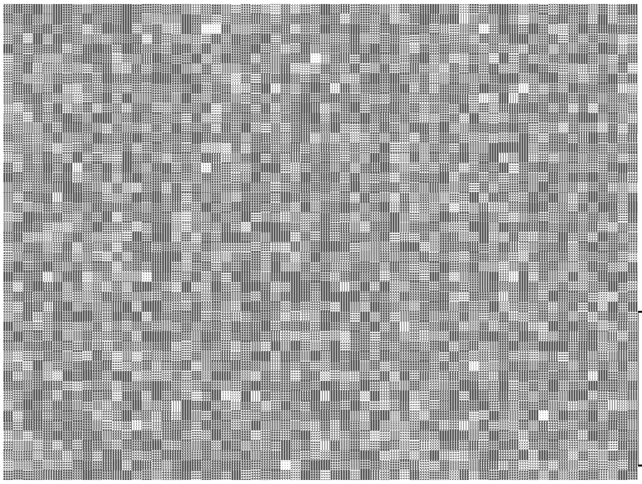


Figure 5. The achromatic random pattern used as the final background pattern for MIND. The average CIELAB L* of the pattern is 60. Average (a*,b*) values are very close to (0,0).

II.c Sampling of Starting Points

Each session of MIND starts from a random point in the sampled population of the color plane. In theory, we can have the entire color plane as our population space, and we would still converge to the adapted white point. However, in order to speed up the selection process, we choose the starting point from a sampled area of the (u*,v*) space centered on the D-illuminant locus. In cases where the CRT temperature was set at D65 and the ambient lighting was about D40 or D50, we sampled 100 points in an ellipsoid with its long axis the straight line connecting the D45 point to the D75 point on the locus. The short axis was half as long. The reason for using this region was that we wanted to start at a reasonably close point to the actual neutral axis and having the ambient and CRT temperatures, we knew that the adaptation point would be roughly in the same region of the plane. Use of an ellipsoid was justified by trying to trace the shape of the locus in (u*,v*) plane. We then converted the 100 sample points to CIELAB space of the CRT, and randomly picked one as the starting central point of the algorithm. Since each observer repeated the experiment at least five times, each time a new random starting point would be selected.

The algorithm starts with a random point from the (u*,v*) ellipsoid as the central point of the four vectors emanating from it. The four vectors form 0,35,90, and 135-degree angles with the center point horizon. We choose 16 points on each vector. This is referred to as vector sampling technique in Gorzynski's method. In the beginning of the experiment every two points are separated by 2 units (in CIELAB plane). Each point such found represents a color patch. Each screen presented to the observer holds 17 patches: 16 sample points along one vector, plus the central point itself. Once the observer selects the most achromatic-looking patch from 4 consecutive screens, MIND finds the mean of these four selections and comes up with a new central point around which the next four screens of patches are constructed. The points along the vectors are now separated by σ, where σ is the standard deviation of the observer selections from the previous four screens. If this standard deviation falls below 0.25 units the algorithm terminates, since we assume that 0.25 CIELAB ΔE units is the least color difference that the average visual system can detect.

Note that the algorithm thus constructed is capable of adjusting for observer inconsistencies. If the

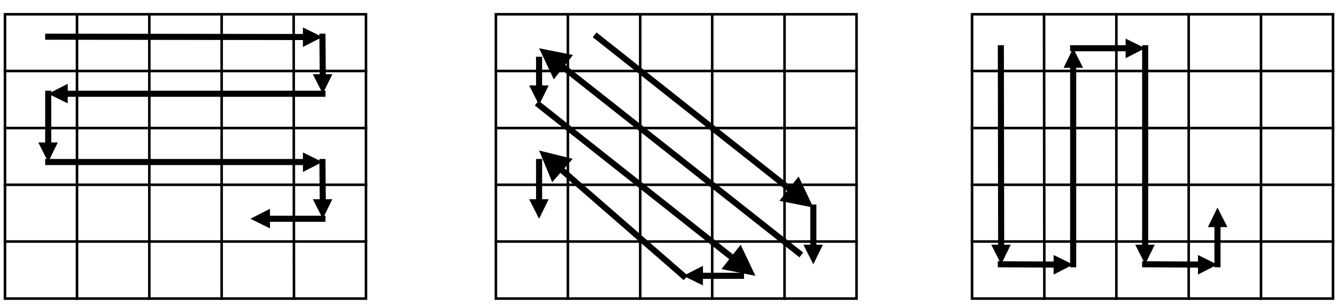


Figure 6. Three of the twelve grid patterns used to place the color patches on the screens presented to the observers. From left to right these are examples of horizontal, diagonal, and vertical patterns respectively. Four of each type of pattern was used, with different directions and/or starting grids. Each screen's grid pattern for the patch display was chosen in random order.

