Journal of Imaging Science and Technology® 68(5): 1–13, 2024.

Monastic Color Reproduction: A Software Tool for Printing and Assessing the Monk Skin Tone Scale

Wei-Chung Cheng
U.S. Food and Drug Administration, Silver Spring, Maryland, United States
E-mail: wei-chung.cheng@fda.hhs.gov

Abstract. *The Monk Skin Tone (MST) scale, comprising ten digitally defined colors, represents a diverse range of skin tones and is utilised by Google to promote social equity. The MST scale also has the potential to address health disparities in the medical field. However, these colors must be printed as physical charts for health practitioners and researchers to conduct visual comparisons. This study examines the colorimetric characteristics of the MST scale and establishes two per-patch acceptance criteria to determine the acceptability of a printed MST chart using a four-tier grading method. A software tool was developed to assist end users in printing accurate MST charts. The tool was tested with four printer/paper combinations, including inkjet, dye sublimation, and laser printers. Results show that a wide color gamut covering the brightest and darkest MST levels is crucial for producing an accurate MST chart, which can be achieved with a consumer-grade inkjet printer and glossy paper.*
Keywords: *calibration, color matching, print quality, Monk skin tone scale, medical application*

[DOI: 10.2352/J.ImagingSci.Technol.2024.68.5.050404]

1. INTRODUCTION

Skin tone refers to the natural color of a person's skin, primarily determined by the amount and type of melanin [1]. Assessing skin tone is crucial in the medical field. For example, in dermatology, different skin tones contain varying levels of melanin, influencing sensitivity to UV radiation. Accurate skin tone assessment is also essential for diagnosing and treating skin conditions [2]. Pulse oximeters are medical devices used to measure blood oxygen saturation levels. They gained unprecedented attention and became ubiquitous worldwide during the COVID-19 pandemic. Inaccurate readings from pulse oximeters have been reported in the literature, particularly for patients with darker skin, leading to potential misdiagnoses and delayed treatments [3–12].

A color test target can be either physical or digital. Physical targets are designed, manufactured, and validated to provide consistent color swatches for visual assessment or optical measurement. For example, the ColorChecker target [13] and Munsell Book of Color [14] are commonly used to communicate colors and calibrate devices. The PanTone SkinTone Guide [15] and Picterus SkinChecker [16] are physical targets designed for skin tone applications [17]. The ColorGauge targets [18] are miniature physical targets that have been used to evaluate color performance of endoscopes [19].

The Monk Skin Tone (MST) scale is a set of ten colors intended to "represent the full spectrum of human skin tones" [20]. However, the MST scale is defined digitally in a color space. When health practitioners or researchers require actual swatches for visual assessment and skin tone evaluation, a physical MST chart must be created. Unfortunately, there is currently no validated physical MST chart available to the public, so users, whether or not they have expertise in color management, must create their own MST charts. The experiment conducted in [21] demonstrated the variability in printing the MST chart using 16 different printers without any color management. The deviation was so pronounced that the authors recommended changing the MST reference values for novice users.

When making an MST chart, the following four intertwined elements must be considered:

- **Reference**: The reference MST chart must be defined colorimetrically (i.e., objectively, quantitatively, and device-independently) such that the same target is used by all users and skin tone data can be correctly collected and interchangeable.
- **Figure of merit**: The printed MST chart needs to be compared with the reference using a rational, standard, and objective metric to evaluate the errors such that the quality can be objectively determined.
- **Acceptance criteria**: The acceptance criteria of the printed MST chart need to be quantitatively established with justifications such that only qualified MST charts are used in data collection.
- **Implementation**: The procedure for calibrating and printing the MST chart needs to be described clearly. Minimum requirements for the printer, paper, instruments, software programs, and expertise need to be specified.

Colorimetry is the presumptive tool for measuring, evaluating, communicating, and reproducing skin tones [22–25]. In dermatology, the utility of colorimetry and instrument-based measurements has been recognized and disseminated as early as 1996 [26]. However, in other fields, colorimetry is not properly utilized to solve real-world problems.

In this manuscript, the characteristics of the MST scale are colorimetrically defined as the reference and analyzed

Received June 3, 2024; accepted for publication Oct. 5, 2024; published online Oct. 28, 2024. Associate Editor: Simone Bianco.
1062-3701/2024/68(5)/050404/13/$25.00

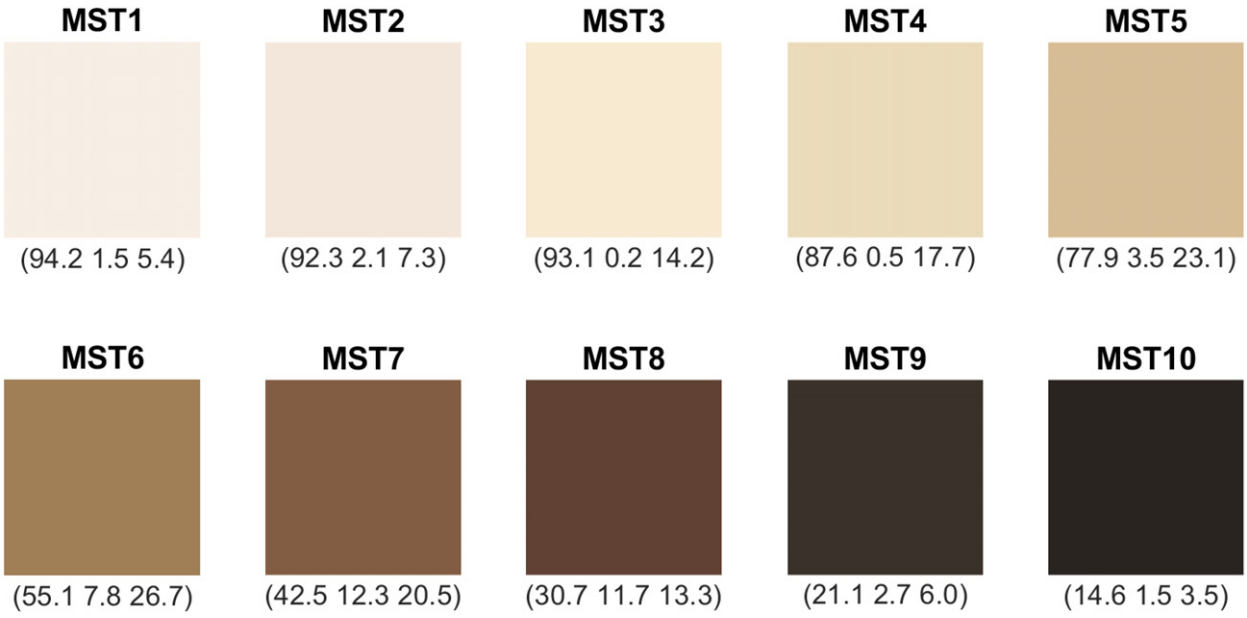

Figure 1. The ten MST levels and their CIELAB values.

in Section 2. Usage of the MST chart is modeled as a color matching task in Section 3.1 using the CIE color difference metric as the figure of merit. Novel acceptance criteria and an associated grading method for MST charts are proposed in Section 3.2. A software tool developed to facilitate the implementation and evaluation of printed MST charts is described in Section 4. The results of applying the software tool to different printers are presented in Section 5.

The following terminologies are used throughout this manuscript to refer to different physical and digital objects. A *color* is a color coordinate that can be defined in the three-dimensional CIELAB color space. The *MST scale* is a set of ten specific colors that will be discussed in Section 2.1. An *MST level* is one of the ten colors defined in the MST scale. A *patch* is a uniform area that contains pixels or material of the same color. An *MST chart* contains ten patches representing the MST scale. A *swatch* is a patch printed on a certain type of paper that needs to be illuminated by an external light source to see its color. A *printed MST chart* is a collection of ten swatches representing the MST scale.

2. REFERENCE MST SCALE

2.1 *Characteristics*

Listed in Table I are the CIELAB and RGB values of the MST scale published by Google [27]. For illustration purposes, the ten MST colors are shown in Figure 1.

The brightest level, MST1, has a lightness value of 94.2L^*, which is considered very high and as such difficult to reproduce. As a comparison, the calibration white tile of some colorimeters (e.g., ColorReader, DataColor [28] and EyeOne Pro, X-Rite [29]) is about 96L^*. Regular paper used by copy machines is lower than 94L^*.

