Federal Aviation Administration

DOT/FAA/AM-18/18
Office of Aerospace Medicine
Washington, DC 20591

Development of a Standard Palette for Color Coding ATC Displays

Kevin M. Gildea[1]
Nelda Milburn[1]
David L. Post[2]

[1]Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

[2] ϕ
Dayton OH 45324

October 2018

Final Report

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-18/18	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Development of a Standard Palette for Color Coding ATC Displays		5. Report Date October 2018
		6. Performing Organization Code
7. Author(s) Gildea KM, [1] Milburn N,[1] Post DL[2]		8. Performing Organization Report No.
9. Performing Organization Name and Address [1]FAA Civil Aerospace Medical Institute P.O. Box 25082, Oklahoma City, OK 73125 [2]ϕ 1333 Meadowlands Drive Dayton OH 45324		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplemental Notes Work was accomplished under approved task BHRR523		

16. Abstract

This project was motivated by the FAA's Air Traffic Organization (ATO) Program Management Organization's (PMO) research requirement to address the need to accommodate air traffic controllers who have color-vision deficiencies (CVDs) to allow them to perform their duties using the latest air traffic control (ATC) color displays. Color has become an integral part of ATC displays, helping controllers to discriminate, identify, locate, and use information efficiently. The color palettes used on contemporary ATC systems were not designed for CVD controllers. The current project provides a recommended color palette that reasonably accommodates the CVD personnel, allowing them to perform their jobs as well as normal color vision (NCV) controllers.

We developed a recommended color palette that provides a suitable set of colors that are proven to be discriminable, recognizable, conspicuous (i.e., easy to locate), and legible for CVD and color-normal controllers. The recommended palette meets these design objectives, and the human factors data and application of vision science provides the rational basis for it.

17. Key Words Air traffic control, color code, color set, conspicuity, deficient color vision, discriminability, display, legend, legibility, palette, recognizability, spectra, sRGB, standard, visual search, reasonable accommodation	18. Distribution Statement Document is available to the public through the Internet: http://www.faa.gov/go/oamtechreports/

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

We thank the following individuals, who contributed significantly to the project and the creation of this report. Pam Della Rocco served as the project technical sponsor in the Air Traffic Organization, Program Management Office. Thomas Chidester, Deputy Director of the FAA Civil Aerospace Medical Institute, provided valuable ideas and guidance that enhanced the project throughout. Kenneth Allendoerfer, Branch Manager, FAA Human Factors Branch, contributed to the Standard Terminal Automation Replacement System (STARS), En Route Automation Modernization (ERAM), and Ocean21 display measurements and provided the tech refresh STARS spectra. Edmundo Sierra, an FAA Scientific and Technical Advisor, collected ambient illumination measurements at the Washington DC and Memphis Centers (ZDC and ZME) that determined the illumination level used in our experiments. William E. Goode, Goode Systems, adapted our color calibration software to run under Windows 7 and wrote our data-collection program. Seve Benincasa, Human Solutions, Inc., contributed to the STARS, ERAM, and Ocean21 display measurements and provided the tech refresh ERAM spectra. Daniel Jack and Shijing Liu, Cherokee CRC, provided statistical support and created all the associated figures. Contract support was provided through the Office of Research & Sponsored Programs at Wright State University. The grant was facilitated by Jennie J. Gallimore of Wright State University's Department of Biomedical, Industrial and Human Factors Engineering.

The research reported herein was conducted under the Air Traffic Program Directive/Level of Effort Agreement between the Human Factors Division (ANG-C1), FAA Headquarters, and the Aerospace Human Factors Research Division of the FAA Civil Aerospace Medical Institute.

TABLE OF CONTENTS

DEVELOPMENT OF A STANDARD PALETTE FOR COLOR CODING ATC DISPLAYS

INTRODUCTION

Color has become an integral part of air traffic control (ATC) displays as they have evolved from monochromatic radar displays to full-color, complex interfaces that require operators to discriminate, identify, and locate many colors to make effective use of the information displayed. That information must be legible too, regardless of its color. In accordance with the Rehabilitation Act of 1973, the FAA and the Civil Aerospace Medical Institute (CAMI) screen normal color-vision (NCV) and color-vision deficient (CVD) ATC candidates to determine if their color vision is sufficient to perform mission tasks. The Air Traffic Color Vision Test (ATCOV; Chidester et al., 2011; Chidester et al., 2013) was developed by CAMI to determine which CVD candidates have sufficient color vision to complete mission tasks on current ATC systems and to screen out those who do not. Medical screening is the province of Aviation Medicine, under the Federal Air Surgeon.

The color palettes used on the En Route Automation Modernization (ERAM), Standard Terminal Automation Replacement System (STARS), and Ocean21 ATC systems were not designed to accommodate CVD controllers. We developed a recommended color palette that provides a suitable set of colors that are proven to be discriminable, recognizable, conspicuous (i.e., easy to locate), and legible for CVD and color-normal controllers. The recommended palette meets these design objectives, and the human factors data and application of vision science provides the rational basis for it.

We conducted human factors experiments to develop a standard palette for color coding critical information[1] on non-tower primary ATC displays that meets the aforementioned requirements. We focused on the windowless ATC viewing environments because ambient illumination levels in towers vary dramatically (see Wilson, Wilson, & Jha, 2007, for measures), thereby complicating palette design significantly. The palette accommodates NCV viewers and most CVD viewers. The palette (documented in Appendix A) contains 11 foreground colors for coding symbols, alphanumerics, and other objects on the display, plus 4 background colors for coding weather severity. The colors are specified in terms of their Commission Internationale de l'Eclairage (CIE) normalized luminances and chromaticity coordinates, as well as their corresponding computer-industry standard red, green, and blue (sRGB) values per IEC (1999). The use of sRGB values will facilitate accurate reproduction of the standard palette in the field.

The project was motivated by the FAA's Air Traffic Organization (ATO) Program Management Organization's (PMO) research requirement to address the need to accommodate air traffic controllers who have color-vision deficiencies (CVDs), to allow them to perform their

[1] Critical information is that which is essential to an operator's ability to accomplish the task in accordance with system or mission requirements, thereby avoiding adverse effects on system safety, effectiveness, efficiency, and reliability (FAA HF-STD-010, 2017).

duties using the latest air traffic control (ATC) color displays. A display color standard was developed (FAA HF-STD-010, 2017) and includes a standard color palette for coding critical information.

We developed the standard palette over the course of two pilot studies. The first used subjective assessments by NVC and CVD participants to choose an initial color set. The second refined that set by measuring NCV and CVD participants' visual search performance and adjusting the colors to improve performance. We concluded with a formal experiment. Its main purpose was to compare NCV versus CVD search performance, in order to judge the palette's suitability as an FAA standard. We also tested for a benefit of displaying a legend that shows all the palette colors and for an effect of a display's RGB spectra on CVD search performance. We required that all the palettes we tested satisfy FAA HF-STD-010 (2017). We assessed the standard's quantitative requirements using a purpose-built program named *Palette Designer* (Post & Goode, 2017). The results for the FAA standard palette are documented in Appendix B.

Why might a display's RGB spectra matter?

The CIE system of photometry and colorimetry uses spectral sensitivity functions that describe normal color vision. Any two spectra that yield identical luminances and chromaticity coordinates when they are weighted by the CIE functions will produce colors that look the same to NCV viewers, even if the spectra differ. The CIE system does not predict color matches accurately for CVD viewers, however, because their spectral sensitivity functions are abnormal. Consequently, differing spectra that produce matching colors for NCV viewers may not match for CVD viewers and vice versa.