four consecutive selections are haphazard, it is well possible that the next central point will be further from the CIELAB neutral point than the previous central point. In which case the step size between color patches sampled on each vector will also increase. This fact in itself is a main reason why the algorithm is low in noise and very robust. Another way to think of this is that MIND forces the observer to stay consistent, and hence the intra-observer variation becomes very small.

We also studied the positioning of the 17 patches on each screen presented to the observer. Our basic premise was that these patches would be superimposed on the background image in a grid-like pattern. If the patch squares are 2cm by 2cm (representing a 2^o observer range of view), the space between every two patches is to be about 2cm as well. With these restrictions, and the size of the background image positioned in the middle of the screen, we can have a 5X5-grid pattern superimposed on the background to use for our patch placement. At first we had the 17 patches randomly distributed in this grid. This again posed noise problems for us: If the patches are similar in chroma and placed very far from each other, the visual system cannot consistently sort the patches and choose the one with the least amount of chroma. We noticed that the observers had a strong blue tendency when asked to select the most achromatic patch from randomly placed patches on the screen. Their selections were also noisy between their five repeats, and observer consistency was still an issue. To try and eliminate these problems, we came up with 12 grid patterns of patch placement where the 17 patches are placed in the same order that they are sampled along each vector. So for example the patches go from a more saturated pink to a neutral looking color, to greens and finally a more saturated green. The grid patterns used are horizontal placements, vertical placements, and diagonal placements. Figure 6 shows examples of these patterns. The observers do not know that the patches are presented in any particular order, and they do not know anything about the grid patterns used. We utilize 12 grid patterns, where each screen assumes one of the patterns in random fashion. Using these patterns the patches similar in chroma will be placed close to each other, and hence distinguishing between them becomes an easier task for the visual system. This scheme has been extremely effective in reducing the noise between the five repeats of each observer.

II.d Threshold

We tried a few different termination criteria for the algorithm. In the beginning of the evolution of MIND the observer would decide when to terminate based on whether he/she could tell apart the patches on the screen, or whether they all looked equally achromatic. As soon as the patches all looked the same, he/she would terminate the process. This technique proved to be too subjective and noisy. We then tried a termination scheme similar to Gorzynski's, where we stop the iteration when the successive means are within two standard deviations (2σ) of each other or when the standard deviation in both a* and b* directions falls below 0.25. This scheme worked fine most of the time, but depending on the starting point and observer selections, it would either stop too soon (due to the means being within the 2σ threshold), or not terminate even after the samples all looked very similar. We tried refining the tolerances to avoid these anomalies by discarding the threshold on the means, and terminating when the average standard deviation of a* and b* directions falls below 0.25 (note that we use the average σ, not the standard deviation in both directions). This modification did improve the consistency of termination points, although we still have rare cases where the algorithm keeps iterating long after the patches stop being different to the eye. The algorithm also asks the observer to confirm the final achromatic choice before the process terminates.

Overall, we have modified the four areas discussed above expressly to reduce noise in the experiment, and as the results will show in the next section, we have been successful in achieving this goal.

III. Results

The experiment was conducted in a specially constructed room, where the walls are painted a non-reflective 18% gray (that is, their luminance is L* 18 out of a white of L*100). There are no windows, and special fluorescent lamps are mounted to simulate D65, D50, D40, or any other illuminants at various intensity levels. All the furniture is painted gray and there are no highly chromatic objects in the room. The observers are also required to wear a black robe over their clothes in order to minimize the reflection of their clothes on the CRT screen. We refer to this room as the gray room.