Table I. CIELAB and RGB values of the MST scale.

MST Level	L^*	a^*	b^*	RGB (sRGB)
1	94.2	1.5	5.4	(246, 237, 228)
2	92.3	2.1	7.3	(243, 231, 219)
3	93.1	0.2	14.2	(247, 234, 208)
4	87.6	0.5	17.7	(234, 218, 186)
5	77.9	3.5	23.1	(215, 189, 150)
6	55.1	7.8	26.7	(160, 126, 86)
7	42.5	12.3	20.5	(130, 92, 67)
8	30.7	11.7	13.3	(96, 65, 52)
9	21.1	2.7	6.0	(58, 49, 42)
10	14.6	1.5	3.5	(41, 36, 32)

The darkest level, MST10, has a lightness value of 14.6L^*, which is considered very low and also difficult to reproduce. In comparison, the darkest patch in the Munsell Book of Color [14] is about 25L^*. The black-out tape (Advance Gaffa, Advance Tapes International) used in optics labs and darkrooms is brighter than 20 L^*. The darkest oil paint (Mars Black, Winsor and Newton) is about 28 L^*.

As hinted by the name "scale," the ten MST shades appear to be arranged in a descending order of lightness. However, MST3 (93.1L^*) is brighter than MST2 (92.3L^*). In other words, MST2 and MST3 are out of order in lightness and fail to preserve the monotonicity in the lightness dimension. Therefore, when using the MST scale to match skin tones, the observer needs to carefully rely on the chromaticity difference, represented by $\Delta a^* = -1.84$ and $\Delta b^* = 6.92$ as shown in Table II, to distinguish these two

Table II. Color differences between adjacent MST levels.

Between	ΔE^*_{ab}	ΔL^*	Δa^*	Δb^*	ΔC^*_{ab}	Δh_{ab}
MST1 – MST2	2.74	−1.94	0.56	1.86	1.94	−0.31
MST2 – MST3	7.21	0.82	−1.84	6.92	6.64	14.94
MST3 – MST4	6.56	−5.52	0.24	3.54	3.55	−0.61
MST4 – MST5	11.47	−9.67	3.01	5.39	5.64	−7.05
MST5 – MST6	23.44	−22.76	4.31	3.60	4.45	−7.70
MST6 – MST7	14.82	−12.67	4.54	−6.21	−3.90	−14.75
MST7 – MST8	13.83	−11.79	−0.66	−7.20	−6.23	−10.21
MST8 – MST9	15.07	−9.61	−8.98	−7.37	−11.18	16.91
MST9 – MST10	7.01	−6.46	−1.21	−2.44	−2.72	1.47

colors. Table II also shows that the smallest color difference, 2.74 ΔE^*_{ab}, is located between MST1 and MST2.

Figure 2 depicts the ten MST levels in the CIELAB color space compared with 3,840 measured skin tones from Xiao et al. [23]. In this study, the authors used spectrophotometers to obtain skin color measurements of four ethnic groups (Caucasian, Chinese, Kurdish, and Thai) at four locations (forehead, cheek, inner arm, and back of hand). The dense cluster near the center of the MST curve shows that the MST scale does not align well with Xiao's data. It is unclear whether the MST CIELAB values were determined by physically measuring human subjects with optical instruments. Nevertheless, design of the MST scale is beyond the scope of this paper.

In summary, the MST scale features extremely bright patches and extremely dark patches, non-monotonicity in the lightness dimension, uneven spacing in all dimensions between the ten patches, and high similarity between the first three MST levels.

2.2 *RGB or sRGB?*

On Google's website, the ten MST levels are specified in five different formats, including HEX, HSL, RGB, CIELAB, and CIELCH [27]. The five formats indicate three different color spaces. The CIELAB and CIELCH data represent the same standard, device-independent color space in the Cartesian and cylindrical coordinates, respectively [30]. The RGB and HEX data are the same numbers written in the decimal and hexadecimal form, respectively. No specific color space is associated with the RGB/HEX data. In other words, the RGB/HEX data represent ten colors in an undefined or device-dependent color space. The HSL data are converted from the RGB/HEX data using the *hue/relative chroma/intensity* color space [31], which is also device-dependent. We confirmed that these RGB/HEX values are actually the sRGB data converted from the CIELAB data using the D65 white point. Although it is well-known that the sRGB is a standard, device-independent color space defined in IEC 61966-2-1 [30] and is different from the device-dependent RGB representation, RGB values provided by Google may mislead some readers into thinking that driving a printer with RGB values produces the corresponding

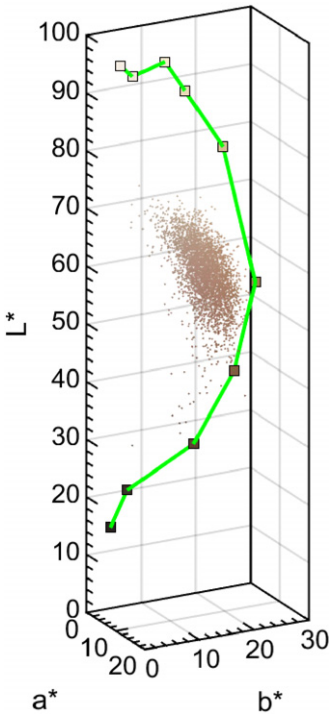

Figure 2. The ten MST colors (squares connected with green lines) in the CIELAB space in comparison with the 3,840 colors (dots) from Xiao's dataset.

CIELAB colors. The consequences of using these sRGB values to print the MST chart will be evaluated in Section 5.3.

3. TOLERANCES FOR THE MST SCALE

3.1 *Color Matching with the MST Chart*

The typical intended use of an MST chart is for a human observer to visually find the best match to a given target skin tone from the ten patches in the MST chart. This process can be modeled as a series of nine two-alternative forced choice trials. A pool initially contains the ten MST patches. Two patches are randomly chosen and compared with the target skin tone. Based on visual assessment, the human observer estimates the color differences for both patches. After comparing the two color differences, the human observer puts the patch with the smaller color difference back to the pool and discards the other. This process is repeated nine times until only one patch is left. In this study, the same color matching method is used to evaluate a printed MST chart except that the visual assessment is substituted with an ideal observer by using the CIE 1976 color difference metric ΔE^*_{ab}.

The individual typology angle (ITA) [26] is used by some researchers to evaluate the MST chart. Using the ITA metric to evaluate a printed MST chart is inadequate because the dimensionality of the color space is reduced from three to one, which is insufficient to identify or compare colors. In addition, calculating the ITA requires the CIELAB information, which is not feasible for a human subject to use in the visual assessment process.

Google also provides ten "MST orbs" for use in research studies [27]. The orbs for MST1 and MST2 are shown in

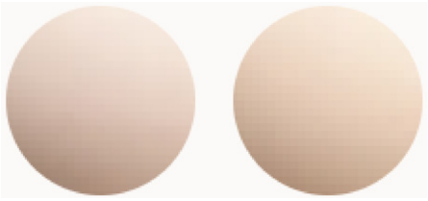

Figure 3. Google's orbs for MST1 (left) and MST2.

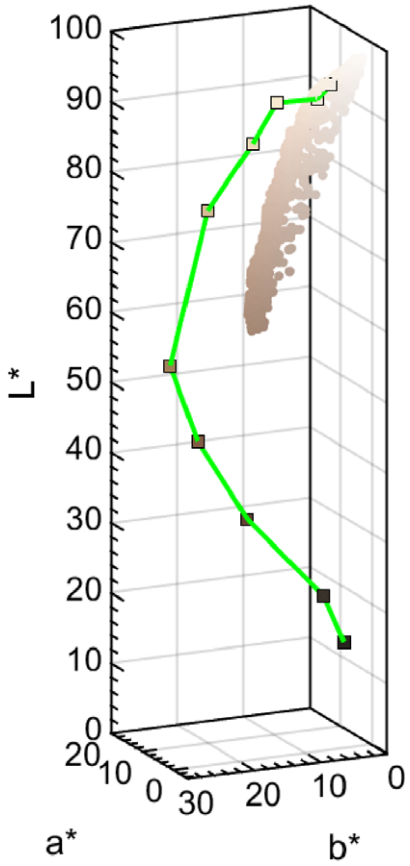

Figure 4. The 1,075 colors contained in Google's MST1 orb (dots) and the MST levels (squares connected with green lines) in the CIELAB space.