This limitation of the CIE system of photometry and colorimetry poses a problem for designing a standard palette that is satisfactory for CVD viewers, regardless of the display in use. Liquid-crystal displays (LCDs), which are the most common display type presently, use differing backlight technologies, which produce differing spectra. Further, those spectra are modified by the color filters that form the R, G, and B subpixels on the display, and those filters often vary from one display model to another. Thus, two different display models having identical backlights can produce different RGB spectra because the color filters differ and two displays having identical color filters can differ because the backlights produce different RGB spectra. Thus, a set of luminances and chromaticity coordinates that is acceptable for CVD viewers on one display may not be acceptable on another, and knowledge of the backlight technology alone cannot predict the outcome.

A fool's errand?

The notion of a color set that CVD viewers can use might seem nonsensical. Protans and deutans have difficulty detecting red-green differences and tritans have difficulty detecting yellow-blue differences. If we eliminate all need to make those discriminations, we are left with nothing but luminance as a color-coding dimension, and even that is problematic because the CIE 1924 spectral weighting function for luminance is inaccurate for protans.

The fact that no color set can be usable in its entirety by all CVD viewers does not mean they cannot benefit from color coding. Most CVD viewers are able to discriminate, recognize, locate, and read at least some of the colors in a well-chosen set. They need a redundant code, however, to obtain information carried by color differences they cannot detect reliably. This is why FAA HF-STD-010 (2017) requires a redundant coding dimension for color-coded critical information. A well-chosen color set for our purposes, then, is one that works well for NCV viewers and contains subsets that work well for the three main types of CVD.

PILOT STUDY 1

Our first pilot study was designed to yield a color set that was promising enough to merit objective testing. We used subjective assessments because they allowed us to test, adjust, and down-select many color candidates quickly.

Participants

Participants in the first pilot study were recruited by a contracted staffing agency from the Oklahoma City greater metropolitan area. Their near and far visual acuity was tested using Snellen charts and was required to be 20/30 or better, with correction if necessary. We used the Color Assessment and Diagnosis test (CAD; Barbur, Rodriguez-Carmona, Evans, & Milburn, 2009) to identify each participant's color-vision category. The CAD is a computerized color vision test that screens for normal color vision, quantifies loss of chromatic sensitivity, and classifies individuals by type and degree of color vision deficiency. Finally, the participants completed informed consent forms (see Appendix C).

We formed four groups, each consisting of one color-normal, one protan (i.e., red weak), one deutan (i.e., green weak), and one tritan (i.e., blue weak). We found only one tritan, so that person was a member of all four groups. (Tritans are rare, constituting less than 0.002% of the population per Wyszecki & Stiles, 1982, p. 464.) Otherwise, each group was unique.

Apparatus

The stimuli were presented on a 22-inch diagonal Dell model 2210 LCD that was connected to a personal computer (PC) running Windows 7. We used the 2210's native resolution, which is 1680 x1050 pixels at a 60-Hz refresh rate. The 2210's color gamut is slightly larger than sRGB's; thus, it can produce any color that is available on an sRGB-compliant display.

Initial foreground-color selection.

Our initial set of foreground colors was the 25 colors shown in Figure 1 (see page 5).[2] Derefeldt and Swartling (1995) found that all 25 colors were recognized reliably by color normals when they were presented as large (roughly 3 x 3-degree) squares on a gray background. The sRGB standard (IEC, 1999) specifies a peak luminance of 80 cd/m^2, which is approximately the lowest peak luminance we encountered when we sampled ERAM, STARS, and Ocean21 displays at an FAA William J. Hughes Technical Center ATC simulation facility. Consequently, we set our

[2] Color figures in this document may not look exactly as intended because of differences in color rendering among displays. Discrepancies are even more likely if paper copies are printed.

white's luminance to 80 cd/m^2 and scaled all other colors' luminances proportionally to that maximum value.

Initial weather-color selection.

Weather severity is depicted on ATC displays using moving shapes that represent storms. We wanted seven weather colors so all seven FAA weather-severity levels (0 to 6) could be represented by a unique color. We chose black to represent level 0 because 0 is the most common severity level, black provides maximum contrast for the foreground objects, and air traffic controllers are accustomed to seeing level 0 represented by a black or dark blue background. The six remaining colors ranged from a desaturated green to a desaturated red to provide an intuitive depiction of increasingly severe weather. They all had low luminances, chosen to provide at least a 3:1 contrast ratio for all foreground colors, as required by FAA HF-STD-010 (2017). Using low-saturation, low-luminance background colors has the advantage of reducing the effects of simultaneous color contrast, that is, the effect of a background color on a foreground color's appearance.

Color calibration.

The Palette Designer (PD) program computes the RGB values needed to produce desired colors on a specific display, given a characterization file for that display. The computations implement the Piecewise Linear interpolation assuming Variable Chromaticity coordinates (PLVC) method described by Post and Calhoun (1989, 2000). If a secondary display is connected to the computer that is running PD and a characterization file is provided, a color-swatch chart like that shown in Figure 1 will be displayed there, using the calculated RGB values so the user can see a colorimetrically accurate rendition of the current color set. Post and Calhoun (2000) found that PLVC yielded an average error of 0.44 and 95th-percentile error of 0.85 distance units in CIELAB space. It is therefore likely that the colors PD produced on our secondary display matched the intended luminances and chromaticity coordinates with sufficient accuracy for purposes of our pilot study.