The monitor used is a Sony GDM 2000TC, which was a state of the art monitor back in 1995 when we purchased it. As mentioned above, despite its high quality, it is still non-uniform in both temporal and spatial domains. It is also non-linear, meaning that the chroma of a gray ramp does not measure a constant value for all steps in the ramp and there is high variation in the (x,y) measurements of the ramp. We allowed a 2 to 3 hour warm up period for the CRT before calibrating and characterizing the monitor. We characterized the monitor twice daily, once before the morning sessions at 9 am, and once before the afternoon sessions at 12 noon. If there were any night sessions scheduled, we would characterize it a third time at around 5pm. During the entire day the monitor would never be turned off, although while not in use we would display a dark screen encompassing the entire CRT screen and hence "preserving" the guns.

The ambient lighting, provided by the special fluorescent lights, also was temporally non-uniform, although the amount of variation was not very much. The most important factor in the stability and constancy of the ambient lighting was the warm-up period. The lights needed at least half an hour to warm up to stable levels of intensity.

We had four viewing conditions set up. Ideally we should have administered these four experiments in a random fashion to avoid any correlation. However, since each setup required installation of a different set of fluorescent lamps, hence resulting in much more setup time than our schedule allowed, we decided to set the gray room up for each of the four setups at the start of the week, and run that particular setup for the entire week. These four setups were:

1. Dark Room: no ambient lighting, monitor set to a sRGB monitor, i.e., D65 monitor at luminance of 80cd/m^2. This setup would be our control case.
2. SRGB Room: set up everything as in the sRGB specifications: ambient lighting is D50 at luminance of 20.4cd/m^2 (equivalent to illuminance of 64 lux) on the CRT faceplate, monitor set to D65 temperature at luminance of 80 cd/m^2, with veiling glare of less than 1%.
3. Office 1: a more typical office setting, where the monitor is same as case 2 above, but the ambient lighting is D40-ish at luminance of 270 cd/m^2 on the desktop (equivalent to illuminance of 850 lux). This is equivalent to ambient luminance of about 85cd/m^2 on the CRT faceplate. The veiling glare is about 8 or 9%.
4. Office 2: the same setup as Office 1 above, except that the CRT is set to D93 temperature.

We had eleven naïve, color-normal observers who undertook the entire 4 weeks (4 viewing conditions) of experiments. We selected observers who had some art or coloring-related background such as painting, weaving, or photography. This was defined very loosely, and the expertise and background of our observers varied greatly. The main purpose for this requirement was due to the fact that in our initial studies we had noticed that the "artists" who knew about colors to some degree, tended to be more consistent in their terminology and hence in their patch selections. We told the observers the general purpose of the experiment: "that the purpose is to see how the visual system changes under different viewing conditions", but that was the extent of what they knew about the experiment. We ended up with 6 female and 4 male observers. Their ages varied from 25 years of age to 63, with the average age of 43, and the median age of 41.

Each week, each observer would be given three sessions of experiments, in random order. Each of these 3 sessions would use patches at a given, fixed lightness level. The lightness levels used for the patches were L* 50, L* 65, and L* 80. We corrected for the Helmolz-Kohlrausch effect [9] so that the perceived hue and lightness of all patches remained constant on each screen. Each observer would repeat the experiment five times at any given session, each time starting at a random point selected from the population of the central starting points (within the ellipsoid around the D locus). So for each lightness level we gathered: 11 observers X 5 repeats = 55 results, and for each viewing condition we had: 3 lightness levels X 55 results = 165 results. Using a spectraradiometer (PR650), we measured the lightness and chroma of each result patch. Hence we had 165 (Y,x,y) values representing what the observers perceived to be achromatic under each viewing condition. Averaging these results over all observers and over all three lightness levels gives us the adapted white point for that particular viewing condition.

Immediately after a session was over, we measured the background image using the PR650. As explained above we display the background image on the screen, divide the central portion of the screen (where the experiment is run) into 9 regions and measure all 9 regions and then average these into what we call our background measurement. The 9-way measurement was due to our studies of the spatial monitor non-uniformity. Since the background measurements were done immediately after a session with an observer was over, the monitor's temporal non-uniformity was not an issue. Therefore we had:
3 lightness levels X 11 observers = 33 sessions each week, which yielded 33 background measurements for each viewing condition, 33 monitor white point measurements, 33 ambient white point measurements, and 33 flare measurements for each viewing condition. By comparing the adapted white point results versus these measurements we were able to draw conclusions on the behavior of adaptation under various viewing conditions. These results will be introduced in the next section.