Figure 3. Each of the ten orbs contains hundreds of different colors, and some colors overlap between adjacent MST levels. After analyzing 8,000 pixels in the MST1 orb, 1,075 different colors were identified and plotted in Figure 4. These colors not only enclose both MST1 and MST2 levels, but also span a wide range that can be matched to MST1 through MST6. Using such a method to match skin tones may lead to confusion among users when distinguishing adjacent MST levels.

3.2 *Acceptance Criteria and Grading for the MST Scale*

Two acceptance criteria are established for evaluating each of the ten printed MST swatches.

3.2.1 *Correct Classification Criterion*

When comparing a printed MST swatch with an ideal MST chart, which has the ten (reference) MST levels, the printed

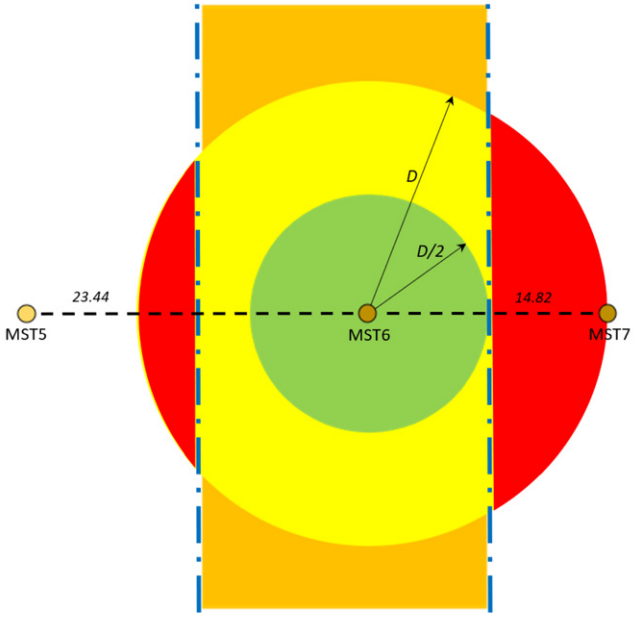

Figure 5. A conceptual two-dimensional CIELAB space for illustrating the acceptance criteria for matching MST6. Swatches inside the green circles are A-grade. Swatches in the yellow zone are B-grade. Swatches between the two blue lines (orange zone) are C-grade. Swatches outside the two blue lines (including the two red zones) are F-grade. See Section 3.2 for details.

MST swatch must match the intended MST level. In other words, among the ten MST levels, the intended level must be the closest to the printed swatch. Let $\Delta E(p1, p2)$ denote the color difference between two colors $p1$ and $p2$. This criterion is defined in Eq. (1) as the *C-criterion*.

$$\Delta E(S_i, MST_i) \leq \Delta E(S_i, MST_j), 1 \leq i, j \leq 10. \quad (1)$$

3.2.2 *Belongingness Criterion*

A printed MST swatch must be closer to the intended MST level than to any of the other nine MST levels. In other words, none of the other nine MST levels should be closer to the intended MST level than the printed MST swatch. This criterion ensures that the printed MST swatch is in the vicinity of the ten MST levels and therefore "belongs" to the MST scale. Otherwise, there is another MST level representing the intended MST level better than the printed MST swatch. This criterion is defined in Eq. (2) as the *B-criterion*.

$$\Delta E(S_i, MST_i) < \Delta E(MST_j, MST_i), 1 \leq i, j \leq 10, i \neq j. \quad (2)$$

The concept of the acceptance criteria is depicted in Figure 5 using MST6 as an example. Colors are drawn as points on a conceptual two-dimensional CIELAB space where the distance between two points represents their color difference. MST6 has two adjacent neighbors, MST5 and MST7, at distances 23.44 and 14.82, respectively, as noted in Table II. Two dotted blue lines are drawn as the boundaries at the midpoints between MST6 and its two adjacent neighbors.

- Following the C-criterion, a printed swatch intended to match MST6 should be located between the two blue lines. Otherwise, it will be classified as either MST5 or MST7. Since MST7 is closer to MST6 than MST5 (14.82 versus 23.44), the right boundary is more critical than the left one. A parameter D for MST6 is assigned with the shorter distance, $D = \min(23.44, 14.82) = 14.82$.
- Following the B-criterion, a printed swatch intended to match MST6 should be located inside the yellow circle of radius D. Otherwise, MST7 is closer to MST6 than the printed swatch. The distance between MST6 and the right boundary is $D/2 = 7.41$. Any color inside the green circle of radius $D/2$ satisfies both the B-criterion and C-criterion. The $D/2$ and D values for each MST level are listed in Table IV.

The consequence of violating the B- or C-criteria is elucidated in the following two examples. Consider two MST charts. Chart #1 has ten ideal, reference MST swatches. Chart #2 is the same as Chart #1 except that the reference MST6 swatch is replaced with a printed MST6 swatch.

- If the printed MST6 swatch violates the C-criterion and is located in the right red area in Fig. 5, a skin tone the same as the printed MST6 swatch will be classified as MST7 by Chart #1 but MST6 by Chart #2.
- If the printed MST6 swatch violates the B-criterion and is located in the top orange area in Fig. 5, a skin tone the same as the reference MST6 level will be classified as MST6 by Chart #1 but either MST5 or MST7 by Chart #2.

Areas partitioned by the two boundaries and two circles form five cases to grade a printed swatch as either A, B, C, or F, ranging from good to bad. The grading method is described in Algorithm 1 and summarized below:

Algorithm 1 Grading method

```
1:  for each printed swatch S_i for MST_i do
2:      if ΔE(S_i, MST_i) ≤ D(i)/2 then
3:          Grade(S_i) = A
4:      else
5:          if ΔE(S_i, MST_i) ≤ D(i) then
6:              if ΔE(S_i, MST_i) ≤ ΔE(S_i, MST_j), i ≠ j then
7:                  Grade(S_i) = B
8:              else
9:                  Grade(S_i) = F
10:             end if
11:         else
12:             if ΔE(S_i, MST_i) ≤ ΔE(S_i, MST_j), i ≠ j then
13:                 Grade(S_i) = C
14:             else
15:                 Grade(S_i) = F
16:             end if
17:         end if
18:     end if
19: end for
```

Table III. Grades and their necessary conditions.

Grade A	Meet both B-criterion and C-criterion; ΔE is bounded by $D/2$
Grade B	Meet both B-criterion and C-criterion; ΔE is bounded by D
Grade C	Meet C-criterion only; ΔE is greater than D
Grade F	Fail both B-criterion and C-criterion

Table IV. Distances to the closest boundary ($D/2$) and neighbor (D) for each MST level.

MST level i	$D(i)/2$	$D(i)$
1	1.37	2.74
2	1.37	2.74
3	3.28	6.56
4	3.28	6.56
5	5.73	11.47
6	7.41	14.82
7	6.91	13.82
8	6.91	13.82
9	3.50	7.00
10	3.50	7.00

- If a swatch is inside the green circle (line 2), it is graded as "A" since it exceeds the minimum requirements set by both the B-criterion and C-criterion. This condition is easier to check because only $\Delta E(S_i, MST_i)$ needs to be considered. It is suitable for an accurate MST chart with very small errors.
- If a swatch is outside the green circle but inside the yellow circle (line 5), there are two cases: (i) If the swatch is between the two boundaries (line 6) and inside the yellow area, it is graded as "B" because both the B-criterion and C-criterion are satisfied. This is a more relaxed condition than grade A. (ii) If the swatch is outside the boundaries and in the red areas, then it is graded as "F" (line 9).
- If a swatch is outside the yellow circle but between the two boundaries (line 12), it is graded as "C" because it satisfies the C-criterion but violates the B-criterion.
- Finally, a swatch is graded as "F" if it is outside the two boundaries (line 15).

Algorithm 1 specifies the sufficient conditions of the four grades. The necessary conditions of the four grades are listed in Table III. Figure 6 shows the relative green, yellow, and orange areas for the ten MST levels.