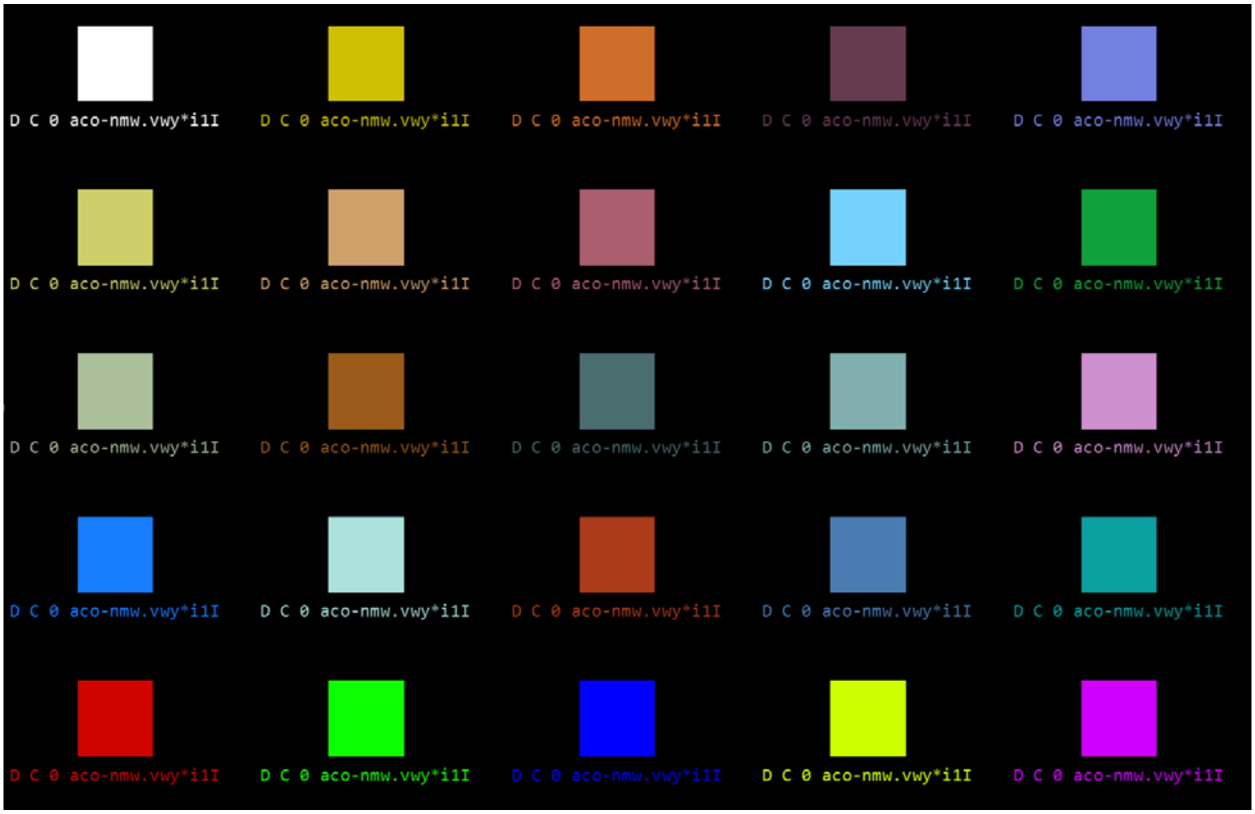

Figure 1. Twenty-five reliably recognized colors from Derefeldt and Swartling (1995).

We produced a characterization file for the Dell 2210 display by measuring 16 evenly spaced values ranging from 0 to 255 for each of the display's R, G, and B channels, using a Photo Research model PR-740 spectroradiometer and custom software. The testing room's lights were turned off and its door was shut during the measurement process. The 2210 served as PD's secondary display, so its characterization file was used by PD's PLVC implementation to calculate the RGB values needed to produce the luminances and chromaticity coordinates we needed.

Ambient illuminance.

The testing room's overhead lighting was adjusted to produce 40 lux of ambient illumination on the display's screen during data collection, as measured by a Photo Research model 524 illumination meter. The 40-lux figure is slightly higher than the upper 99% confidence interval (38 lux) for measurements taken in two FAA en route facilities (E. A. Sierra, personal communications, January 28 and February 19, 2014). Our intent was to test under the worst-case amount of ambient illumination apt to occur in the field.

Procedure

The LCD was placed on a table with four chairs arrayed in front of it. The participants were encouraged to change their viewing distances and rotate the display horizontally so it faced them directly, as needed, while evaluating the color-swatch chart shown on the display.

Group 1 examined all 25 foreground-color swatches and matching character strings, presented against all seven weather colors, initially. They also examined all six non-black weather colors against a black background, presented in another swatch chart. Group 1's CVD participants identified many color pairs that were indistinguishable for one or more of them. Guided by the group's feedback and colorimetric figures of merit PD computed and displayed, the experimenter adjusted problematic colors, using PD's user interface, and eliminated some when it became apparent that adjusting a color to make it distinguishable from a second made it indistinguishable from a third for at least one group member. The experimenter also ensured all colors remained within the sRGB color gamut.

Group 2 examined the color set chosen by Group 1, plus a few extra colors the experimenter thought promising. Group 2 performed the same process as Group 1, resulting in elimination of some colors and adjustment of others. Group 3 examined the set chosen by Group 2, eliminated one more color, and wanted minor adjustments of a few others. Group 4 examined the set chosen by Group 3 and pronounced the set fine as is. We measured the final color set on the display, using the PR-740 spectroradiometer to determine each color's final luminance and chromaticity coordinates as accurately as possible.

Results

We ended with the set of 11 foreground colors shown in Figure 2 and the weather colors shown in Figure 3. The participants deemed the foreground colors discriminable, recognizable, conspicuous, and legible in all foreground/weather-color combinations. They deemed the weather colors discriminable, recognizable, and conspicuous.

Figure 2. The 11 foreground colors that resulted from Pilot Study 1.

Figure 3. The six non-black weather colors that resulted from Pilot Study 1.

PILOT STUDY 2

The second pilot study tested the foreground and weather colors resulting from the first pilot study using objective performance measures. We anticipated that the subjective assessments would not predict objective performance reliably and, consequently, further changes to the colors would probably yield a better set. Our experimental procedure was similar with the response-surface exploration we used for Pilot Study 1.

Participants

Participants were recruited by the same personnel agency as in Pilot Study 1. They were screened for visual acuity and color-vision deficiencies and then signed informed consent forms. The group included 54 color normals, 12 protans, and 25 deutans (no tritans were available), including 1 normal, 2 protans, and all 4 deutans from Pilot Study 1.

Apparatus

Data collection took place in a large, windowless, light-controlled room. The room was equipped with 19 workstations, each of which included a PC and display, allowing us to test multiple participants at once. The displays consisted of 22-inch diagonal Dell model 2208, 2210, and 2213 LCDs. We ran all of them at their native 1680 x 1050-pixel resolution at a 60-Hz refresh rate. All three models have color gamuts that are slightly larger than sRGB's. White was set to 80 cd/m^2 and all other colors' luminances were scaled proportionally to that maximum on all displays.

The foreground and weather colors were calibrated on each display to tolerances of +/- 2.5% in luminance and distance error ≤ 0.0025 on the CIE 1976 u′v′-chromaticity diagram, using the PR-740 spectroradiometer and custom software. That software uses the measure-and-adjust algorithm introduced by Post and Calhoun (1989) to achieve those tolerances. The room's lights were turned off and its doors were shut during the calibration process.

After the calibrations were complete, we adjusted the room's ambient illumination by positioning free-standing lamps so that 40 lux was produced at each display's screen without creating specular glare for the viewers.

Procedure

We used an iterative approach to refine the foreground and weather colors, similar to the one we used for the subjective assessments. After running a small group of participants that included at least one color normal, one protan, and one deutan, we examined the data and modified colors for which CVD performance was clearly inferior to the color normals. We repeated the process until it seemed no substantial performance improvements were likely.

We used two search tasks to evaluate performance for the color set. The tasks are modeled after ones that are used in the ATCOV and are designed to test the colors' discriminability, recognizability, and conspicuity. The ATCOV is an occupational color-vision test, used to determine whether ATC candidates who have a color-vision deficiency can nonetheless perform adequately with the color sets used by ERAM, STARS, and Ocean21. CVD candidates must pass the ATCOV for at least one ATC system to become controllers. We randomized the order in which participants performed the two tasks, each of which was preceded by practice trials.

For each trial, participants responded to stimuli using the PC's mouse and we recorded two dependent measures: (1) An accuracy score, computed as the percentage of correct target identifications minus the percentage of incorrect identifications (i.e., false alarms); and (2) The time required to complete each trial. These are the same performance measures the ATCOV records. We used the accuracy score as a measure of the colors' discriminability and recognizability. We used response time as a measure of conspicuity.

Weather-color search (see Figure 4).

Each trial of the weather-color search task was preceded by presentation of a target cue, consisting of a small, solid-colored square having one of the weather colors. That cueing presentation identified the target color for the upcoming search. After the participant clicked a "Go" button on the screen, a 6 x 8 array of small, solid-colored squares was presented. Each of those squares was assigned one of the weather colors randomly, subject to the constraint that at least one used the target color. The participant's task was to use the computer mouse to click on each target square and then click a "Next" button, signaling the end of the trial and causing presentation of the next target cue.