In Appendix A, we itemize the process of a typical experiment session in order to present the chronology of the events and the measurements.

We can then plot the distance of the adapted white point from the background, monitor white point, and the ambient white point for each observer, each lightness level, and each viewing condition. We can also further analyze the results by taking average of all measurements and come up with the "average" observer results. These results are presented in the following discussion. In this paper we will only focus on results which support our claim that MIND is a low-noise, low-error, robust algorithm for psychophysical applications of same kind.

Figure 7 shows the results of the adapted white point (AWP) for the "average" observer with regards to the background, monitor, and ambient measurements. We have presented the results in delta coordinates, where delta means the difference of adapted white point from the other measurements. This in effect shows the distance between adapted white point and each of the other measurements. The closer to zero this delta value is, the more similar are the adapted white point and the other measurement. For example in the dark viewing condition, our control case, we expect the observers to completely adapt to the monitor white point because the CRT is the only source of illumination. We can extend this expectation and say that since the background is achromatic, and hence has the same chroma as the monitor white point, the observers should also completely adapt to the background measurement under the dark viewing conditions. In other words, we expect the AWP to be same as the monitor white point which is theoretically same as the achromatic background (chroma) measurement:

Monitor WP - AWP = Background - AWP = 0

In CIELAB plane (a*,b*), we can say that we expect the (rms.) distance from AWP to monitor white point and to the background to be zero. First row of Figure 4 shows the results in this delta coordinate space. The differences in a* direction are a lot smaller than in the b* direction, but nonetheless they are not zero. The difference of AWP from the background measurement under the control case of dark viewing condition is then our measure of accuracy. In other words, the 95% confidence interval for the mean of the (Background – AWP) is:

$$E_{(Back-AWP)} \pm 2\sigma_{(Back-AWP)} \Rightarrow \quad \sqrt{(0.1)^2+(-0.97)^2} \ \pm\ 2\times\sqrt{(0.1)^2+(0.18)^2} = 0.95 \pm 0.4$$

Hence our accuracy is between 1.35 and 0.55 CIELAB ΔE. This is a small error for a psychophysical experiment, but it is not zero. The sources of error in the experiment can be contributed to monitor instability, non-uniformity, and non-linearity. Also colorimetric errors due to both the measurement instrument (PR650) and our calculations (Yxy to CIELAB) can cause errors.

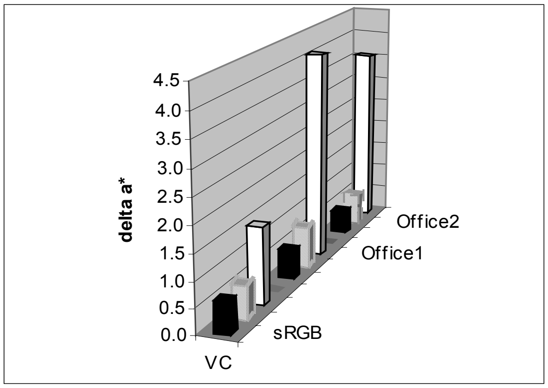
The low standard deviation (and variance) also shows the repeatability of the algorithm. The standard deviation is consistently low for all four viewing conditions tested. This indicates low noise level present in the algorithm. We believe that the high accuracy and low noise level are due to the many modifications we made to the algorithm, specifically to the four areas we discussed above, namely the background selection, the measurement process, the starting point of the experiment and the selection process of the next center point, and the threshold used to terminate the process.

Figure 7 also shows that in the sRGB viewing condition, if we adjust the difference between AWP and the monitor white point for the level of accuracy of the algorithm, the difference is practically nil. Hence we claim that for the sRGB viewing condition we can use the monitor white point, i.e., the white point of D65 illuminant as our adapted white point. This is a powerful result. It says that if we are in the sRGB viewing condition, we do not need to make any measurements in order to apply the CIECAM97 appearance model to the data, and that we can directly use D65 white point as the adapted white point argument into the model.

