2018

Defining Consistent Color Appearance for Print Images

Elena Fedorovskaya

Robert Chung

David Hunter

Pierre Urbain

Don Hutcheson

Defining Consistent Color Appearance for Print Images

Elena Fedorovskaya[1], Robert Chung[1], David Hunter[2], Pierre Urbain[2] ,
and Don Hutcheson[3]

Keywords: consistent color appearance, gray balance, tonality,
95th percentile ΔE_{00}, reference printing conditions

Abstract

Perceptually, consistency of color appearance (CCA) can be defined as the degree of visual consistency or shared visual appearance that a set of images possesses in the presence of visual differences. The CCA varies according to printing conditions and can be affected by paper substrates, inks, environments, printing devices and printing parameters, such as, color balance and color gamut. ISO 15339-2 [1] specifies the CRPC1~CRPC7 reference printing conditions for different CMYK devices to maintain (a) same hue angle of CMYRGB solids, (b) same high-light-to-midtone tonality, and (c) same neutral gray balance as defined by CGATS TR015 [2], and exhibit good CCA despite substantial differences in color gamut and substrate color. Although the assumption of consistency of color appearance among these reference printing conditions was not experimentally tested, it was based on expertise and printing practice accumulated in printing industry.

Motivated by the approach used for the CRPCs specifications, a previous study [3] demonstrated that alterations of tonality and gray balance diminished the CCA for the set of images compared to the changes in color gamut only. In the present paper, we focused on further quantifying the CCA for print images by 1) selectively changing gamut volume, tonality and gray balance in a sequence of colorimetrically measurable steps and 2) conducting psychometric evaluations to derive perceptual CCA scale. The results show that the CCA of the image set with changes due to gamut variation only, appears to be higher compared to the CCA for the types of changes that involve tonality and gray balance. Device-based 95th percentile ΔE_{00} values for adjacent datasets were shown to correlate with the consistency of color appearance in the present experiment. We also observed a discrepancy between

[1]Rochester Institute of Technology; [2]Pilot Marketing Group; [3]HutchColor

experts and novices when judging CCA. Additional experiments are needed to evaluate the effects of pictorial scene on CCA.

Introduction

The present study is carried out as a project conducted within the CIE TC 8-16 "Consistency of Colour Appearance within a Single Reproduction Medium".
Consistent color appearance (CCA) refers to a perceived attribute of a set of images, reproduced in different printing conditions, having similar visual characteristics. The CCA gives a sense of identity for the images in the set, notwithstanding the perceptible differences among them.

The ISO 15339-2 [1] describes seven datasets developed to maintain consistency of color appearance among different printing conditions. To provide initial experimental validation for this assertion, [3] demonstrated that alterations of tonality and gray balance diminished the CCA for the set of images compared to the changes in color gamut only. The goal of the present study is to further quantify the CCA for print images by selectively changing gamut volume, tonality and gray balance in a sequence of multiple colorimetrically measurable steps and conducting psychometric evaluations to derive perceptual CCA scale.

CRPC Datasets

ISO 15339-2 [1], in support of printing from digital data across multiple technologies, specified seven characterized reference printing conditions (CRPC1 ~ CRPC7). These CRPCs, having similar hue angles in primaries and two-color overprints, are suitable for calibrating CMYK output devices. Color images, separated for one of the CRPCs, can be printed in other CRPC printing conditions and preserve color appearance. Figure 1 shows the CRPCs and corresponding gamut sizes, from the smallest (CRPC1) to the largest (CRPC7).

CRPC	CRPC name
1	ColdsetNews
2	HeatsetNews
3	PremUncoated
4	SuperCal
5	PubCoated
6	PremCoated
7	Extra Large

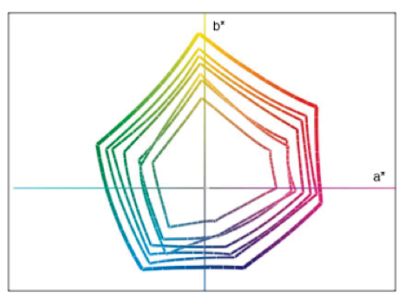


Figure 1. *Seven CRPCs and corresponding gamut sizes.*

In addition to similar CMYRGB hue angles, the CRPCs also have similar gray balance characteristics (see Figure 2) and similar highlight-to-midtone tonality (see Figure 3).

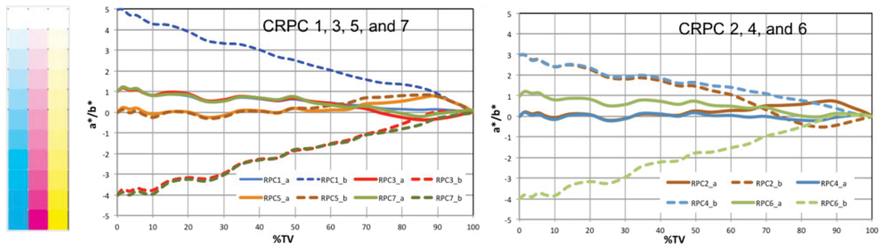


Figure 2. *Gray balance characteristics for CRPC1~ CRPC7. Gray reproduction of pre-defined CMY triplets are plotted with a* and b* of the triplets as a function of the % Tone Value (TV) (Cyan). Beginning with paper colors, all CRPCs show linear converging patterns toward the 100 CMY solid.*

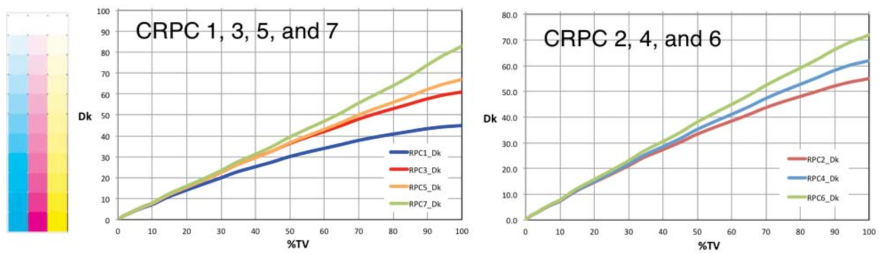
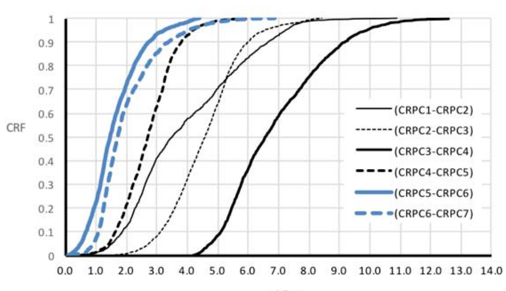


Figure 3. *Tonality characteristics for CRPC1~CRPC7. Tonality curves (TRCs) are plotted as Darkness, Dk (100 - L*) vs %TV (Cyan). If darkness is converted to neutral density, these curves are known as NPDC according to G7 terminology. All CRPCs show similar highlight-to-midtone TRCs.*

To compare adjacent CRPCs, Cumulative Relative Frequency of ΔE_{00} (CRF) and the 95th percentile ΔE_{00} are calculated and are shown in Figure 4. These measures were adopted by CGATS TR016 [4] to assess colorimetrical color difference and print conformance. As can be seen from Figure 4, CRF and the 95th percentile ΔE_{00} between adjacent CRPCs are unequal.