Each square subtended 6 arc-minutes visually at a 50-cm viewing distance because the ATCOV uses that size to represent the smallest weather stimuli that appear on ATC displays (Chidester et al., 2011, p. 5). Participant viewing distance was not controlled, however. Instead, like the ATCOV, the participants were allowed to change viewing distance freely, just as air traffic controllers do. Each square was centered within a larger, solid-colored 1-degree square that used one of the other weather colors (chosen randomly) and served as a background. Each of the 7 colors appeared once as a target against the other 6 colors, yielding 42 trials/session.

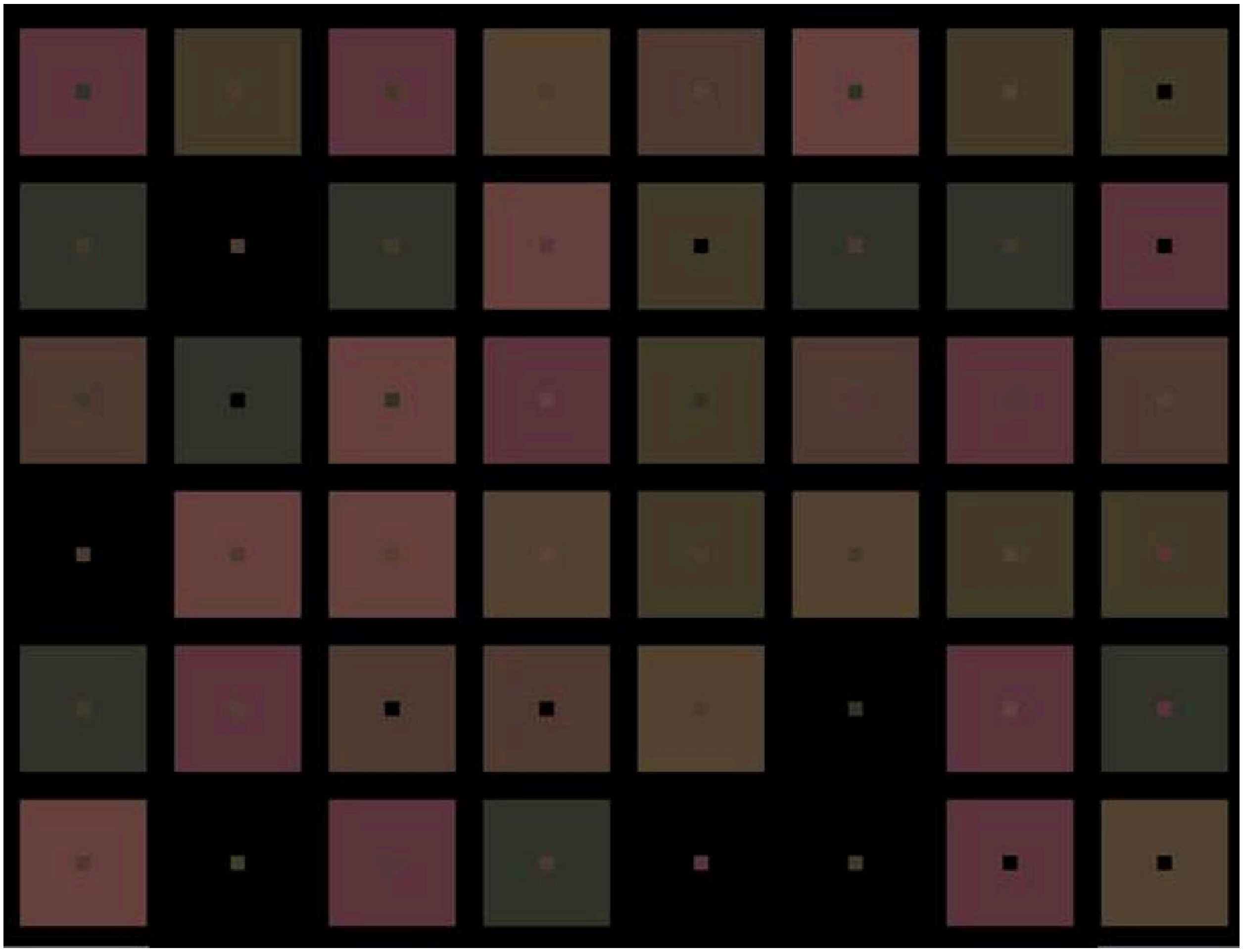

Figure 4. Example of a weather-color search trial.

Foreground-color search (see Figure 5)**.**

Each trial of the foreground-color search task was preceded by presentation of a target cue, consisting of a monochrome ATC datablock, using one of the foreground colors. That cueing presentation identified the target color for the upcoming search. A 6 x 6 array of datablocks was presented next, after the participant clicked a “Go” button. Those datablocks used the same character strings as the cueing presentation and were assigned one of the foreground colors randomly, subject to the constraint that at least one used the target color. Each of the 11 foreground colors was the target color once for each of the 4 weather colors, yielding a total of 44 trials per session. The participant’s task was to use the computer mouse to click on each target datablock and then click a “Next” button, signaling the end of the trial and causing presentation of the next target cue. The datablocks were rendered in the Consolas font, which is a sans serif font that resembles the ones used on ERAM, STARS, and Ocean21. The characters subtended 20 arc-minutes visually at a 50-cm viewing distance, as required by FAA HF-STD-010 (2017).

Figure 5. Example of a foreground-color search trial.

Results

Weather colors.

We found that the weather colors' low luminance and saturation made it difficult for CVD participants to discriminate, recognize, and locate all seven colors. We therefore eliminated three and adjusted the remaining ones, leaving black, green, yellow, and red, as shown in Figure 6. That set is sufficient for ERAM, which shows only four weather severities. For STARS, a second coding dimension, such as patterning, would be needed to depict all seven severity levels uniquely.

Figure 6. The three non-black weather colors that resulted from Pilot Study 2.

It occurred to us that presenting the weather colors on randomly chosen backgrounds made the search task unrealistically difficult. The severity of a storm is usually greatest at its center; therefore, if a storm (or part of one) is depicted as a single square, it is most likely surrounded by squares denoting the next lower weather severity. This means that if a small square is noticed its color can be inferred from the background's color. Consequently, we modified the weather-color search task for our main experiment by constraining the background colors as shown in Table 1.

Table 1. Target and background weather-color pairings used in the main experiment.

Target square	Background square
Wx-Green	Black
Wx-Yellow	Wx-Green
Wx-Red	Wx-Yellow

For the modified weather-color task, each of the 3 possible target/background combinations was used as the target 14 times, yielding a total of 42 trials per session.

Foreground colors.

We found that the participants – especially CVD ones – had difficulty distinguishing the two yellows and two greens in the foreground-color set. We therefore replaced one of the yellows and one of the greens with an aqua and a brown and adjusted some of the remaining colors. These changes left us with a foreground-color set that nearly matches the 11 "focal" color names identified by Berlin and Kay (1969) and Crawford (1982) and studied later by Boynton and his colleagues (e.g., Boynton & Olson, 1987; Smallman & Boynton, 1993). Our final foreground-color set consisted of a red, pink, orange, yellow, brown, green, blue, purple, black, white, and gray, as shown in Figure 7. The focal colors, on the other hand, include black, which we reserved for coding weather severity zero and therefore could not use as a foreground color, and do not include an aqua, although we found that search performance for our aqua was acceptable. Colorimetric specifications and graphical depictions of the final color set are provided in Appendix A.