A printed MST chart, consisting of ten independent swatches, is graded based on the worst swatch with the the lowest grade. An A-grade MST chart has ten A-grade swatches. A B-grade MST chart has at least one B-grade swatch and probably some A-grade swatches, but no C-grade or F-grade swatches. Similarly, a C-grade MST chart has at least one C-grade swatch but no F-grade swatches.

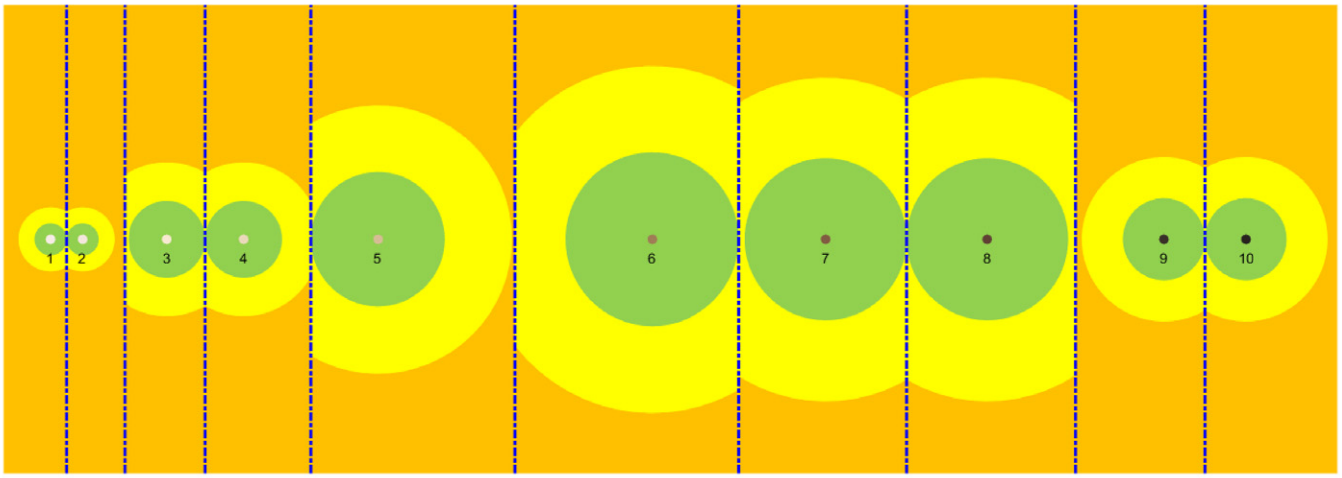

Figure 6. The to-scale green, yellow, and orange areas for the ten MST levels.

An F-grade MST chart has at least one F-grade swatch. Choosing either an A-graded, B-graded, or C-graded MST chart depends on the requirements of the user's application. However, an F-graded MST chart should not be used.

4. IMPLEMENTATION OF THE MST SCALE

4.1 *Materials*

The color charts were printed (see Figure 7) with a consumer-grade inkjet printer (Expression Photo XP970 Small-in-One, Epson) that used six inks: cyan, light cyan, magenta, light magenta, yellow, and black (Epson Claria Photo HD 277). The charts were printed on 4" × 6" glossy photo paper (Epson Ultra Premium Photo Paper Glossy) in a Windows 10 environment. Within the printer driver, the color correction option was disabled using the "no color adjustment" setting. The Microsoft Photos application was used to view and print the color charts. MATLAB code (MATLAB R2023a, MathWorks Inc., 2023) was written to generate the color charts in the PNG file format without embedded ICC color profile.

The printed color charts were measured with a spectrophotometer (CM700d, Konica Minolta, calibrated on September 8, 2023), which measured spectral reflectance between 400 nm and 700 nm at an interval of 10 nm. The lighting/viewing geometry was diffused illumination/8° viewing angle (de/8°). The CIELAB data reported by the spectrophotometer were based on CIE D50 illuminant and 2-degree observer. The CIE 1976 color difference formula (ΔE^*_{ab}) is used in this study and is annotated as ΔE. Code written in MATLAB code was used to send commands to the spectrophotometer through the USB serial port to trigger measurements and receive data.

A student-level computer numerical control (CNC) router (3018-Pro, Genmitsu) was modified to automate the measurement process. As shown in Fig. 7, the spectrophotometer was mounted to the spindle holder via a custom-made, 3D-printed adapter. Code written in MATLAB was used to send commands in the G-code language [32] for moving the spectrophotometer along the X, Y, and Z axes.

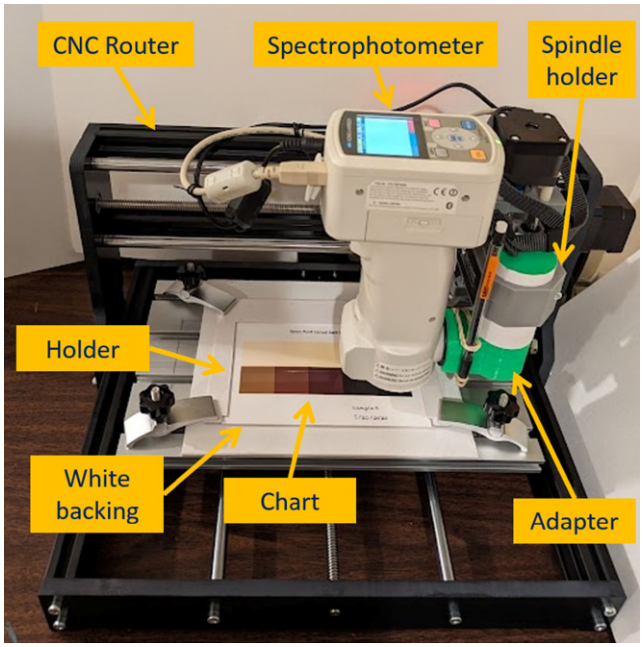

Figure 7. The measurement set-up. The spectrophotometer was mounted to a CNC router with an adapter. The chart was held down by a holder made of card boards and a white background.

A custom-made holder was used to hold the printed chart in place. Three pieces of the same blank photo paper were used as the white substrate to control the reflection from the background during the measurement.

4.2 *Method*

The printing process consisted of four stages: checking the color gamut of the printer, searching for the best-matching RGB values for the ten MST levels, validating the final printed chart, and quality control.

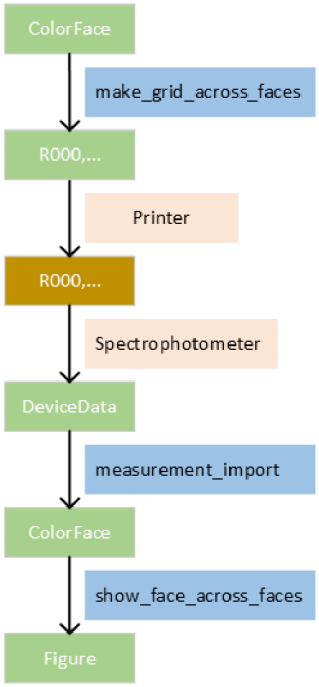

Figure 8. Flowchart for checking the color gamut. Green boxes represent software objects. The brown box represents a physical chart. Pink boxes represent external hardware devices. Blue boxes represent software modules.

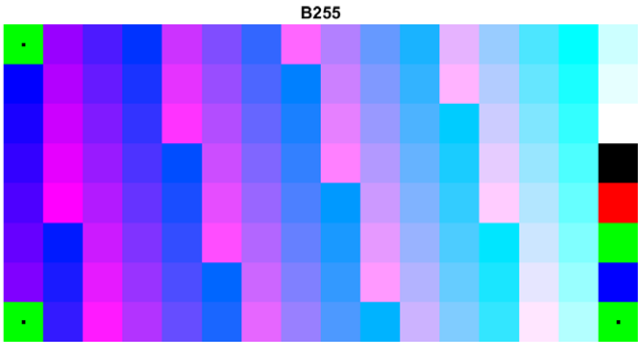

Figure 9. One of the six charts used to determine the color gamut.

4.2.1 *Color Gamut Check*

Since the MST scale spans a very wide range in the color space, it is important to confirm that the color gamut of the printer and paper combination is sufficient before proceeding. The method defined in the IEC 62977-2-1 standard [33] is used to calculate the color gamut envelope, which is a cube with six sides. The set of colors to measure consists of an evenly spaced 11×11 grid on each of the six sides of an RGB cube. In total, at least 602 ($6 \times 112 - 12 \times 11 + 8$) patches need to be measured.