95th percentile ΔE_{00}	
CRPC1-CRPC2)	7.24
CRPC2-CRPC3)	6.56
CRPC3-CRPC4)	9.86
CRPC4-CRPC5)	4.18
CRPC5-CRPC6)	3.26
CRPC6-CRPC7)	4.03

Figure 4. *Cumulative Relative Frequency of ΔE_{00} (CRF) and the 95th percentile ΔE_{00} for adjacent CRPCs.*

Study Goals

In the previous study [3] observers were presented with 2 control images, and a pair of test images of the same scene and asked to select a test image that appeared to yield a higher color consistency for the triplet. The two test images for each trial were chosen from the set of three, differing in tonality, gray balance or none of these. The images were produced using datasets with the 95th percentile ΔE_{00} of 3 between the reference and sample datasets. It was shown that CCA decreases significantly when TR and GB deviate from the reference dataset (CRPC4).

The goals of the present study were as follows: 1) Test the hypothesis that CCA for a set of multiple images is the highest when based on datasets with varying gamut volumes, while having consistent tonality, gray balance and hues relative to substrate (control group), compared to the opposite condition, when the gamut volumes are kept constant, but tonality and gray balance are varied (test groups). The number of datasets for all variations was increased from 3 to 7.

2) Further examine the suitability of the 95th percentile ΔE_{00} as the measure to quantify differences in CCA. The differences between adjacent datasets in the control and test groups were decreased to the 95th percentile ΔE_{00} = 3. This value is considered to be within the tolerance for a good color reproduction - Level II, CGATS TR016 [4].

Experimental Methodology: Sample Preparation Procedure

CRPC5 reference was used as a starting point to create 7 datasets differing in chroma and gamut volume whereby adjacent datasets are 3 ΔE_{00} at 95% CRF apart (Control group). To generate the datasets for the Control group, the following procedure was used.

1. Modify CRPC5 by linear CIEXYZ scaling to get seven datasets with 3 ΔE_{00} 95th percentile gaps.
2. Restore G7 gray scale compliance using Curve4 Virtual Press Run.
3. Repeat steps 1-2 (as needed) to obtain required number of 3 ΔE_{00} 95th percentile intervals.

Table 1 shows the final target values for the control datasets. Reference dataset is designated as 0. -1 ~ -3, and +1~+3 are datasets with decreasing and increasing gamut volumes using one, two or three intervals of 3 95th percentile ΔE_{00}, respectively.

		White			Black			
Datset	L	a	b	L	a	b	Chroma	95th tgt
-3	83	0	1	33.69	0	0	1	3
-2	85	0	0	29.56	0	0	1	3
-1	87	0	0	25.03	0	0	1	3
0	89	0	0	20.03	0	0	1	3
+1	91	0	0	14.26	0	0	1	3
+2	93	0	-2	7	0	0	1	3
+3	95	0	-5.5	4	0	0	1.06	3

Table 1. *Target Lab, Chroma and 95th percentile ΔE_{00} values for Control group datasets.*

Subsequently, one reference dataset (G7+2d, see Table 2) from the control group, was used to alter gray balance and tonality (TR) and create systematically distorted datasets whereby adjacent datasets are 3 ΔE_{00} at 95% CRF apart (Test group).

ID	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6	Sample_7
Control	G7-3d	G7-2d	G7-1d	G7_0d	G7+1d	G7+2d	G7+3d
GB_C-R	GB C+3d	GB C+2d	GB C+1d	G7+2d	GB R+1d	GB R+2d	GB R+3d
GB_M-G	GB M+3d	GB M+2d	GB M+1d	G7+2d	GB G+1d	GB G+2d	GB G+3d
GB_Y-B	GB Y+3d	GB Y+2d	GB Y+1d	G7+2d	GB B+1d	GB B+2d	GB B+3d
TR_S	TR S-3d	TR S-2d	TR S-1d	G7+2d	TR S+1d	TR S+2d	TR S+3d
TR_TVI	TR TVI-3d	TR TVI-2d	TR TVI-1d	G7+2d	TR TVI+1d	TR TVI+2d	TR TVI+3

Table 2. *Datasets used in the experiment.*

In Table 2 the gray balance distorted datasets are listed linearly in rows by three sets of complimentary colors: Cyan-Red (C-R); Magenta-Green (M-G); and Yellow-Blue (Y-B). The implemented numbers of 3 95th percentile ΔE_{00} intervals with respect to the reference set are designated as 0d, +/- 1d ~ +/- 3d. G7®_0d is the reference dataset for the control group, while G7®+2d is the reference dataset for the test group datasets.

The steps and adjustment functions used for creating the test group datasets are described below.

1. Excel: Modify IT8.7/5 CMYK percentages with TVI, gray balance and S-Curve adjustments.
2. ColorThink Pro: assign reference dataset and export absolute LAB.
3. Excel: replace original IT8.7/5 CMYK values.
4. Curve4: find 95th percentile ΔE_{00} error vs. reference dataset.
5. Excel: Multiply strength of original CMYK modification by desired/actual 95th percentile fraction.
6. Repeat steps 2 - 5 (as needed) to obtain required number of 3 95th percentile ΔE_{00} intervals.
7. Export as CGATS .txt.

CMYK Adjustment Functions for tonality variations

TVI = sin (TV 0.008726646); [= sin (TV/2 Pi()/180); S-Curve]
S-curve = sin (TV -0.062831853); [= -sin (TV/100 360 Pi()/180)];
where TV = Tone Value, TVI = Tone Value Increase.

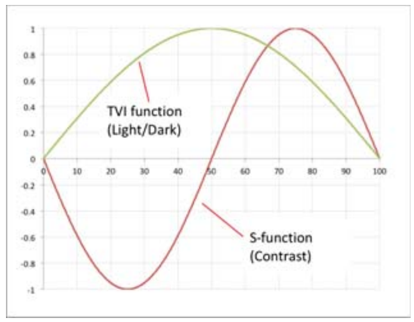
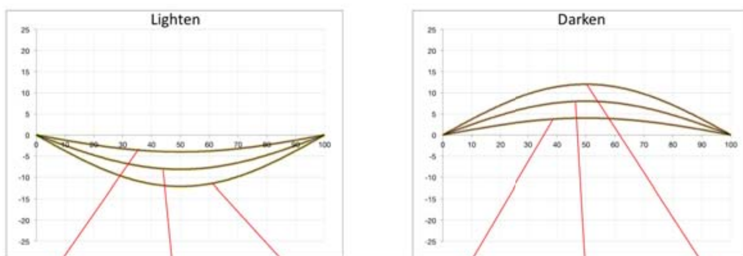


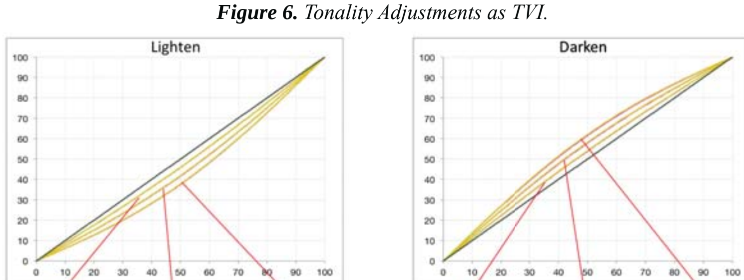
Figure 5. *TVI and S-curve functions.*

Datasets varying in tonality were created with the TVI or S-curve functions (Figure 5), equally for CMYK. Specific numerical implementations for tonality adjustments are presented below in Figures 6-9.