We conclude that our subjective assessment method produced a good starting set of colors, but that set was a local minimum rather than a global one. That is, small adjustments to the Pilot Study 1 set either produced no improvement or degraded performance because large changes were needed to converge to a better set.

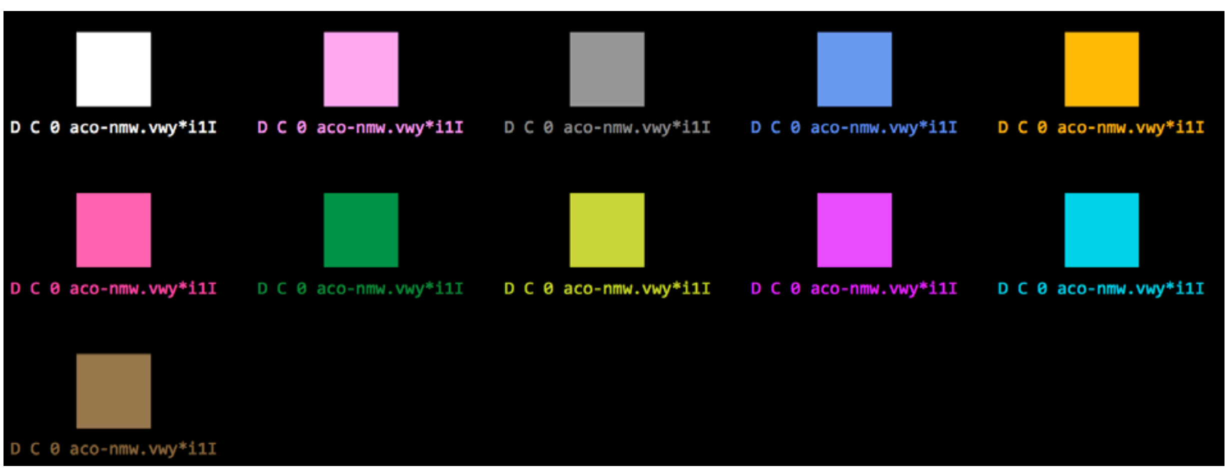

Figure 7. The 11 foreground colors that resulted from Pilot Study 2.

MAIN EXPERIMENT

Participants

Participants were recruited and screened as before. The group included 155 people (140 males) ranging from 20 to 55 years in age (average = 29 years) and consisted of 103 color normals, 20 protans, 31 deutans, and the tritan from Pilot Study 1, including 2 protans and 4 deutans who also participated in Pilot Study 1. Overlap with Pilot Study 2 consisted of 7 protans and 11 deutans. One of the deutans and two of the protans did not pass the ATCOV for any ATC system. Consequently, they would have been rejected if they had sought to become controllers. We retained them because we wanted to see how well our palette accommodates CVD viewers who are rejected presently by the ATCOV.

Histograms of our NCV and CVD participants' CAD scores are shown in Figures 8 and 9. They are very similar to corresponding figures shown in Chidester et al. (2013, p. 5). We conclude that our participants' color vision is similar to those from Chidester et al. (2013).

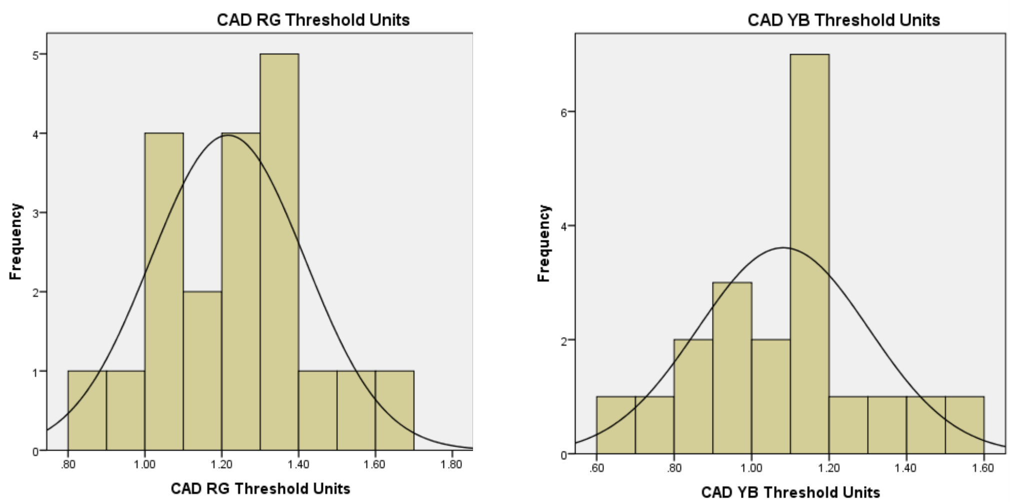

Figure 8. CAD RG and YB threshold scores for our NVC participants.

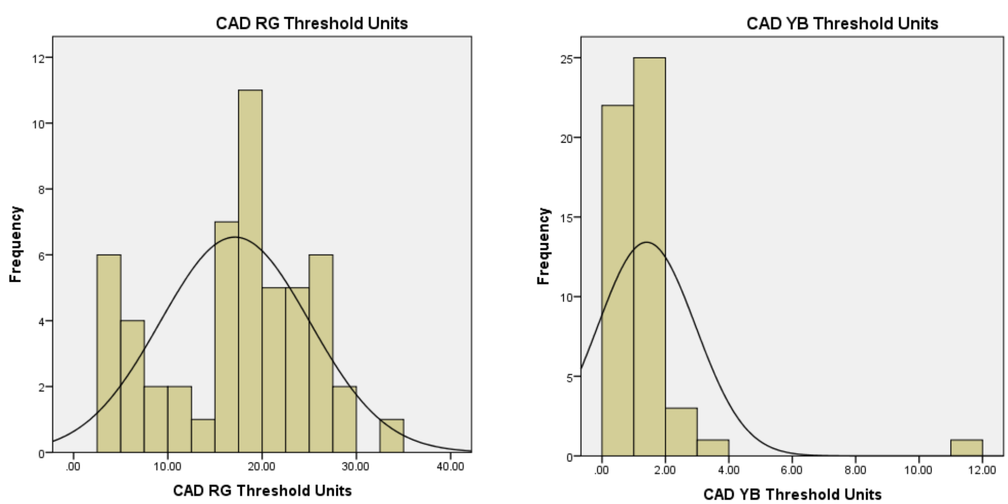

Figure 9. CAD RG and YB threshold scores for our CVD participants

Apparatus

The apparatus and data-collection room were the same as for Pilot Study 2, but we added a Dell model 2408 LCD so we could assess the effects of the displays' RGB spectra. The 2408 is a 24-inch diagonal display with a color gamut that is larger than sRGB's. We used its native 1920 x 1200-pixel resolution at a 60-Hz refresh rate. As before, white was set to 80 cd/m^2 and all other colors' luminances were scaled proportionally to that maximum on all displays. The foreground and weather colors were calibrated on each display to tolerances of +/- 2.5% in luminance and distance error ≤ 0.0025 on the CIE 1976 u′v′-chromaticity diagram. Ambient illumination on the display screens was then set to 40 lux.