Figure 8 shows the steps for calculating the color gamut envelope. The software tool generates and exports six charts for the printer to print (R000, R255, G000, G255, B000, and B255). Each chart characterizes one of the six sides of the color gamut envelope, as shown in Figure 9. The six charts are measured by the spectrophotometer driven by the CNC router. The measurement data are then fed back to the original data structure. Finally, the measurement data can be visualized in 3D to determine whether the printer is capable of printing the MST scale, as shown in Figure 10.

Each of the six charts includes 128 patches. For example, as shown in Fig. 9, the B255 chart contains 121 patches of $RGB = (\frac{255}{10}i, \frac{255}{10}j, 255)$, $0 \le i, j \le 10$. Three green patches, with a center dot, are placed on the corners for an automatic chart reader to position the patches. In the rightmost column, the black, red, green, and blue patches are included for consistency check. The white patch is included in the 121 patches as well as the background. When printed on the 4" × 6" paper, each patch occupies an area of 7.5 mm × 8.0 mm. Adjacent shades, as defined in CIELAB, are arranged in the same column to create larger color differences across columns when the chart is read horizontally. Measuring each of the six color gamut charts took approximately 18 minutes.

According to Fig. 10, MST1, MST2, MST9, and MST10 are enclosed by the color gamut but are very close to the boundary. If any of the ten MST levels are out of the color gamut, accurate reproduction of the MST scale will not be possible.

4.2.2 *Octal Search*

Similar to the traditional binary search method used to find a target number in a series of sorted numbers in one-dimensional space, an iterative *octal search* strategy (see Figure 11) is used to find the target color in three-dimensional color space. The octal search method picks one color as the starting point and constructs a $3 \times 3 \times 3$ grid with spacing K in the CIELAB space. The spacing K is a user-specified parameter for determining the search range, or the size of the $3 \times 3 \times 3$ grid. The CIELAB values are converted into sRGB values and sent to the printer. After measuring the 27 data points, one can decide whether the target color is contained within the grid and which of the 27 colors is closest to the target. If the closest color is an acceptable solution based on the B- or C-criteria, then the search can stop. Otherwise, the closest color will be used as the center to construct a new $3 \times 3 \times 3$ grid in the next iteration. The spacing of the grid can be adjusted depending on the residual errors and whether the target has been contained. Increasing the spacing expands the search range so that the target will be contained faster. Decreasing the spacing helps when the target is already contained but the error is too high.

Figure 12 shows the last iteration of an octal search. The octal search started with a grid of $3\Delta E$ and stopped at $2\Delta E$ when a solution was found. The $3 \times 3 \times 3$ grid with spacing of $2\Delta E$ appears distorted because the printer's RGB space is not sRGB. The target color, represented by the bigger sphere with a green outline, is already contained in the grid. The closest color is the bigger sphere with a blue outline. The residual error, which is equal to the distance between these two spheres, is less than $D(6)/2 = 7.41$, so the search can stop. The swatch will be A-grade.

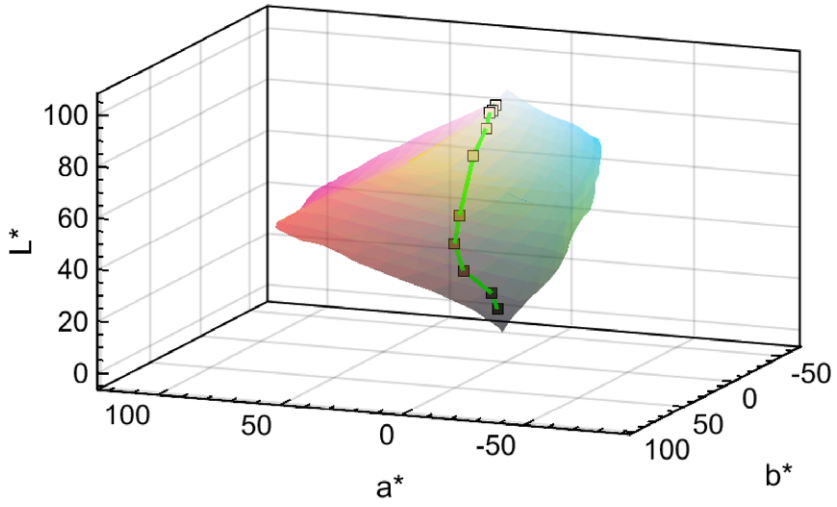

Figure 10. Comparison of the MST scale (brown boxes connected with green lines) with color gamut envelope of the Epson printer and glossy paper. MST1 is the brightest one on top.

Table V. An A-grade MST chart printed with Epson XP970 on glossy paper.

MST	Matched RGB	Measured CIELAB	ΔE^*_{ab}	ΔL^*	Δa^*	Δb^*	ΔC^*_{ab}	Δh_{ab}	ΔRGB	Grade
1	(254,251,230)	(93.60, 2.42, 5.95)	1.22	0.61	−0.92	−0.53	−0.80	6.64	16.25	A
2	(254,248,225)	(92.69, 2.68, 7.95)	1.00	−0.41	−0.62	−0.67	−0.82	2.82	21.12	A
3	(255,249,219)	(93.04, 1.93,13.21)	1.98	0.05	−1.71	0.99	0.86	7.44	20.25	A
4	(240,239,202)	(87.12, 1.58,17.74)	1.21	0.45	−1.12	0.01	−0.06	3.61	27.07	A
5	(229,214,175)	(77.37, 4.84,21.49)	2.21	0.53	−1.37	1.65	1.37	4.16	38.03	A
6	(191,158,114)	(55.83, 7.85,27.39)	0.95	−0.69	−0.07	−0.65	−0.64	−0.24	52.62	A
7	(163,115,093)	(42.95,12.43,21.08)	0.74	−0.48	−0.11	−0.55	−0.53	−0.45	47.90	A
8	(137,089,073)	(32.13,12.18,14.11)	1.72	−1.45	−0.51	−0.77	−0.92	−0.38	51.94	A
9	(090,076,055)	(21.76, 3.54, 6.27)	1.14	−0.69	−0.85	−0.31	−0.66	5.17	43.84	A
10	(056,061,035)	(15.73, 2.63, 3.97)	1.66	−1.12	−1.15	−0.45	−0.94	10.72	29.31	A

Design of the octal search charts is shown in Figure 13. For each MST level, the chart consists of 32 patches. Three patches in red, green, and blue are located on the three corners as markers for the automatic chart reader to position the patches. The black patch and the white background can be used to determine the black and white points, respectively. The RGB and black/white patches are included in all test charts for evaluating consistency of the printing process. Above the black patch is the color generated using the original RGB values recommended by Google (i.e., the sRGB values). The purpose of this patch is to show the outcome when color management is not applied. The remaining 27 patches are the $3 \times 3 \times 3$ grid spanned around the center color `[191 158 114]`. Also indicated in the caption are the target CIELAB values, the target MST level, and the spacing ΔE. Measuring each octal search chart took approximately 6 minutes.

The search results are shown in Table V. As indicated in the fourth column, the ΔE values range between 0.74 and 2.21, all less than the $D(i)/2$. Therefore, the printed chart is A-grade.

In the second rightmost column, the ΔRGB values represent the Euclidean distance between the sRGB values and matched RGB values. The differences range from 16.25 to 52.62. The pronounced differences indicate that using the sRGB value directly without color management to print the MST scale will not generate the desired shades.