	TVI-1 25	50	75	TVI-2 25	50	75	TVI-3 25	50	75	TVI+1 25	50	75	TVI+2 25	50	75	TVI+3 25	50	75
C	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8
M	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8
Y	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8
K	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8

Figure 6. *Tonality Adjustments as TVI.*



	TVI-1 25	50	75	TVI-2 25	50	75	TVI-3 25	50	75	TVI+1 25	50	75	TVI+2 25	50	75	TVI+3 25	50	75
C	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8
M	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8
Y	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8
K	-3	-4	-3	-6	-8	-6	-9	-12	-9	+3	+4	+3	+6	+8	+6	+8	+12	+8

Figure 7. *Tonality Adjustments as Tone Curves.*

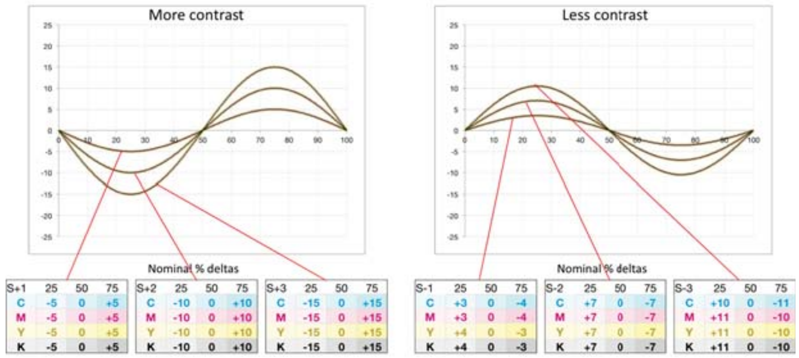


Figure 8. *Contrast modifications(S-Curve) as TVI.*

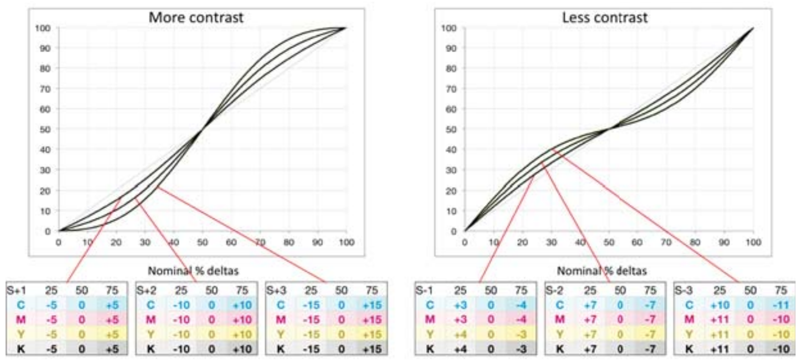
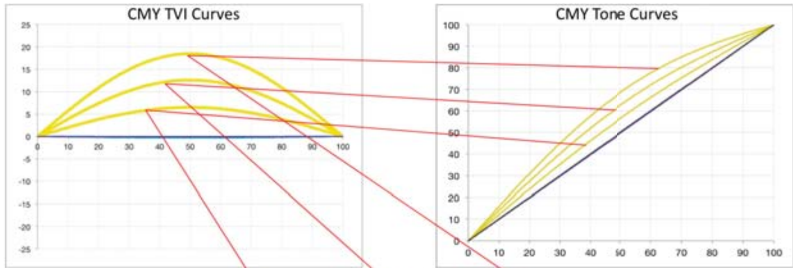


Figure 9. *Contrast modifications(S-Curve) as Tone Curves.*

Gray Balance Adjustment

Datasets varying in gray balance were created with separate TVI functions for CMY with no change in K. TVI curves were applied separately to C, M and Y. To minimize tonality side-effects, average C and M tone curves were maintained by subtracting half the difference between desired M TVI and C TVI from CMY target TVI values. For example, to implement 15 C TVI, changes to 7.5 C, -7.5 M, -7.5 Y were used. Blue cast in cases of B+1d, B+2d, and B+3d, was created by lightening Yellow TVI (not darkening M and C TVI). The gray balance adjustment functions are shown in Figures 10-15.

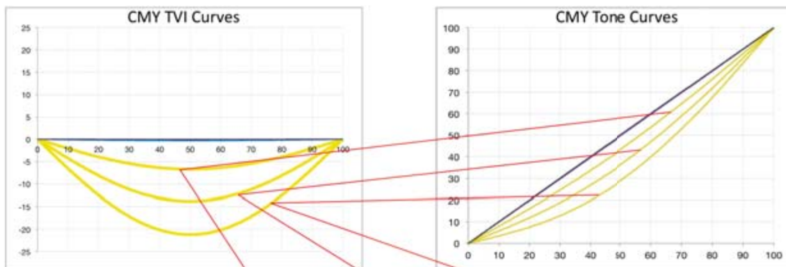


Nominal % deltas

Note: yellow adjustments typically have little or no effect on neutral tonality

Y+1	25	50	75	Y+2	25	50	75	Y+3	25	50	75
C	0	0	0	C	0	0	0	C	0	0	0
M	0	0	0	M	0	0	0	M	0	0	0
Y	+5	+7	+5	Y	+9	+13	+9	Y	+13	+19	+13
K	0	0	0	K	0	0	0	K	0	0	0

Figure 10. *Yellow-Biased Gray Balance Adjustment.*

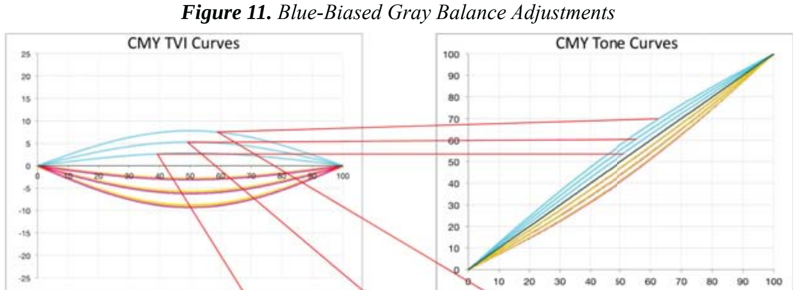


Nominal % deltas

Note: blue distortion achieved by reducing yellow to avoid unwanted neutral tonality effects

B+1	25	50	75	B+2	25	50	75	B+3	25	50	75
C	0	0	0	C	0	0	0	C	0	0	0
M	0	0	0	M	0	0	0	M	0	0	0
Y	-5	-7	-5	Y	-10	-14	-10	Y	-15	-21	-15
K	0	0	0	K	0	0	0	K	0	0	0

Figure 11. *Blue-Biased Gray Balance Adjustments*



Nominal % deltas

Note: cyan and magenta curves approximately cancel each other's effect on neutral tonality

C+1	25	50	75	C+2	25	50	75	C+3	25	50	75
C	+2	+3	+2	C	+4	+5	+4	C	+6	+8	+6
M	-2	-3	-2	M	-4	-6	-4	M	-7	-9	-7
Y	-2	-3	-2	Y	-4	-6	-4	Y	-6	-9	-6
K	0	0	0	K	0	0	0	K	0	0	0