Our decision to test for RGB-spectra effects by comparing Dell 2210 and 2408 displays was based on spectral analyses. Specifically, we evaluated differences in RGB spectra for the displays used presently by ERAM, STARS, and Ocean21, the new STARS and ERAM displays that have been selected during their recent tech refreshes, and displays we had available for testing. Our procedure (suggested by J. Neitz, personal communication, Nov 11, 2015) focused on protans because their luminosity functions differ substantially from color normals, deutans, and tritans. The procedure was: (1) Measure each display's R, G, and B spectra; (2) Normalize the spectra so their CIE luminances are equal; (3) Convolve each spectrum with a protan luminosity function derived from physiological data; (4) Integrate the results to get each spectrum's protan luminance; and (5) Form ratios to determine which two displays produce the largest difference in protan luminance.

The largest difference we found between ATC displays was for Ocean21 versus the current ERAM display: The ERAM R primary is 34% more luminous for protans than the Ocean21 R primary. The next largest difference is between Ocean21 and the new ERAM displays: The new ERAM R primary is 16% more luminous for protans. The displays in our inventory that came nearest to matching those differences were Dell model 2210 and 2408 LCDs. The protan luminance of the 2210's R primary is 21% greater than the 2408's R primary. The 2408 produces the same RGB spectra as the 30-inch diagonal Dell model 3007 LCD, which is the Ocean21 primary display. We established their equivalence by measuring 2408 and 3007 RGB spectra, using the PR-740 spectroradiometer, and comparing them.

Procedure

Our main experiment compared the performance of NCV versus CVD viewers when performing search tasks involving the palette resulting from Pilot Study 2. All participants performed the (modified) weather-color and foreground-color search tasks from Pilot Study 2, plus two more in random order. The two additional tasks were as follows.

Shape search (see Figure 10)**.**

The shape search task tested the foreground colors' legibility. Each trial was preceded by presentation of a three-character alphanumeric string, using white characters. That cueing presentation identified the target string for the upcoming search. After the participant clicked a "Go" button, a 6 x 6 array of datablocks was presented next. At least one of the datablocks began with the target string. Each of the 11 foreground colors was the datablock color once for each of the 4 weather colors, yielding a total of 44 trials per session. The participant's task was to use the computer mouse to click on each target datablock and then click a "Next" button, signaling the end of the trial and causing presentation of the next target cue.

Figure 10. Example of a shape search trial.

Redundant coding search (see Figure 11)**.**

The redundantly coded search task used color and shape to encode targets, thereby satisfying an FAA HF-STD-010 (2017) requirement that color-coded critical information be coded redundantly. Each trial was preceded by presentation of a three-character alphanumeric string using one of the foreground colors. After the participant clicked a "Go" button, a 6 x 6 array of datablocks was presented next. At least one of the datablocks began with the target string and used the same foreground color as the cuing stimulus. Participants could, thus, locate targets by recognizing the string or by spotting the datablock color. Each of the 11 foreground colors was the target color once for each of the 4 weather colors, yielding a total of 44 trials per session. The participant's task was to use the computer mouse to click on each target datablock and then click a "Next" button, signaling the end of the trial and causing presentation of the next target cue. Performance on this and the weather-color search task was our primary basis for deciding whether CVD performance with the palette justifies making it an FAA standard.

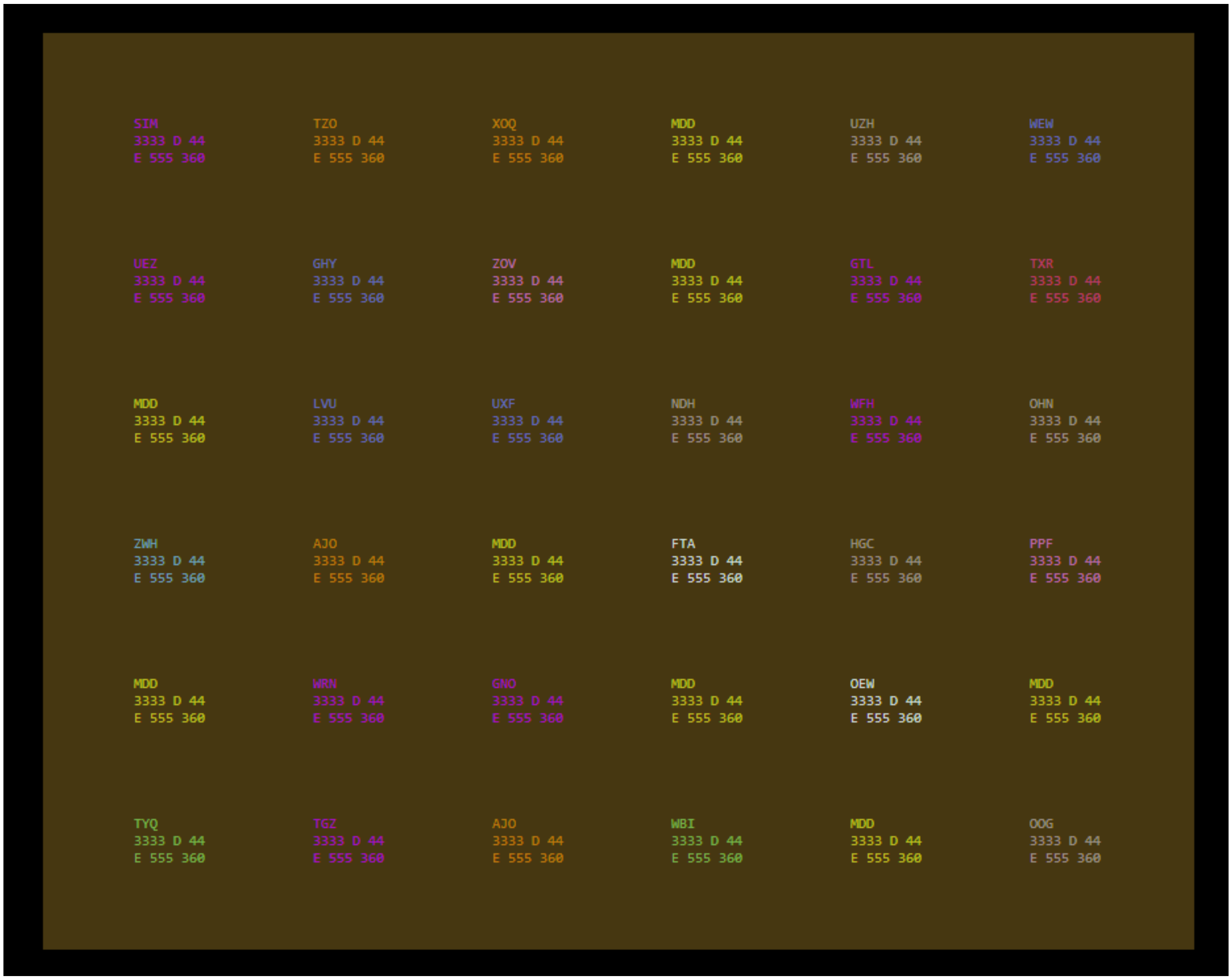

Figure 11. Example of a redundantly coded (color + shape) search trial.

Effect of a legend (see Figure 12)**.**

We tested the possibility that displaying a legend, showing all the palette colors, would improve search performance. We thought that having an ever-present reference might help participants – especially CVD ones – discriminate and recognize the colors. Toward this end, all participants performed the weather-color, foreground-color, and redundant search tasks with and without having a legend present during trials. Presentation order of the tasks and legend conditions was randomized.