		Backgroud - AWP		***Monitor -AWP***		***Ambient - AWP***	
Viewing Condition		da*	db*	da*	db*	da*	db*
Dark	Mean	.10	-.97	.02	-.16		
	Std. Error	.10	.18	.09	.17		
sRGB	Mean	-.06	-1.21	.06	-.67	1.51	7.89
	Std. Error	.08	.18	.08	.19	.07	.18
Office1	Mean	.09	-1.66	.69	-4.63	4.11	39.73
	Std. Error	.08	.22	.08	.22	.08	.22
Office2	Mean	.24	-1.58	.50	-5.49	3.55	53.44
	Std. Error	.06	.17	.06	.17	.06	.18

Figure 7. Table of results from all four viewing conditions. The means and standard deviations are calculated over all three L* levels, and over all 11 observers.

Figure 8 is a plot of the differences between AWP and background, monitor, and ambient white points for the sRGB and the two office viewing conditions. It shows that the differences in b* direction are more significant than those in a* direction (range of values in b* direction is 60 units versus the a* direction which has a range of 4). It also shows that the difference between the observer results (the AWP measurements) and the background measurement (in black) is considerably less than that between the AWP and the monitor (in orange) or the ambient (in blue). This is a powerful result since we can base the adapted white point estimation on the background measurements.



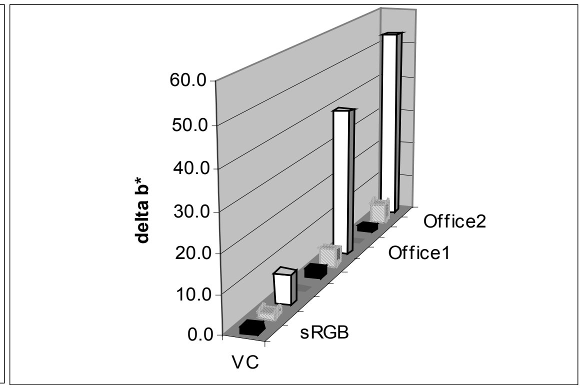


Figure 8. Plots of the difference between AWP results and the other VC measurements in CIELAB (a*,b*) plane. Black bars are difference from background measurement. Cross-hatched bars are difference from monitor WP, and white bars are difference from ambient WP. In all three viewing conditions the difference from background is much smaller. Note the scale of delta a* plot versus that of the delta b*. Most of the differences lie in the b* direction.

Another important result that confirms the robust-ness of the algorithm is that all eleven observers hold the same trends across the board. This is depicted in Figure 9, where the distance (ΔE) of individual observer results from the three measurements (background, monitor, and ambient white points) is plotted. Figure 9 shows results for the three lightness levels in each of the three viewing conditions with ambient illumination (sRGB, Office1, and Office2). Although the observers were from varied backgrounds, ages, and color expertise, they all hold the trend of having results furthest from ambient white point, and closest to the background measurements. Even though we do have outliers in terms of their individual results and their distance from the average AWP, Figure 9 proves that they all hold the trend displayed in Figure 8. This is valuable confirmation on the fact that the algorithm is a solid measure of how any observer responds to the viewing conditions regardless of how accurate his/her visual system is.

IV. Further Usage

The MIND algorithm is proven to be accurate, repeatable, and robust. We have devised it for the purpose of white point determination. In a similar vein, we can apply it to find the entire neutral axis of the CRT under a given viewing environment. Instead of applying it to 3 distinct lightness levels, (the way it is applied now), we can apply it to many lightness levels, hence giving us more points along the neutral axis of the CRT gamut. Being able to pinpoint the neutral axis without any measurements can have potential uses in tone reproduction and monitor characterization.

A major problem with monitor characterization is that there are always some measurements involved. Section II above, described the way we characterize our monitor. The measurement of a series of gray patches of varying lightness is a typical method of characterization. In addition we need to measure red, green, blue patches, and we need to know the white point of the monitor. Armed with all these info we can then apply a regression method to determine the gamma, gain, and offset of each of the red, green, blue guns. Using MIND we can in effect backward engineer this process. MIND has two shortcomings here: it cannot give us the CRT white point, and it is not equipped to give us information on the red, green, blue patches. However, there are other visual tools that can provide us with this info. For example, the maximum value of the red, green, blue guns of the CRT are usually fixed with respect to the type of CRT in use. We can also use a simple measuring tool to get the CRT white point. The advantage of using MIND for monitor characterization over the standard methods is that MIND will characterize the monitor in the viewing environment, hence the ambient lighting and the flare and any other contributing factor is taken into account. In effect, the user can run himself through the MIND experiment, and once he has his results, he then has characterized his viewing environment. The current monitor characterization techniques are incapable of taking the viewing conditions into account as well.