Figure 12. *Cyan-Biased Gray Balance Adjustments*

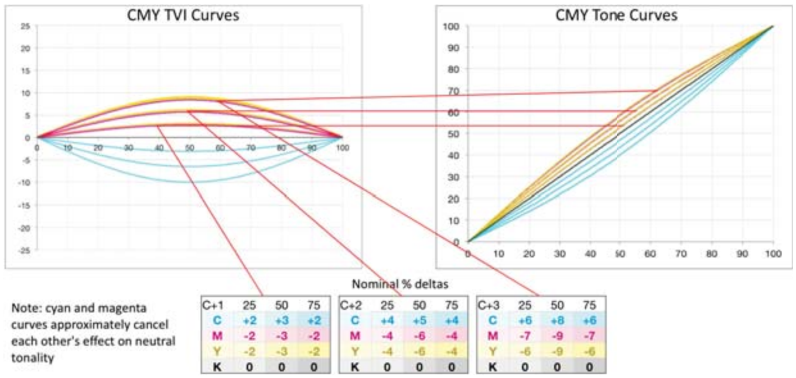


Figure 13. *Red-Biased Gray Balance Adjustments*

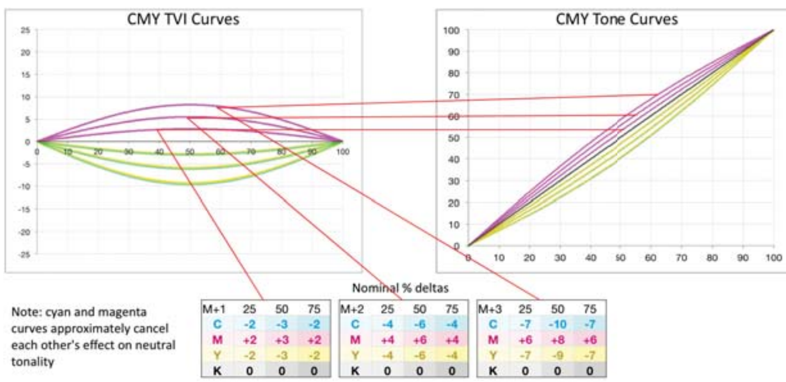


Figure 14. *Magenta-Biased Gray Balance Adjustments*

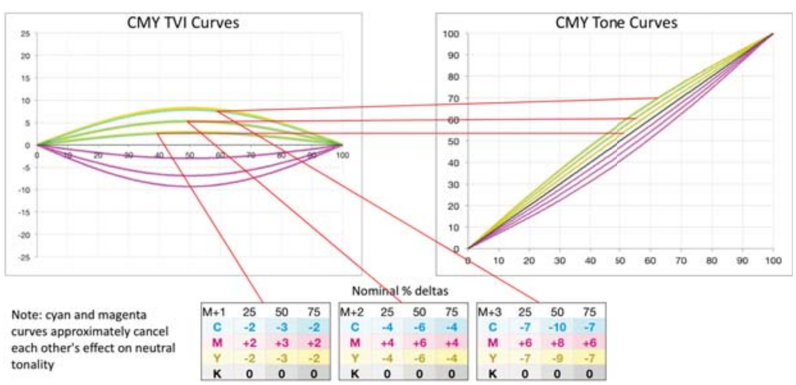


Figure 15. *Green-Biased Gray Balance Adjustments*

The curves and percentage values shown in Figures 6 -15 were the nominal initial adjustments. TVI or tone curve analysis of final datasets and/or ICC profiles could produce different CMYK percentages due to a number of factors, such as iterations to optimize 95th percentile difference, difficulty of calculating legacy TVI from CIEXYZ data, rounding errors within profiling software. Therefore, the verification procedure was implemented to confirm that the resulting datasets satisfy the designed requirements.

Experimental Methodology: Sample Verification Procedure

All datasets and profiles were verified in the following manner.

I. Use ColorThink Pro software and ICC profiles to compare 2-D gamut volume in a*b* plane.
   1) Choose 'Graph in 2D' view.
   2) Open an ICC profile in ColorThink Pro.
   3) Add other ICC profiles of interest to the 2D graph.
   4) Capture all 2-D gamuts.

II. Use Adobe Photoshop to simulate SCID pictorial color images. The RIT test block (CMYK, legacy images) was used in the experiment and the simulations.
   1) Assign an ICC profile, generated from the dataset of interest, to the RIT test block.
   2) Convert the image to the Lab space using the absolute colorimetric rendering intent.
   3) Save the converted image.
   4) Repeat the above procedures for all dataset conditions.
   5) Arrange the display using 'Tile All Vertically.'
   6) Capture all output conditions side by side.

III. Use ColorThink and Excel to analyze datasets of interest.
   1) Open the RIT Test Block in ColorThink Pro.
   2) Assign an ICC profile, generated from the dataset of interest, and select the absolute colorimetric rendering.
   3) Custom sample the 4 x 6 grid and save the color list (.txt).
   4) Copy and paste the color lists (reference and sample) to the Excel template to produce the comparison charts for a*b* gamut; gray balance of pre-defined triplets; and TVI of CMYK.

IV. Use ColorThink and Excel to analyze the color difference between adjacent datasets.

1) Open the reference dataset and a sample dataset in ColorThink Pro.
2) Select the ΔE_{00} function to find their color difference. Save the color list (.txt). Remove the sample dataset.
3) Repeat the above for all datasets.
4) Copy and paste the color lists to the Excel template to find out color differences of the adjacent datasets at the 95th percentile corresponding to the 81 largest value out of 1,617 values.

The ICC profiles are then applied to the test images and the output hard copies are produced using Epson P5000 printer as shown on the flowchart of Figure 16.

Finally, the hard copies of the Idealliance ISO 12647-7 [5] control strip (84 patches) were measured and the 95th percentile ΔE_{00} values between adjacent datasets were calculated.

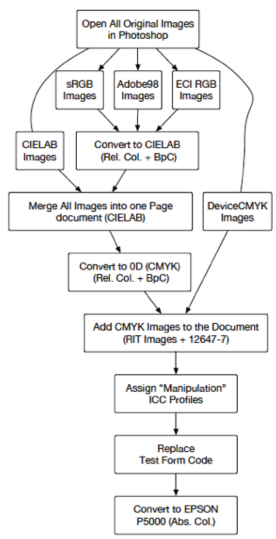


Figure 16. *Flowchart of the steps employed in the production of the hard copy output.*

Experimental Methodology: Psychometric Testing

Two psychometric tests were designed to evaluate consistency of color appearance of printed images for different datasets. The schematic of the first test is shown below (Figure 17). The observers' task was to rank candidate images according to their capacity to yield the best consistency of color appearance for the Control group of images, where one image (the reference G7® +2d) was removed from the sequence. The second test consisted of comparing all pairs of image sets in terms of their CCA and providing the CCA ratings for each set on the 1-5 scale, where 1 corresponds to an excellent CCA and 5 – to an unacceptable CCA.

Figure 17. *A procedure for the psychometric test: "There is a 'hole' in the (Control) group of images. Rank the candidate images according to the resulting consistency of color appearance of the group when a candidate image fills the gap, from the highest to the lowest."*

Participants were provided with the definition of CCA as the degree of visual consistency and/or shared visual appearance that a set of images possesses in the presence of visual differences. The CCA was further defined as an attribute of an image set that makes the images from the set belong to the same family. The experiments were conducted under the standard viewing conditions using the gti, ISO 3664:2009 [6] compliant, viewing booth. 12 participants, 8 male and 4 female observers, with the normal color vision and visual acuity took part in the experiments. Among the participants, there were 6 experts, who practice principles of color management and color process control in print and digital media, and 6 novices unfamiliar with printing and color reproduction. 6 sample sets listed in Table 3 served as experimental stimuli.

Results: Sample Verification and Simulation

Figure 18 demonstrates visual simulation for the Control group of datasets in comparison with the CRPC1 through CRPC7. The control group consisted of 7 new datasets derived from CRPC5 by scaling white point, black point and chroma while maintaining constant primary hue angles, G7® tonality and gray balance. The images from the control group exhibit more systematic changes from left to right compared to the CRPC images, and therefore appear more consistent as a group.