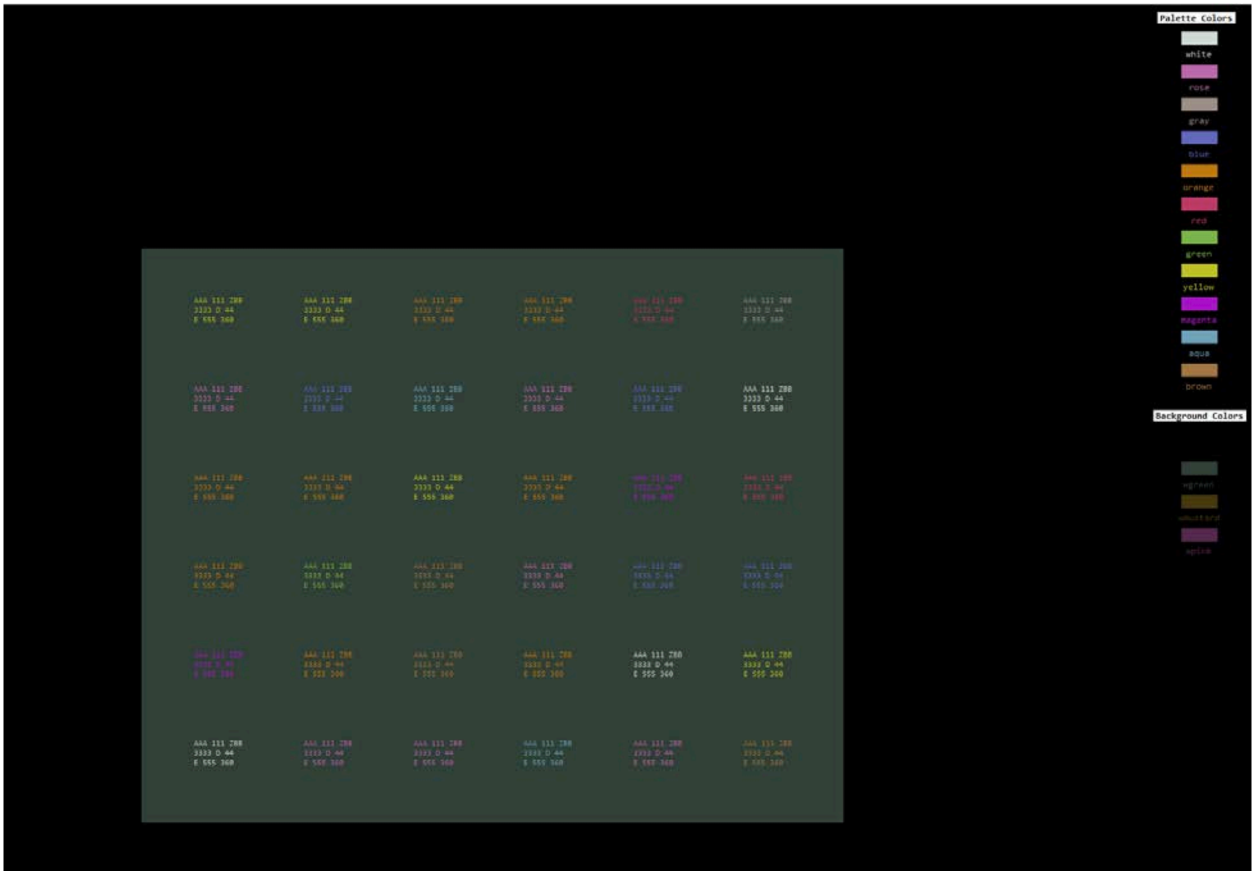

Figure 12. Example of a foreground-color search trial with the legend present.

Results

Data distributions.

Figures 13 – 18 are histograms and corresponding boxplots for our color-normal participants' search performance. The figures plot average scores, response times, and response speeds (i.e., inverse response time, or responses/s) and are representative of figures for the other CVD diagnoses.

Average score shows a ceiling effect that produced negative skew. These features are apparent in the ATCOV histograms presented by Chidester et al. (2011, 2013), also. Average response time exhibits positive skew. Its inverse is more symmetric, so we used response speed ("Speed") rather than response time as our temporal performance measure. All the boxplots include observations that lie beyond the "whiskers," the lengths of which denote ±1.5 times the interquartile range. We regarded those observations as outliers and omitted them from analysis.

Power considerations.

The crux of our main experiment was to determine whether CVD performance with the standard color palette compares acceptably with NCV performance. If that criterion is met, the palette is suited for use as an FAA standard. Meeting that criterion implies failing to reject null hypotheses, though, which poses a hazard: If the statistical power for a comparison is low, we might take a resulting failure to reject the null as evidence the palette is acceptable when it is not;

that is, we might commit a Type II error. Fielding a flawed palette might be worse than not fielding a good one; therefore, to boost power and reduce the likelihood of Type II error, we set alpha equal to 0.1 and used one-tailed tests where possible when we compared NCV and CVD performance. Otherwise, we set alpha equal to 0.05 and used two-tailed tests.

We used Faul, Erdfelder, Lang, and Buchner's (2007) *G*Power* program to compute power. We computed achieved power for average score ("Score") by setting the effect size to 1 divided by the standard deviation of the differences. We chose 1 as the critical difference because the ATCOV uses integer criteria and rounds average scores to the nearest integer when reporting whether a candidate passes or fails, so a difference of 1 can determine whether a candidate passes.

The ATCOV Alert Detection task uses a 2-s time limit. The rationale for choosing that value as a critical time difference is presented in Chidester et al. (2011, p. 26). We adopted the same criterion for judging the adequacy of our participants' temporal performance.

Evaluating the tritan.

We could not use a single participant as a unique level of color-vision diagnosis ("Diagnosis") in our ANOVAs. We therefore evaluated the tritan's performance separately, by comparing his means with confidence intervals for the color normals.