Of course there are still problems with the above-proposed usage for MIND. Issues such as how exactly to measure the monitor white point and how many lightness levels are needed to run through the MIND experiment have not been resolved yet. At the minimum, if we know that the CRT is shipped with sRGB settings, and if we get the neutral axis of the monitor via MIND, we can then compute the offset or shift from sRGB to the current setting of the monitor, and use this result in adjusting the tone maps. MIND is just another weapon in our arsenal of visual and measurement tools, and its power lies in its ability to accurately predict or determine the adapted white point, or the neutral axis of the RGB cube for the entire viewing environment and not just for the CRT.

V. Conclusions

The method of multi-stimuli interactive neutral-point determination (or MIND) has been designed based on work done by Gorzynski using method of constant stimuli. We have made numerous adjustments and modifications to the method in order to make it robust, noise-free, accurate, and repeatable. Based on the results we have presented, we believe that we have achieved all these objectives. The accuracy of the algorithm in particular far surpasses any other psychophysical experiment of this nature to date.

Furthermore, having applied the algorithm to determine the adapted white point (AWP) under various viewing conditions, we have confirmed the fact that the MIND algorithm enables observers to be accurate and consistent. Hence the intra-observer noise has been greatly reduced and the repeatability of the algorithm is established. The inter-observer behavior is now consistent across observers and follows the same pattern for all observers. The AWP experiment has shown us the error of the MIND algorithm to be about $1.5\Delta E$ at its worst. This is unprecedented in psychophysical experiments of this nature.

Recently Katoh et al [10] have also studied the problem of adaptation under mixed mode illumination. His approach has been slightly different than ours however. While we have addressed the problem of CRT adaptation where the observer is continuously or near-continuously looking at the CRT screen, Katoh has studied the problem of hardcopy and softcopy adaptation. In his method, the observer shifts his focus from the CRT monitor to the print and vice versa during the course of the experiment. Hence the ambient lighting becomes more influential in such a simultaneous viewing situation. Although both cases fall under “adaptation under mixed mode illumination”, the viewings are done in very different ways. Both scenarios are valid although their context is very different. We have assumed a successive viewing model, where the user first looks at the monitor, and then moves over to another area (perhaps under the same illumination, perhaps not) to view his print of the same image on the screen. Katoh's model is more of a simultaneous viewing case, where the user looks at both the print and the screen image at the same time. They each have their merits, and we will be confirming Katoh's results under his assumptions in future.