Figure 18. *Visual simulation of the RIT test block images using 7 control group datasets and 7 CRPC datasets. With the equal 95th percentile ΔE_{00} of 3 between adjacent datasets in the control group, the group of images appear more consistent compared to the CRPCs-based images.*

The test group consisted of 12 datasets varying in tonality: 3 lighter, 3 darker, 3 lower contrast, 3 higher contrast datasets; and 18 datasets with gray balance (3 each +CMYRGB) variations from one reference control dataset (G7®+2d), with 3 95^{th} percentile ΔE_{00} between any two adjacent datasets. The images from the test group datasets are illustrated in Figure 19 for tonality variations and Figure 20 for gray balance variations.

Figure 19. *Visual simulation for the datasets varying in tonality.*

Figure 20. *Visual simulation for the datasets with gray balance variations. Images are arranged in groups of complementary hue angle biases.*

As can be seen from the Figures 18-20, the step-wise changes in tonality and gray balance produce systematic changes in the appearance of images in regular steps, with no visible hue angle changes in the case of tonality variations, and no visible changes in lightness and contrast in the gray balance variation sets.

The gamut projections for the Control group (Figure 21) demonstrate more even changes in gamut sizes between the sets and higher hue angle consistencies compared to the CRPC group, supporting visual observations from Figure 18.

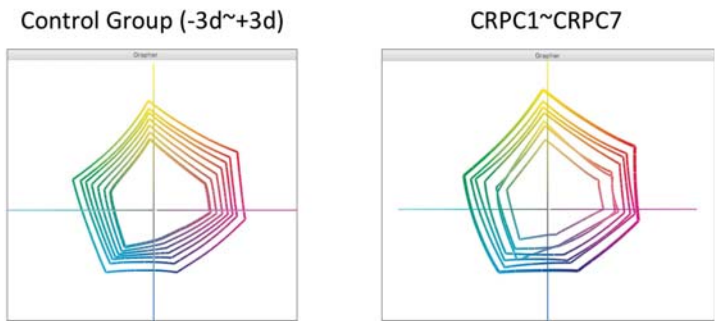


Figure 21. *Gamut projections on a*b* plane for the Control group and CRPCs datasets.*

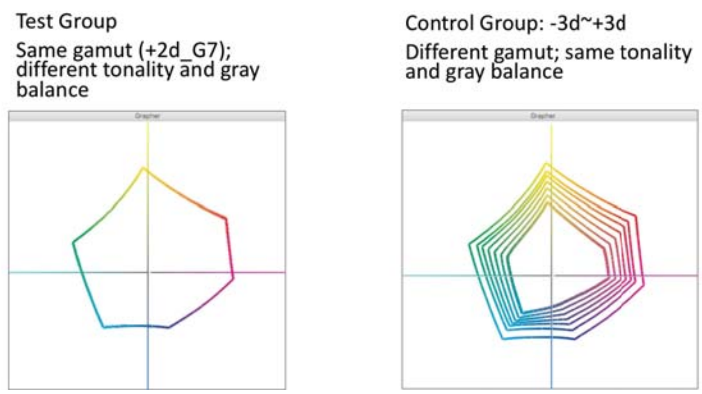


Figure 22. *Gamut projections on a*b* plane for the Test and Control group datasets.*

For all test groups the gamut sizes are constant as shown in Figure 22.

Table 3 compares the 95th percentile ΔE_{00} of adjacent datasets between seven Control group datasets and seven CRPC datasets defined by ISO 15339-2 [1]. This value is constant for the control group and is equal to 3.1 (0), but varies considerably for the CRPCs. Average 95th percentile ΔE_{00} for this group is 5.9 (2.5).

Control group

95th ΔE00	-3d_G7	-2d_G7	-1d_G7	0d_G7	+1d_G7	+2d_G7	+3d_G7
-3d_G7	-----						
-2d_G7	3.1	-----					
-1d_G7	6.2	3.1	-----				
0d_G7	9.2	6.2	3.1	-----			
+1d_G7	12.3	9.3	6.2	3.1	-----		
+2d_G7	15.2	12.2	9.2	6.2	3.1	-----	
+3d_G7	16.8	13.8	10.8	7.9	5.5	3.0	-----

CRPC1~CRPC7

95th ΔE00	CRPC1	CRPC2	CRPC3	CRPC4	CRPC5	CRPC6	CRPC7
CRPC1	-----						
CRPC2	7.2	-----					
CRPC3	8.8	6.6	-----				
CRPC4	12.5	6.2	9.9	-----			
CRPC5	14.4	8.2	11.8	4.2	-----		
CRPC6	15.2	9.0	12.3	6.3	3.3	-----	
CRPC7	19.0	12.5	15.5	7.5	5.3	4.0	-----

Table 3. *The 95th percentile ΔE_{00} between adjacent datasets for Control group and CRPCs.*

Figures 23-25 show gamut projections, gray balance and TVI curves for several test datasets in comparison with the reference G7®+2d to illustrate the datasets conformance with the design criteria. For the cyan-bias gray balance adjustments, the gamut and TVI curves remain constant while gray balance characteristics show distinct difference relative to the reference dataset (Figures 23 and 24). Moreover, these differences increase when two intervals of 95th percentile $\Delta E_{00} = 3$ are applied (compare middle plots of Figures 23 and 24). When tonality adjustments are used, for example, contrast S-curve modifications, gamut projections and gray balance characteristics remain constant, however TVI curves have changed (Figure 25).

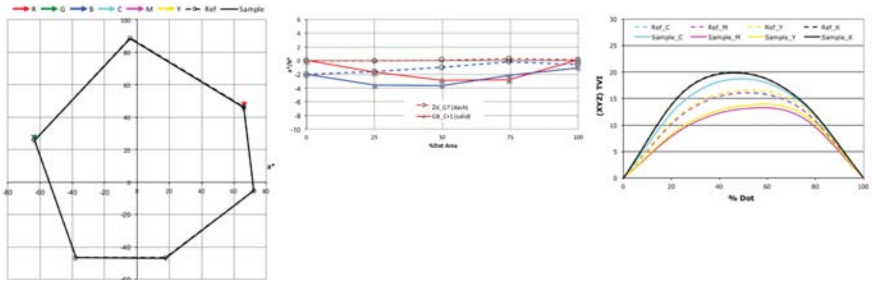


Figure 23. *Gamut projections (left panel), gray balance characteristics (middle panel) and TVI curves (right panel) for GB C+1d (Cyan-biased) and G7+2d (reference) datasets. There is a clear difference in gray balance between GB C+1d and G7+2d datasets.*

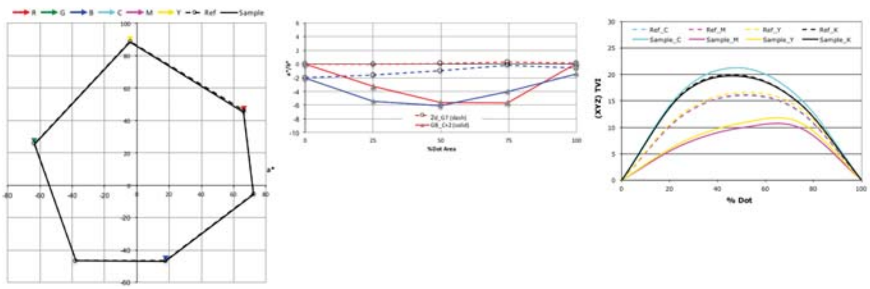


Figure 24. *Gamut projections (left panel), gray balance characteristics (middle panel) and TVI curves (right panel) for GB C+2d (Cyan-biased) and G7+2d (reference) datasets. The difference in gray balance between GB C+1d and G7+2d datasets increases.*