4.2.3 *Print and Validate*

The best-matching RGB values are plugged into a pre-arranged template designed for the intended application. Besides the MST level and CIELAB data, the template includes a traceable serial number that includes information of the printer set-up, source data of the patches, print date, sequence number, etc. A URL is also included for providing additional documentation for using the software tool. The digital chart should be printed soon after characterization, to avoid the deviation of the printing environment from its

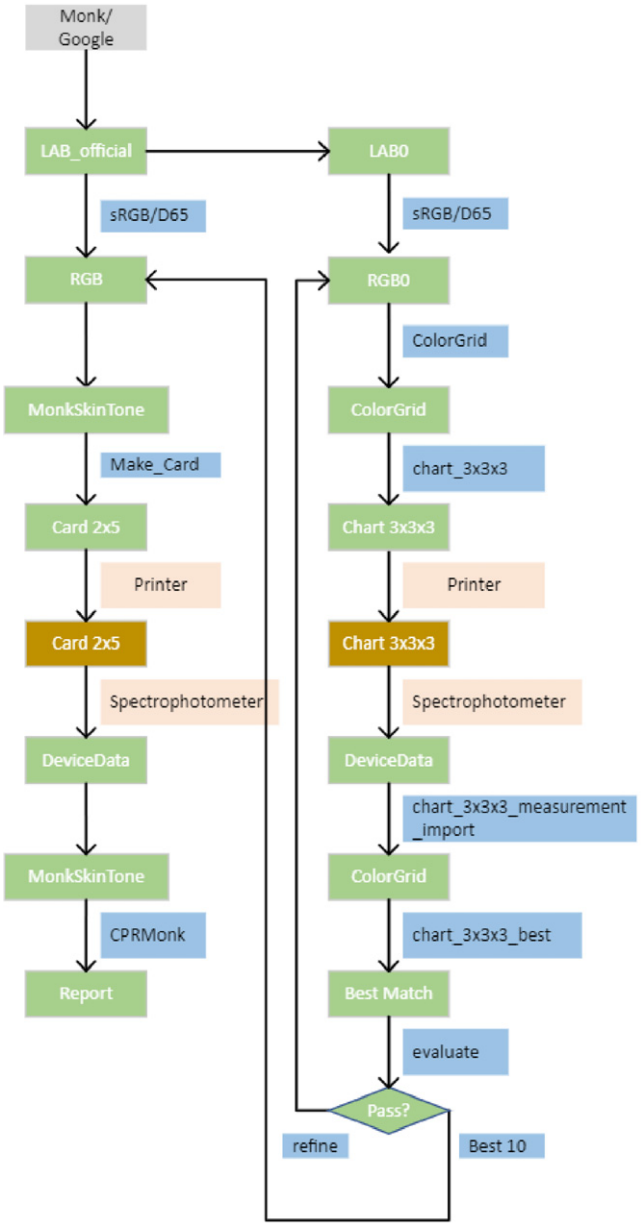

Figure 11. Flowchart of the octal search (right branch) and printing/validation (left branch).

previously-characterized status. After printing, the ink needs at least 24 hours to dry for the chart to achieve its intended, stable color.

The printed chart can then be measured with the same instrument to grade the ten MST swatches. Each swatch should be labeled with the grading result. A validated chart should be signed and dated by the tester for certification purposes.

4.2.4 *Quality Control*

When not in use, the printed chart should be properly stored according to the manufacturer's instructions (e.g., under appropriate temperature and humidity, light and chemical

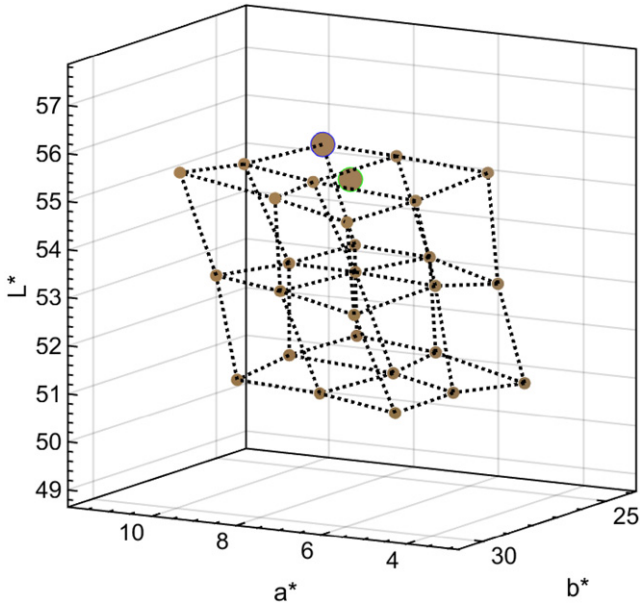

Figure 12. The 3 × 3 × 3 grid used to search the target shade (green-outlined sphere). The best match is highlighted with a blue circle.

proofing, etc.). Based on the lifetime of the ink and paper, the printed chart should be re-validated periodically.

5. CASE STUDIES

Three more printer/paper combinations were tested to evaluate the effects of the paper and printing technology. Matte paper was tested based on the impression that matte paper represents skin tones better due to its reflection properties. Dye sublimation is another popular printing technology for printing long-lasting photos. Color laser printers are ubiquitous and recommended in [21].

5.1 *Inkjet Printer with Matte Paper*

To compare glossy paper with matte paper, the same Epson XP970 inkjet printer was tested with matte paper (Photo Quality Ink Jet Cards, Epson). As shown in Figure 14, the matte paper has a smaller color gamut that cannot cover the MST scale, specifically MST1, MST9, and MST10. The evaluation results are shown in Table VI. The matched RGB values for the matte paper are different from those for the glossy paper, although the printer and inks are the same. Six patches are A-grade (MST2 through MST8) and two patches are B-grade (MST1 and MST9). MST10 is F-grade because it would be classified into MST9 ($\Delta E = 8.03$) instead of MST10 ($\Delta E = 8.74$). Therefore, this MST chart is F-grade and cannot be used.

5.2 *Dye Sublimation Printer*

The Canon SELPHY CP1500 is a consumer-grade portable photo printer based on the dye sublimation technology. It uses a single-use cartridge to transfer the cyan, magenta, and yellow (CMY) dyes onto the specific paper (RP-108 High-Capacity Color Ink/Paper Set, Canon). As shown in

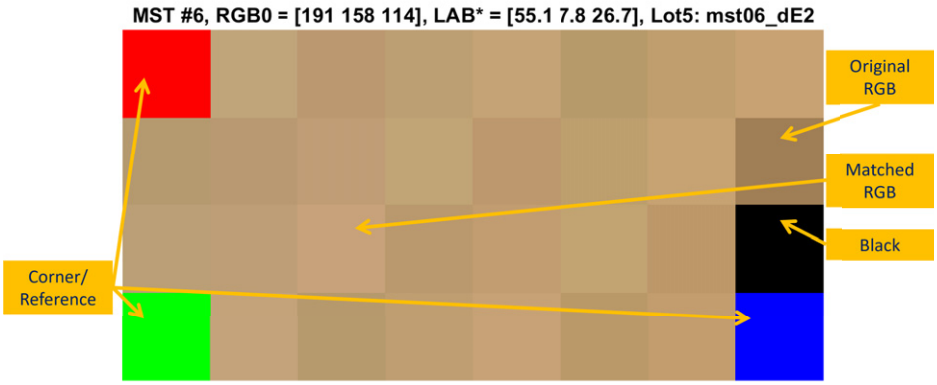

Figure 13. Image of the octal search chart for MST6 with spacing of $\Delta E = 2$.

Table VI. An F-grade MST chart printed with the Epson XP970 on matte paper.

Intended MST	Input RGB	Measured CIELAB	ΔE^*_{ab}	Grade	Incorrectly matched MST (ΔE^*_{ab})
1	(246,237,228)	(92.77, 0.55, 5.23)	1.74	B	–
2	(243,231,219)	(91.30, 2.60, 6.66)	1.27	A	–
3	(247,234,208)	(91.93,−0.42,13.95)	1.35	A	–
4	(234,218,186)	(85.83, 2.34,18.52)	2.68	A	–
5	(215,189,150)	(75.60, 3.78,22.49)	2.41	A	–
6	(160,126,086)	(53.36, 8.94,26.70)	2.13	A	–
7	(130,092,067)	(41.49, 9.62,20.37)	2.88	A	–
8	(096,065,052)	(32.76, 8.05,12.31)	4.30	A	–
9	(058,049,042)	(25.80, 1.52, 3.26)	5.57	B	–
10	(041,036,032)	(21.56,−1.16,−1.07)	8.74	F	**9 (8.03)**

Table VII. A B-grade MST chart printed with the Canon CP1500.