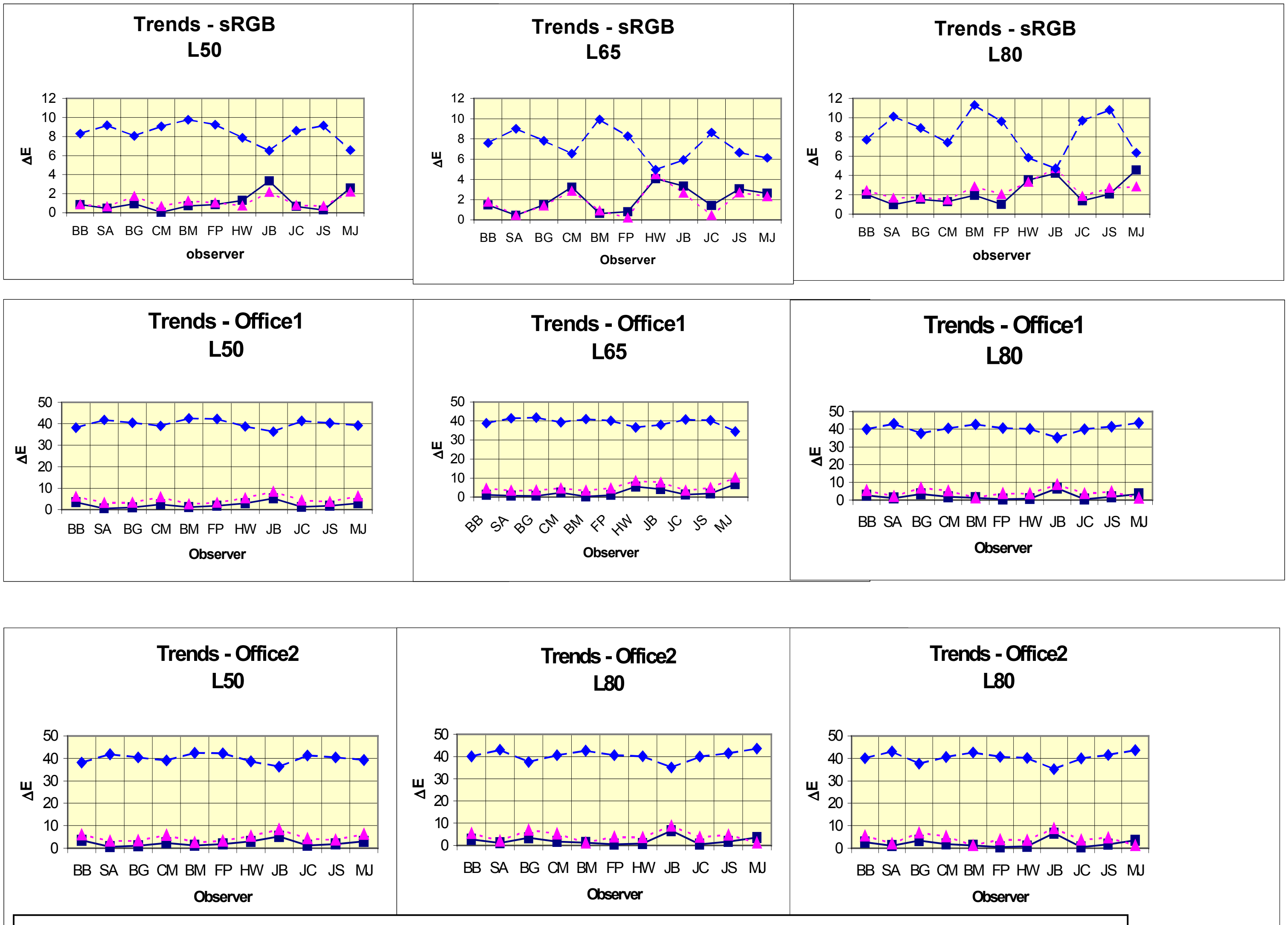


Figure 9a. Plots of distances between each observer AWP and the background, monitor, and ambient white points. Top row is for sRGB, middle row is for Office1, and bottom row is for Office2 viewing conditions. Each row, from left to right depicts the L*50, L*65, and L* 80 results. In each plot the horizontal axis represents the 11 observers, and the vertical axis is the distance from the observer adapted white point result. The large dashed line (with diamonds) is distance between ambient white point and the AWP, the small dashed line (with circles) is the distance between monitor white point and the AWP, and the black line is the distance between background chroma and the AWP. Due to the scale of the plots, the background and monitor differences seem to lie almost on top of each other. In all cases and for all observers the trend shown in Figure 8 holds. This proves the consistency and robustness of the MIND algorithm.

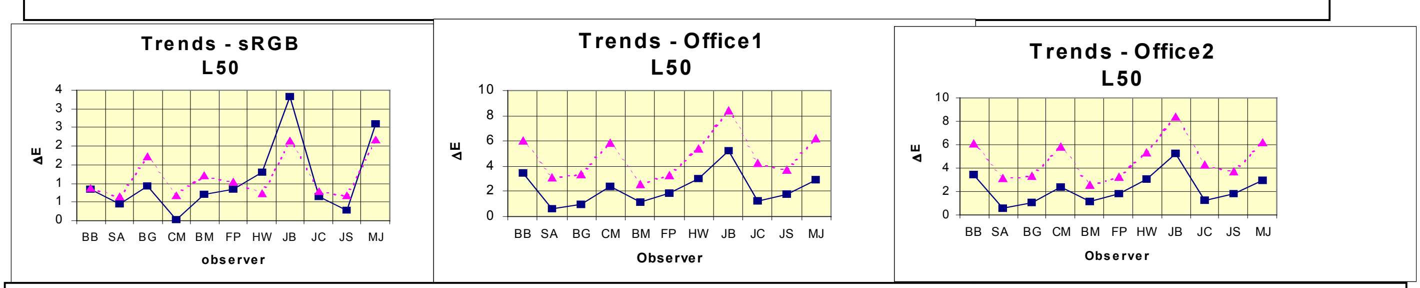


Figure 9b. Same plots as in Fig. 9a), but shown only for background and monitor differences, and only for one of the three lightness levels, L*50. Due to the scale change, we can now see that the background (black solid line) and the monitor (dashed line) differences are significant. Still, almost all observers follow the same trends.