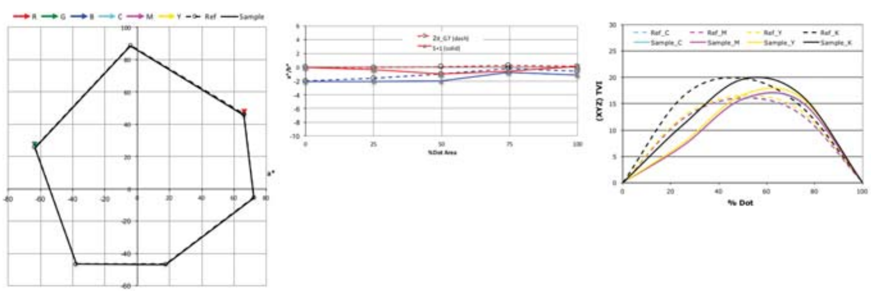


Figure 25. *Gamut projections (left panel), gray balance characteristics (middle panel) and TVI curves (right panel) for TR S+1d (tonality via contrast S-curve modifications) and G7+2d (reference) datasets. The differences between the datasets are apparent with respect to TVI curves.*

The sample verification procedure showed that the average 95th percentile color difference between adjacent datasets in the Control group was equal to 3.1 ΔE_{00}. The average 95th percentile color difference between adjacent datasets in the test group was 3.0 ΔE_{00}. The average 95th percentile ΔE_{00} values between the Control dataset (G7® +2d) and gray balance distorted group were 3 ΔE_{00}, 6 ΔE_{00}, or 9 ΔE_{00}. These results were in an agreement with the target values. Table 4 shows computed

95^{th} percentile ΔE_{00} values between adjacent datasets for the GB_Y-B and TR_TVI modifications.

GB_Y-B (from Y+3d to B+3d)

95th ΔE_{00}	3Y	2Y	1Y	0	1B	2B	3B
3Y	----						
2Y	3.0	----					
1Y	6.1	3.0	----				
0	9.1	6.1	3.0	----			
1B	12.2	9.1	6.1	3.0	----		
2B	15.2	12.1	9.1	6.1	3.0	----	
3B	18.1	15.1	12.1	9.1	6.0	3.0	----

TR_TVI (from TVI-3d to TVI+3d)

95th ΔE_{00}	-3TVI	-2TVI	-1TVI	0	+1TVI	+2TVI	+3TVI
-3TVI	----						
-2TVI	3.0	----					
-1TVI	6.1	3.0	----				
0	9.0	5.9	3.0	----			
+1TVI	12.0	8.9	6.0	3.0	----		
+2TVI	14.9	11.9	8.9	6.0	3.0	----	
+3TVI	17.7	14.9	11.9	9.0	6.0	3.0	----

Table 4. *The 95^{th} percentile ΔE_{00} between adjacent datasets for GB_Y-B and TR_TVI modifications. 0 corresponds to the reference G7®+2d dataset.*

In addition to computing ΔE_{00} between adjacent datasets with the examples presented in Tables 3 and 4, the CIELAB measurements of the Idealliance ISO 12647-7 [5] control wedge from the Epson P5000 proof prints were used to obtain the 95^{th} percentile ΔE_{00} values. The comparisons between these two measurement approaches are presented in Table 5.

Control group
G7-3d ~ G7+3d

95th ΔE_{00}	-3d_G7 / -2d_G7	-2d_G7 / -1d_G7	-1d_G7 / 0d_G7	0d_G7 / +1d_G7	+1d_G7 / +2d_G7	+2d_G7 / +3d_G7	Ave.	Stdev
Dataset	3.1	3.1	3.1	3.1	3.1	3	3.1	0
Proof	3.5	2.8	3.3	3.0	3.2	4.5	3.4	0.6

Test group
GB B+3d ~ GB Y+3d

95th ΔE_{00}	3Y / 2Y	2Y / 1Y	1Y / 0	0 / 1B	1B / 2B	2B / 3B	Ave.	Stdev
Dataset	3.0	3.0	3.0	3.0	3.0	3.0	3.0	0
Proof	3.1	2.9	3.1	3.0	3.3	3.3	3.1	0.2

Test group
TVI -3d ~ TVI+3d

95th ΔE_{00}	-3TVI / -2TVI	-2TVI / -1TVI	-1TVI / 0	0 / +1TVI	+1TVI / +2TVI	+2TVI / +3TVI	Ave.	Stdev
Dataset	3.0	3.0	3.0	3.0	3.0	3.0	3.0	0
Proof	2.8	3.2	4.1	2.3	3.0	2.7	3.0	0.6

Table 5. *Device color difference as the 95^{th} percentile ΔE_{00} between adjacent datasets. Subset of 21 datasets.*

When the subset of 21 datasets was used to calculate the device color differences, the results were close to the intended, i.e., 3 ΔE_{00} (95^{th} percentile) with ~0 standard deviation. When the 21 sets of color measurement data from the proofs were used to assess the device color differences, the results were also close to the intended, i.e., 3 ΔE_{00} (95^{th} percentile). The variability was about 0.5 ΔE_{00} (one standard deviation). The sources of variation include device link accuracy, proofer media, proofer calibration, and color measurement.

Results: Psychometric Testing

As described in the previous sections, the experiments consisted of two tests conducted under the standard viewing conditions using the gti viewing booth.

The goal of the first experiment was to develop a relative CCA scale for the set of Control images “perturbed” by alternate reference images based on the alteration method and its magnitude assessed via the ΔE_{00} measure. The goal of the second experiment was to compare different image groups and therefore underlying sample preparation methods with respect to the resulting CCA.

Among 12 observers, 6 participants were experts familiar with the principles of color management and color quality control, and 6 participants were novices without prior knowledge related to color reproduction and color management. 6 sample sets listed in Table 3 served as experimental stimuli.

In the first test, participants were presented with the set of images from the Control group organized in a sequence from G7®-3d to G7®+3d, where the reference G7®+2d image was removed from the sequence. The participants were given the candidate test images for filling the missing image and instructed to rank the test images according to the resulting consistency of color appearance of the set, from the highest to the lowest (see illustration on Figure 26 1). The test images were comprised of all altered images from the test groups plus the reference G7®+2d image. Since there were many candidate images - reference image plus 30 different images from the test groups, prior to ranking, the participants were asked to discard obvious misfits that would produce the low resulting CCA for the group, and thus, reduce the test image set to about 10 potential candidates.

In the second test, the participants were asked to compare and rate sample sets in accordance with the apparent consistency of color appearance. Each sample set consisted of images from the same test group: Control, TVI, S, GB C-R, GB M-G, GB Y-B. For the second test, the sample sets from Table 2 were reduced from 7 images to 5 by removing the flanks (Figure 26 2).

1) Rank samples that fit in the image set for best CCA

2) Compare and rate sample sets for demonstrating CCA.

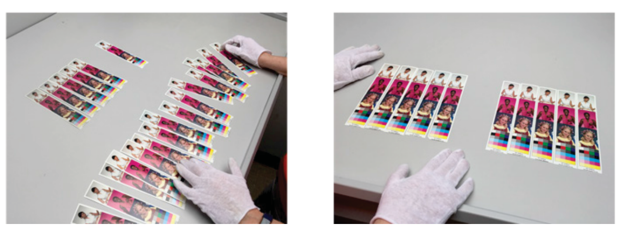


Figure 26. *Photos illustrating the experimental procedures.*

Participants were provided with the definition of CCA as an attribute that makes images appear to belong to a “family”, that is having visual consistency and/or shared visual appearance in the presence of perceptible visual differences.