Intended MST	Input RGB	Measured CIELAB	ΔE^*_{ab}	Grade
1	(248,230,207)	(90.20, 2.25, 4.94)	4.11	B
2	(243,231,219)	(90.86, 1.72, 1.67)	5.80	B
3	(255,228,186)	(90.29,−0.20,13.76)	2.87	A
4	(218,212,174)	(85.19, 0.03,16.46)	2.74	A
5	(195,181,135)	(74.65, 3.14,21.43)	3.69	A
6	(147,114,065)	(51.86, 8.47,23.91)	4.39	A
7	(131,082,047)	(41.21,12.38,18.23)	2.62	A
8	(093,057,033)	(29.84,10.41,12.00)	2.02	A
9	(060,049,027)	(21.97, 2.18, 2.80)	3.33	A
10	(049,037,025)	(17.19, 2.38,−1.80)	5.98	B

Figure 15, compared with the sRGB, the color gamut is larger in the green direction ($-a^*$) but smaller in the yellow direction ($+b^*$). As a result, the color gamut cannot properly cover MST1, MST2, MST3, MST9, and MST10. The evaluation result is shown in Table VII. Seven patches (MST3 through MST9) are A-grade and the remaining three are B-grade. Therefore, this MST chart is B-grade.

5.3 *Laser/professional Printers without Color Management*

A set of new CIELAB values were proposed in [21] to replace the original MST scale based on the experiment results. The authors sent the raw sRGB values to different printers and took the average CIELAB values from five printers that were more expensive than USD $2,000. Analysis of the proposed CIELAB values is shown in Table VIII. Four patches (MST5

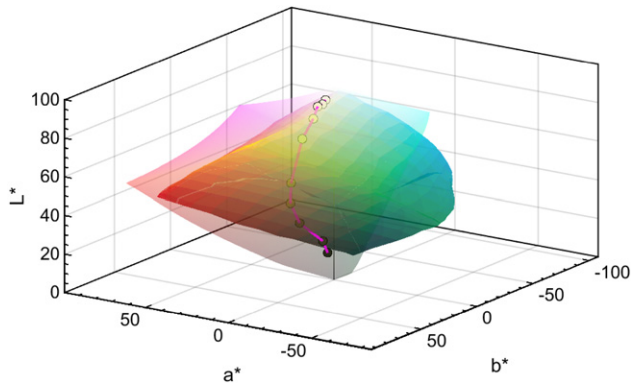

Figure 15. The MST scale (spheres connected by green lines) in comparison with the color gamut of the Canon CP1500 printer (inner) and sRGB (outer). MST1, MST2, MST3, MST9, and MST10 are out of the color gamut despite the green portion exceeding sRGB.

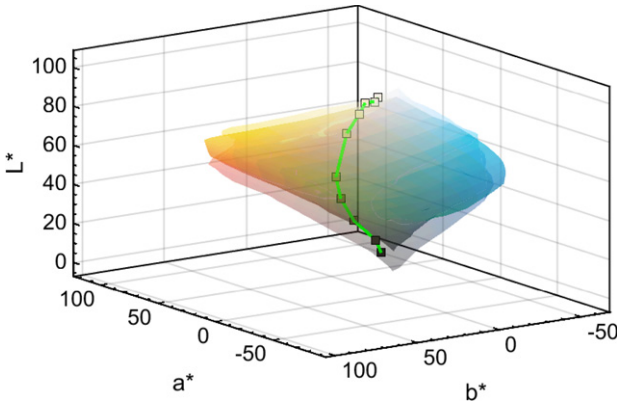

Figure 14. The reference MST scale (spheres connected by green lines) inside the color gamut of the Epson XP970 printer with glossy paper (outer) and matte paper (inner).

Table VIII. Evaluation result of the MST scale recommended in [21].

Intended MST	Measured CIELAB	ΔE^*_{ab}	Grade	Incorrectly matched MST (ΔE^*_{ab})
1	(88, 3,−1)	9.06	C	–
2	(86, 3, 1)	8.93	C	–
3	(86, 2, 7)	10.27	F	**2 (6.28)**
4	(81, 2, 10)	10.28	C	–
5	(73, 4, 18)	7.12	B	–
6	(53, 7, 21)	6.18	B	–
7	(43, 8, 14)	7.85	B	–
8	(33, 7, 8)	7.46	B	–
9	(27, 1, 2)	7.33	C	–
10	(24, 0, 0)	10.14	F	**9 (7.17)**

Table IX. An F-grade MST chart printed with the Canon imageRunner without color management.

Intended MST	Measured CIELAB	ΔE^*_{ab}	Grade	Incorrectly matched MST (ΔE^*_{ab})
1	(86.79, 7.50,−3.82)	13.28	C	–
2	(84.98, 7.39,−2.25)	13.13	C	–
3	(85.95, 8.15, 2.19)	16.07	F	**2 (10.15)**
4	(80.40, 7.77, 3.71)	17.38	F	**2 (13.65)**
5	(74.26, 8.64,12.08)	12.74	C	–
6	(60.40,12.40,21.30)	8.86	B	–
7	(49.62,16.56,17.10)	8.99	B	–
8	(39.23,14.04,11.42)	9.08	B	–
9	(32.48, 3.80, 4.13)	11.61	C	–
10	(28.53, 2.13, 0.34)	14.29	F	**9 (9.36)**

through MST8) are B-grade. Four patches (MST1, MST2, MST4, and MST9) are C-grade. MST3 and MST10 are F-grade because they would be incorrectly classified as MST2 ($\Delta E = 6.28$) and MST9 ($\Delta E = 7.17$), respectively.

To repeat the same experiment, we sent the raw sRGB values listed in Table I directly to a high-end color laser printer (Canon imageRunner Advance C7565 III, price > USD $3,000) loaded with regular copy paper (Hammermill, Copy Plus, 20 lb, 92 brightness). The evaluation result is shown in Table IX. Three patches (MST6 through MST8) are B-graded. Four patches (MST1, MST2, MST5, and MST9) are C-grade. MST3, MST4, and MST10 are F-grade because they would be incorrectly classified as MST2 ($\Delta E = 10.15$), MST2 ($\Delta E = 13.65$), and MST9 ($\Delta E = 9.36$). Therefore, this MST chart is F-grade and cannot be used.

The color gamut volumes of the four tested printers/paper are compared in Table X. All four color gamut volumes are much smaller than the sRGB color space. Thus, these printers are not sRGB-calibrated. For the same inkjet printer, the matte paper's color gamut was 17% smaller than the glossy paper. Although the dye sublimation printer has a larger color gamut volume, it could not cover the MST scale completely. Therefore, size of the color gamut is not the only factor to consider. Finally, the color gamut of the laser printer and plain paper is 52% smaller than that of the inkjet paper and glossy paper. The combination of laser printer and plain paper should not be recommended for printing accurate MST charts.

6. DISCUSSION

In medical applications, a high-grade MST chart should be used when possible. A low-grade (e.g., F-grade) MST chart may incorrectly classify a patient into the wrong group, possibly leading to inappropriate diagnosis and/or treatment.

In this study, the octal search stopped when the minimum requirements were met. Production of the optimal MST chart for a given printer/paper combination was not investigated because the objective was to find a valid MST chart rather than the best one. Nevertheless, this can be done by tightening the acceptance criteria and continuing the octal search until the result converges.

Table X. Comparison of color gamut volumes.

Printer and Paper	ΔE^3
Epson inkjet and glossy paper	515,428
Epson inkjet and matte paper	428,705
Canon dye sublimation	531,259
Canon laser and plain paper	251,232
sRGB	809,724

The ICC color profile approach was tested but not adopted in this manuscript for a few reasons: (1) the ICC color profile may not include the absolute colorimetric intent that is required to print accurate patches, (2) the ICC color profile is optimized for all colors rather than the ten specific MST patches, and (3) high-quality color printer profilers are not free and are expensive.

7. CONCLUSION

Without the right tool, manually making an accurate MST chart can be a tedious task that requires monastic dedication to accurate color reproduction. We presented a framework consisting of four key elements. For the ground truth, the Monk Skin Tone scale was colorimetrically defined and analyzed to show its unique features. For the figure of merit, the color matching task was procedurally described and modeled as a two-alternative forced choice experiment where the CIE color difference metric can be used to simulate an ideal observer. For the acceptance criteria, the per-patch B- and C-criteria were established based on whether a printed MST chart could be used to correctly classify the subject's skin tone by an ideal observer. A four-tier grading system was also proposed to grade the accuracy of a printed MST chart. Finally, for the implementation, a software tool was developed to create an accurate MST chart. The software measures 768 patches to determine the color gamut of the printer/paper combination, followed by measuring 320 or more patches to find the best input RGB values using the octal search strategy. Four printer/paper combinations were tested in the experiment. Results show that a wide color gamut covering the brightest and darkest MST levels is crucial for making a high-grade MST chart. This work is based on elementary colorimetry that is well known in the color research community. In addition to providing a practical tool for the general users outside the community, we hope, by introducing this problem to the community, more experts will contribute to solving real-world color problems in the medical field.