VI. References

1. M. R. Luo and R. W. Hunt, *The structures of the CIE 1997 colour appearance model (CIECAM97s)*, Color Res. Appl., **23**, 138-146 (1998).
2. N. Moroney, *Model based color tolerances*, Proceedings of the 7th IS&T/SID Color Imaging Conference, 93-96 (1999).
3. D. H. Brainard and K. Ishigami, *Factors influencing the appearance of CRT colors*, proceedings of the 3rd IS&T/SID Color Imaging Conference, 62-66 (1995).
4. M. D. Fairchild, *Chromatic adaptation to imaging displays*, TAGA Proceedings, **2**, 802-823 (1992).
5. T. Seim and K. Richter, *Changes in Cone Excitations as Basis for Analyzing New Color Scaling and Threshold Data*, Proceedings CIE 21st Session, (1987).
6. M. E. Gorzynski, *Achromatic perception in color image displays*, M.S. Thesis, Rochester Institute of Technology, 1992.
7. R. Berns, M. Gorzynski, and R. Motta, *CRT Metrology and Colorimetric Characterization Techniques*, Technical report. Munsell Color Science Laboratory, Rochester Institute of technology, 1991.
8. P. Oskoui and E. Pirrotta, *Influence of Background Characteristics on Adapted White Points of CRTs*, Proceedings of the 6th IS&T/SID Color Imaging Conference, 22-26 (1998).
9. M. D. Fairchild and E. Pirrotta*, Predicting the lightness of chromatic object colors using CIELAB*, Color Research and Application, **16**, 385-393 (1991).
10. N. Katoh, K. Nakbayashi, M. Ito, and S. Ohno, *Effect of ambient light on the color appearance of softcopy images: Mixed chromatic adaptation for self-luminous displays*, Journal of Electronic Imaging, 794-806 (1998).

VII. Appendix A
Step by Step Process of the AWP Experiment Using MIND

1- Turn on the CRT and let it warm up for a couple of hours.
2- Calibrate the CRT to sRGB specs.
3- Characterize the CRT. The characterization file will be used in mapping correct RGB values to the CRT during the experiment.
4- Turn on the ambient lights to full intensity. Let it warm up for at least half an hour.
5- Calibrate the ambient lights to the viewing condition specs. Since the CRT needs to be turned off while making the adjustments, and we do not of course want to turn the CRT off, we instead either cover the CRT screen with a thick black cloth, or display a totally black screen-sized image on the CRT screen. Use a halon disc to measure the ambient white point on the faceplate and the desktop.
6- Having the CRT screen covered or blacked out, we also measure the veiling glare on the CRT faceplate.
7- The viewing conditions are now ready for the experiment to start. Display the background image on the screen, and invite the observer into the room.
8- If this is the observer's first time, instruct him on the procedure, and give him one practice repeat.
9- Once the experiment starts, the algorithm is set to give the observer a 1-minute adaptation period where his/her visual system is adapting to the environment. The background image is on the screen, but no color patches. Once the minute is over, the experiment starts. Currently it is set to repeat the experiment five times, each time starting from a random spot in the starting point population (pre-calculated). If this is observer's first time taking the experiment, he/she will be given six repeats of the experiment, first one being the so-called practice run. Note that the lightness of the patches during the entire experiment is randomly selected to be one of our set of lightness values (L* of 50, 65, or 80).
10- Once all five (or six) repeats are completed and the algorithm stops, the observer exits, and the administrator displays the five resulting choices of the observer and measures them under the same viewing conditions (using the PR650). The average of these five results provides the adapted white point of that observer.
11- The background's nine regions are then measured under the same conditions, using PR650. The average value from these nine measurements provides the background measurement for each session.