Consistency of Color Appearance from Ranking images to complete the set

To analyze the first experiment, interval scale from the rank data was calculated using Thurstone’s law of comparative judgment [7]. The frequency of assigned ranks and calculations leading to the derivation of interval CCA scale are shown in Table 6. Overall 22 images were ranked by the participants, whereas each individual participant ranked only 10 images after discarding the misfits. Images that were not selected for ranking by some participants (but ranked by others) were assigned the rank 11 for those participants.

rank	Control G7 +2d	TR S+1d	TR S-1d	GB R+1d	GB M+1d	TR TVI-1d	TR TVI+1d	TR TVI+2d	GB B+1d	GB Y+1d	TR S-2d	GB C+2d	GB C+1d	GB G+1d	TR S-3d	GB B+2d	TR TVI-3d	TR TVI-2d	GB R+2d	GB G+2d	TR S+2d	GB M+2d
1	3	0	1	3	0	2	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0
2	4	2	0	1	0	0	1	0	1	2	0	0	0	1	0	0	0	0	0	0	0	0
3	2	1	2	0	0	0	3	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0
4	0	1	0	2	0	4	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	0
5	0	1	2	0	1	0	1	1	2	0	0	1	2	0	0	0	0	1	0	0	0	0
6	1	1	0	0	1	2	0	0	1	2	0	1	2	0	0	0	0	0	0	0	0	1
7	1	1	0	0	0	0	1	1	0	0	0	0	3	1	0	0	1	0	2	0	1	0
8	1	0	2	1	2	1	0	1	0	1	0	0	0	2	0	0	0	1	0	0	0	0
9	0	0	2	0	0	0	0	2	3	1	0	2	0	1	1	0	0	0	0	0	0	0
10	0	0	0	1	2	0	0	0	2	2	0	0	1	1	1	0	0	0	0	1	1	0
11	0	5	3	4	6	3	6	7	1	4	11	7	3	3	10	11	11	8	9	11	10	11
frequency	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Mean rank (Mr)	3.17	7.00	7.00	6.25	9.42	5.92	7.42	9.58	6.67	8.08	10.33	9.17	7.25	7.08	10.75	10.42	10.67	9.00	9.75	10.92	10.58	10.58
Proportions (Nr-Mr)/(Nr-1)	0.78	0.40	0.40	0.48	0.16	0.51	0.36	0.14	0.43	0.29	0.07	0.18	0.38	0.39	0.03	0.06	0.03	0.20	0.13	0.01	0.04	0.04
z score	0.78	-0.25	-0.25	-0.06	-1.00	0.02	-0.36	-1.07	-0.17	-0.55	-1.50	-0.90	-0.32	-0.27	-1.96	-1.57	-1.83	-0.84	-1.15	-2.40	-1.73	-1.73

Table 6. *Ranking frequency and z-scores as the CCA scale for the first experiment.*

The z-scores correspond to values of image sets on an interval scale of relative CCA. The CCA scores for the sets based on the ranked images are plotted in Figure 27. For simplicity, TR, GB and d designations are left out.

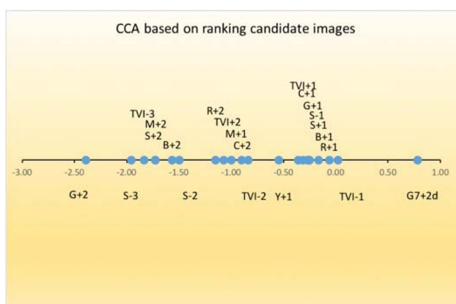


Figure 27. *Color Consistency scale obtained in the ranking experiment. It represents the CCA of the control group completed by the images labeled using the type of manipulation and the number of intervals from the reference G7+2d dataset.*

From Figure 27 it can be seen that the set with the reference image (G7®+2d) has the highest CCA score, followed by the images that deviate by one 95-percentile ΔE_{00} interval, and then by two and three 95 percentile ΔE_{00} intervals. Notably, the images that are altered by one interval tend to form a group (with the exception of M+1, that is 1 step - Magenta-biased GB shift). Groupings can be also seen for images that are two intervals apart from the reference. The interval scale shows how far the sets are from each other in terms of CCA. Figure 28 demonstrates the correlation of the relative CCA scores and the measured ΔE_{00} (95 percentile) between adjacent sets using printed Idealliance ISO 12647-7 [5] control strips. The high R^2 value exhibits a good agreement between the measured data and the experimentally obtained CCA scores.

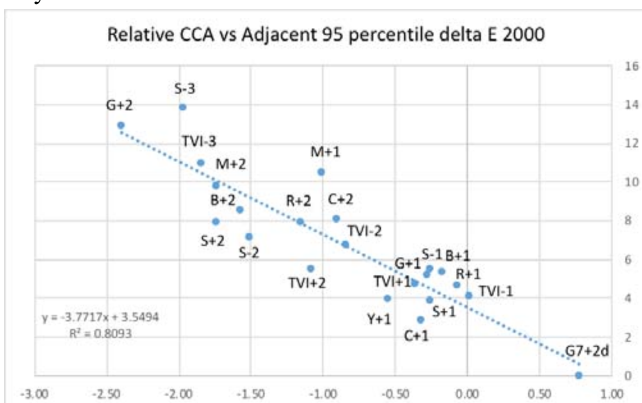


Figure 28. *Color Consistency scale from the ranking experiment (x-axis) vs measured 95-percentile ΔE_{00} between consecutive adjacent datasets (y-axis).*

Consistency of Color Appearance from Rating image sets

In the second experiment participants rated the CCA of image sets presented in pairs, where a rating of 1 corresponded to the highest, excellent CCA, and 5 – to the lowest, unacceptable CCA. The data obtained in the second experiment were analyzed using one-way ANOVA to test the influence of the image set factor on the CCA ratings. The results are highly significant – the image group factor, which corresponds to the method of altering images, significantly affects the perceived consistency of color appearance (F-ratio =15.30; $p< 0.0001$). The control group had the highest CCA (lowest mean rating), followed by tonality altered groups – TVI and S, and gray balance altered groups – Cyan-Red biased set, GB C-R, Magenta-Green biased set, GB M-G, and Yellow-Blue biased set, GB Y-B (see Table 7 and Figure 29). When comparing the ratings for the pairs of individual sets using Tukey's test , the Control and TVI groups had significantly higher CCA than the rest of the sets, although there was no statistically significant difference between the Control and TVI groups themselves. The lowest score was observed for the Yellow-Blue-biased image set. The CCA for this set was significantly lower than for the S, Control and TVI groups (Table 7).

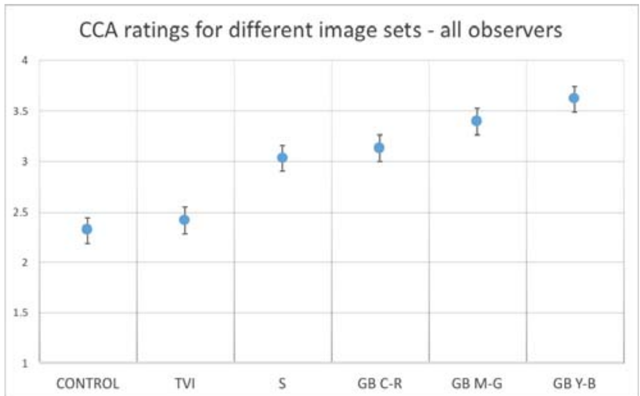


Figure 29. *Mean Color Consistency ratings for different image sets. Data for all 12 observers.*

LSMeans Differences Tukey HSD

α= 0.050 Q= 2.8654

Level				Least Sq Mean
GB Y-B	A			3.62
GB M-G	A	B		3.40
GB C-R	A	B		3.13
S		B		3.03
TVI			C	2.42
Control			C	2.32

Levels not connected by same letter are significantly different.