ACKNOWLEDGMENT

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services. This is a contribution of the U.S. Food and Drug Administration and is not subject to copyright.

REFERENCES

[1] R. R. Anderson and J. A. Parrish, "The optics of human skin," J. Investigative Dermatology **77**, 13–19 (1981).
[2] S. Del Bino and F. Bernerd, "Variations in skin colour and the biological consequences of ultraviolet radiation exposure," Br. J. Dermatology **169**, 33–40 (2013).
[3] M. W. Sjoding, R. P. Dickson, T. J. Iwashyna, S. E. Gay, and T. S. Valley, "Racial bias in pulse oximetry measurement," New England J. Med. **383**, 2477–2478 (2020).
[4] A.-K. I. Wong, M. Charpignon, H. Kim, C. Josef, A. A. De Hond, J. J. Fojas, A. Tabaie, X. Liu, E. Mireles-Cabodevila, L. Carvalho, and R. Kamaleswaran, "Analysis of discrepancies between pulse oximetry and arterial oxygen saturation measurements by race and ethnicity and association with organ dysfunction and mortality," JAMA Netw. Open **4**, e2131674 (2021).
[5] O. E. Okunlola, M. S. Lipnick, P. B. Batchelder, M. Bernstein, J. R. Feiner, and P. E. Bickler, "Pulse oximeter performance, racial inequity, and the work ahead," Respiratory Care **67**, 252–257 (2022).
[6] N. R. Henry, A. C. Hanson, P. J. Schulte, N. S. Warner, M. N. Manento, T. J. Weister, and M. A. Warner, "Disparities in hypoxemia detection by pulse oximetry across self-identified racial groups and associations with clinical outcomes," Crit. Care Med. **50**, 204–211 (2022).
[7] C. Shi, M. Goodall, J. Dumville, J. Hill, G. Norman, O. Hamer, A. Clegg, C. L. Watkins, G. Georgiou, A. Hodkinson, and C. E. Lightbody, "The accuracy of pulse oximetry in measuring oxygen saturation by levels of skin pigmentation: A systematic review and meta-analysis," BMC Med. **20**, 267 (2022).
[8] M. V. Plaisime, "Invited commentary: Undiagnosed and undertreated–the suffocating consequences of the use of racially biased medical devices during the covid-19 pandemic," Am. J. Epidemiology **192**, 714–719 (2023).
[9] R. Al-Halawani, P. H. Charlton, M. Qassem, and P. A. Kyriacou, "A review of the effect of skin pigmentation on pulse oximeter accuracy," Physiol. Meas. **44** (2023).
[10] A. K. Khanna, J. Beard, S. Lamminmäki, J. Närväinen, N. Antaki, and H. O. Yapici, "Assessment of skin pigmentation-related bias in pulse oximetry readings among adults," J. Clin. Monit. Comput. **38**, 113–120 (2024).
[11] S. Hao, K. Dempsey, J. Matos, C. E. Cox, V. Rotemberg, J. W. Gichoya, W. Kibbe, C. Hong, and I. Wong, "Utility of skin tone on pulse oximetry in critically ill patients: a prospective cohort study," medRxiv (2024).
[12] G. Leeb, I. Auchus, T. Law, P. Bickler, J. Feiner, S. Hashi, E. Monk, E. Igaga, M. Bernstein, Y. C. Chou, and C. Hughes, "The performance of 11 fingertip pulse oximeters during hypoxemia in healthy human participants with varied, quantified skin pigment," EBioMedicine **102** (2024).
[13] C. S. McCamy, H. Marcus, and J. G. Davidson, "A color-rendition chart," J. App. Photogr. Eng. **2**, 95–99 (1976).
[14] Pantone, "Munsell book of color, matte edition." https://www.pantone.com/munsell-book-of-color-matte-edition, Accessed: 2024-05-21 (2024).
[15] Pantone, "Pantone skintone guide." https://www.pantone.com/skintone. Accessed: 2024-06-04.
[16] M. Amani, H. Falk, O. D. Jensen, G. Vartdal, A. Aune, and F. Lindseth, "Color calibration on human skin images," *Int'l. Conf. on Computer Vision Systems* (Springer, Cham, 2019), pp. 211–223.
[17] A. Aune, G. Vartdal, G. J. Diaz, L. M. Gierman, H. Bergseng,, and E. Darj, "Iterative development, validation, and certification of a smartphone system to assess neonatal jaundice: Development and usability study," JMIR Pediatrics and Parenting **6**, e40463 (2023).
[18] Image Science Associates, "ColorGauge micro target," https://www.imagescienceassociates.com/colorgauge-micro-target.html. 2024. Accessed: 2024-07-30.
[19] W.-C. Cheng, "Color performance review (cpr): A color performance analyzer for endoscopy devices," J. Imaging Sci. Technol. **67**, 1–9 (2023).
[20] C. M. Heldreth, E. P. Monk, A. T. Clark, C. Schumann, X. Eyee, and S. Ricco, "Which skin tone measures are the most inclusive? an investigation of skin tone measures for artificial intelligence," ACM J. Responsible Comput. **1**, 1–21 (2024).
[21] W. Verkruijsse, A. Brancart, M. B. Jaffe, and S. Groen, "Variability of printed monk skin tone scales may cause misclassification of clinical

study participants: Caveats on printing," Anesthesia & Analgesia **138**, e43–e44 (2024).
[22] H. Zeng and M. Luo, "Skin color modeling of digital photographic images," J. Imaging Sci. Technol. **55**, 30201–1 (2011).
[23] K. Xiao, J. M. Yates, F. Zardawi, S. Sueeprasan, N. Liao, L. Gill, C. Li, and S. Wuerger, "Characterising the variations in ethnic skin colours: A new calibrated data base for human skin," Skin Research and Technology **23**, 21–29 (2017).
[24] Y. Wang, M. R. Luo, M. Wang, K. Xiao, and M. Pointer, "Spectrophotometric measurement of human skin colour," Color Res. Appl. **42**, 764–774 (2017).
[25] M. Wang, K. Xiao, M. R. Luo, M. Pointer, V. Cheung, and S. Wuerger, "An investigation into the variability of skin colour measurements," Color Res. Appl. **43**, 458–470 (2018).
[26] A. Chardon, I. Cretois, and C. Hourseau, "Skin colour typology and suntanning pathways," Int'l. J. Cosmetic Sci. **13**, 191–208 (1991).
[27] E. Monk, "Monk skin tone scale," (2019) https://skintone.google. Accessed: 2024-10-01.
[28] Datacolor, "Colorreader pro." https://www.datacolor.com/colorreader/products/colorreader-pro/. Accessed: 2024-06-04.
[29] X-Rite, Inc., "EyeOne pro." https://www.xrite.com/categories/calibration-profiling/i1pro-3. Accessed: 2024-05-21.
[30] International Electrotechnical Commission, "IEC 61966-2-1: Multimedia systems and equipment-Colour measurement and management-Part 2-1: Colour management-Default RGB colour space-sRGB", Standard IEC 61966-2-1, *International Electrotechnical Commission*, (1999).
[31] G. H. Joblove and D. Greenberg, "Color spaces for computer graphics," *Proc. 5th Ann. Conf. on Computer Graphics and Interactive Techniques* (ACM, New York, NY, 1978), pp. 20–25.
[32] GRBL, "GRBL - an open source, embedded, high performance g-code-parser and cnc milling controller." https://github.com/gnea/grbl. Accessed: 2024-05-21.
[33] International Electrotechnical Commission, "IEC 62977-2-1:2021 Electronic displays - Part 2-1: Measurements of optical characteristics - Fundamental measurements", Standard IEC 62977-2-1, *International Electrotechnical Commission*, (2021).