Table 7. *Mean CCA ratings for different image sets and their comparison using Tukey method. Data for 12 observers.*

To evaluate the potential influence of knowledge related to print color reproduction on the ratings, an analysis was conducted for experts and novices' data separately. The analysis results for the expert group are shown in Figure 30 and Table 8, and for the novice group – in Figure 31 and Table 9, accordingly.

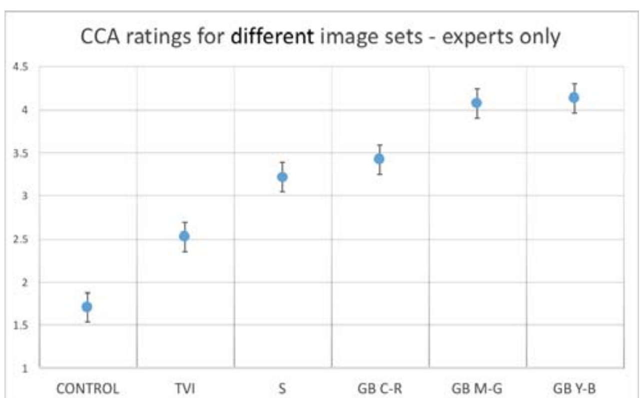


Figure 30. *Mean Color Consistency ratings for different image sets. Data for 6 expert observers.*

LSMeans Differences Tukey HSD

α= 0.050 Q= 2.88174

Level						Least Sq Mean
GB Y-B	A					4.13
GB M-G	A	B				4.07
GB C-R		B	C			3.43
S			C			3.23
TVI				D		2.53
Control					E	1.70

Levels not connected by same letter are significantly different.

Table 8. *Mean CCA ratings for different image sets and their comparison using Tukey method. Data for 6 expert observers.*

As in the case of all participants, the image group factor was significant in affecting the perceived CCA ratings by expert observers (F-ratio =31.72; p< 0.0001). The order of the scores was also the same: the control group images had the highest CCA (lowest mean rating), followed by the tonality altered groups – TVI and S, and the gray balance altered groups – the Cyan-Red biased group, the Magenta-Green biased group, and the Yellow-Blue biased group (see Table 8 and Figure 30). However, the differences in the CCA scores between the image sets are larger for the experts, which can be seen from Table 8. Comparisons between image sets using Tukey's test show that the Control group has significantly higher CCA than all other groups, the TVI group has the second highest CCA, also significantly different from the rest, while the Yellow-Blue biased group (GB Y-B) has significantly lower score, than other image groups (Table 8).

The results for the novices show smaller differences between image sets (Figure 31), where only the difference between the TVI and the Yellow-Blue biased group reaches the level of statistical significance (Table 9). The significance level for the image group factor was also smaller (F-ratio =2.36; p=0.04).

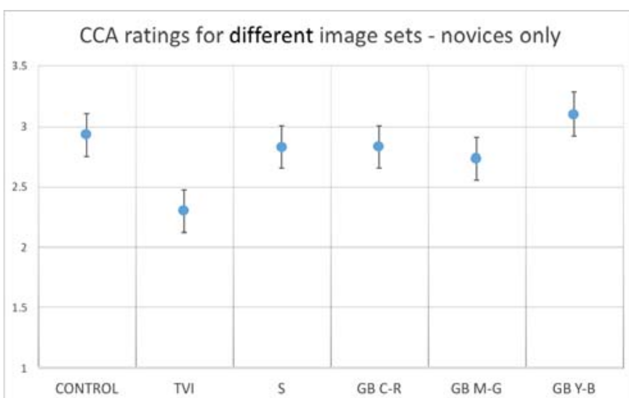


Figure 31. *Mean Color Consistency ratings for different image sets. Data for 6 novice observers.*

LSMeans Differences Tukey HSD

α= 0.050 Q= 2.88174

Level			Least Sq Mean
GB Y-B	A		3.10
Control	A	B	2.93
GB C-R	A	B	2.83
S	A	B	2.83
GB M-G	A	B	2.73
TVI		B	2.30

Levels not connected by same letter are significantly different.

Table 9. *Mean CCA ratings for different image groups and their comparison using Tukey method. Data for 6 novice observers.*

The results of the experiments support the hypothesis that CCA for a set of multiple images is the highest when based on datasets with varying gamut volumes, while having consistent tonality, gray balance and hues relative to substrate (Control group), compared to the condition, when the gamut volumes are kept constant, but tonality and gray balance are varied (Test groups). This is particularly evident from the data obtained in the second experiment when the CCA ratings of the expert participants were analyzed separately. Additionally, image sets altered by changing tonality (TVI and S sets) were on the second and third place based on the CCA image sets rating task, while image sets with the changes in gray balance were judged as having lower CCA. The first experiment also demonstrated that the CCA of the images from the Control group was the highest when completed by the reference image belonging to the same group (G7®+2d), based on the relative CCA scale. Moreover, images from the TVI and S groups were chosen more frequently, and proportionally are better represented on the CCA scale derived based on the results of the first test.

The comparison of the measured colorimetrical differences and the constructed scale values provides experimental evidence in support of using the 95th percentile ΔE_{00} as the measure to quantify differences in CCA in the present experiment. However, further experiments are needed to overcome limitations of the present study in terms of image selection and manipulation and to develop a metric that does not depend on specific dataset creation methods and image content.

Conclusions

A methodology for studying Consistent Color Appearance for a set of images is proposed.

Psychometric tests show that the color appearance of an image set with chroma changes due to gamut variations appears to be more consistent than due to changes in tonality and gray balance reproduction.

There is a discrepancy between experts and novices when judging CCA which may be attributed to image quality preferences.

The device-based 95th percentile ΔE_{00} is shown to be a good metric, along with the reproduction factors such as tonality, gray balance, etc., for quantifying Consistent Color Appearance in various ink-paper-press conditions. The 95th percentile $\Delta E_{00} \sim 3$ differences were perceptible in terms of CCA evaluations.

Additional experiments are needed to develop a general CCA model to evaluate the effects of pictorial scene on the CCA.

Acknowledgements

We wish to thank RIT students, faculty and staff for their involvements in the psychometric experiments.

References

1. ISO/TS 15339-2:2015 (2015). Graphic technology — Printing from digital data across multiple technologies — Part 2: Characterized reference printing conditions, CRPC1 - CRPC7.
2. CGATS/IDEAlliance TR 015 (2015). Graphic technology — Methodology for Establishing Printing Aims Based on a Shared Near-neutral Gray-scale.
3. Fedorovskaya, E.A., Chung, R., Hunter, D., ans Urbain, P. (2017). Effects of Gray Balance and Tone Reproduction on Consistent Color Appearance. TAGA 2017.
4. CGATS TR016 (2014). Graphic technology —Tolerance and Conformity Assessment.
5. Idealliance ISO 12647-7 (2013). 3-Row Digital Control Wedge.
6. ISO 3664:2009 (2009). Graphic technology and photography -- Viewing conditions.
7. Thurstone, L.L. (1927). A law of comparative judgement. Psychological Review, 34, 273-286.