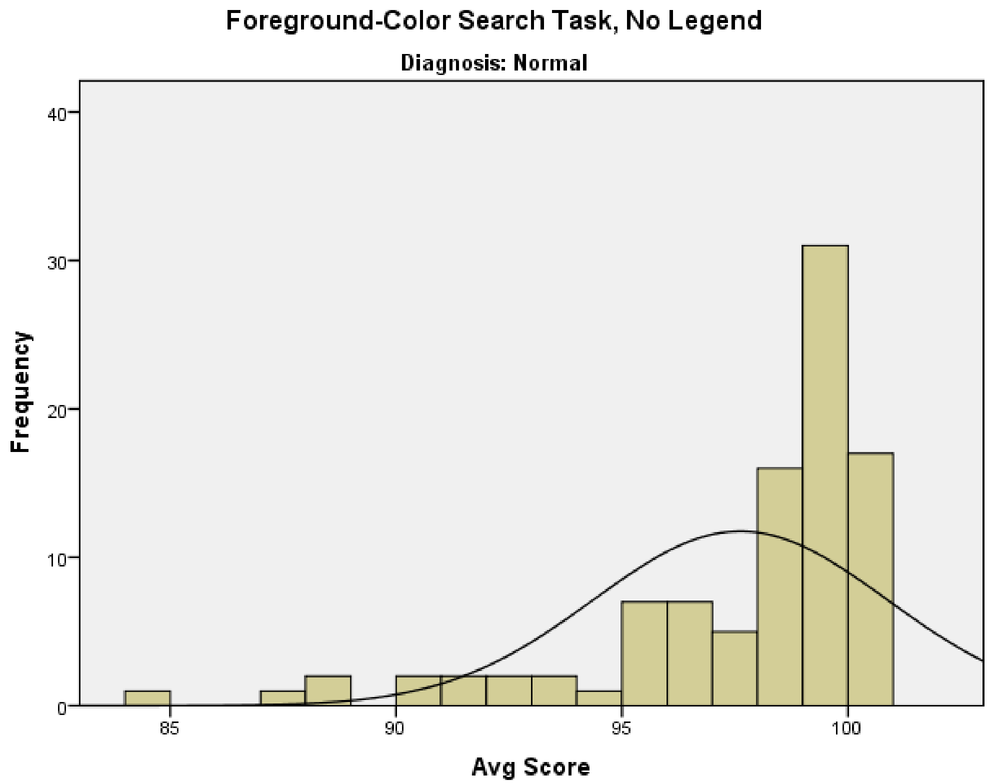

Figure 13. Histogram of color normals' average scores on the foreground-color search task without a legend present.

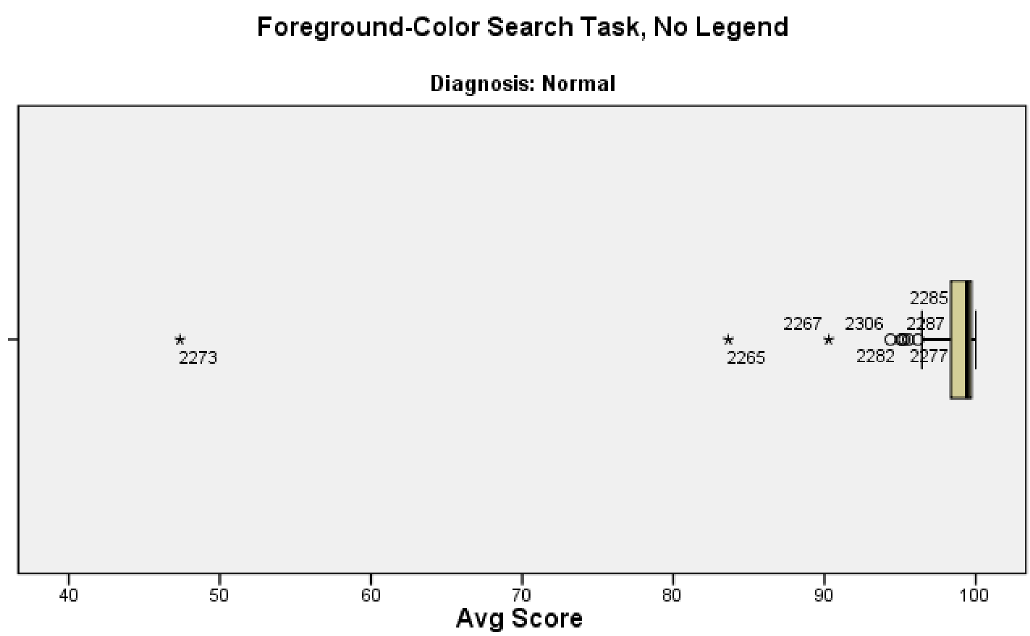

Figure 14. Boxplot of color normals' average scores on the foreground-color search task without a legend present. Numbered points are outliers.

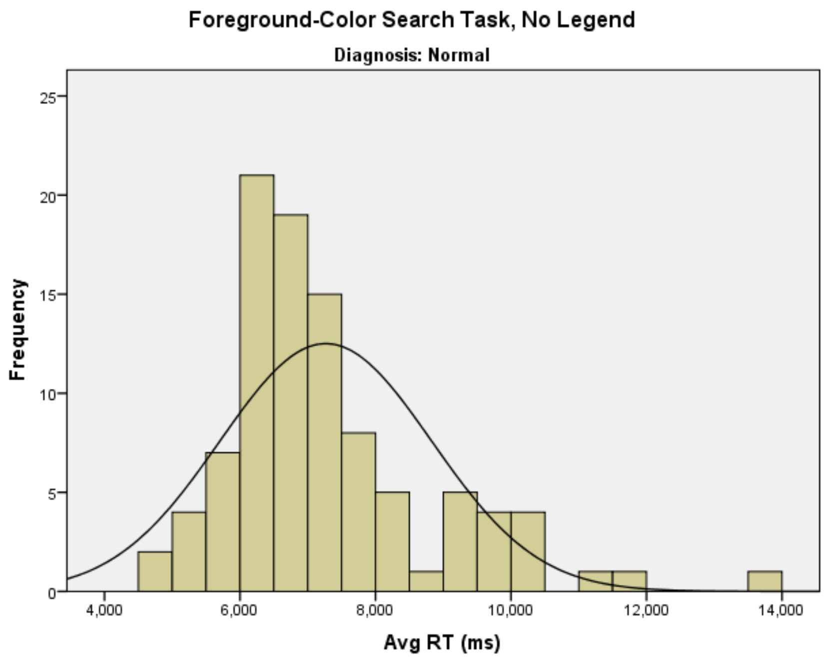

Figure 15. Histogram of color normals' average response times on the foreground-color search task without a legend present.

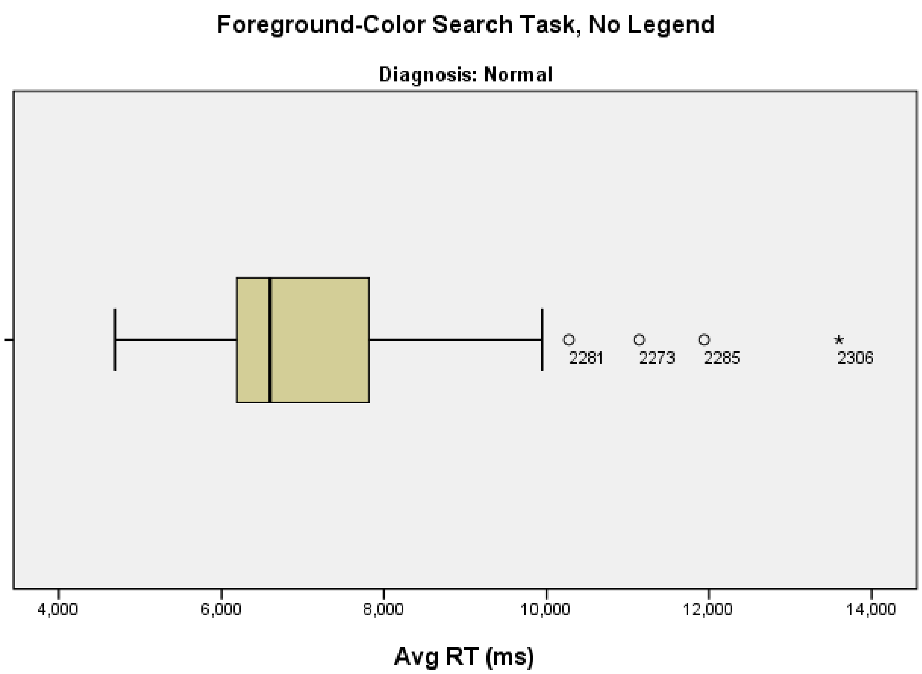

Figure 16. Boxplot of color normals' average response times on the foreground-color search task without a legend present. Numbered points are outliers.

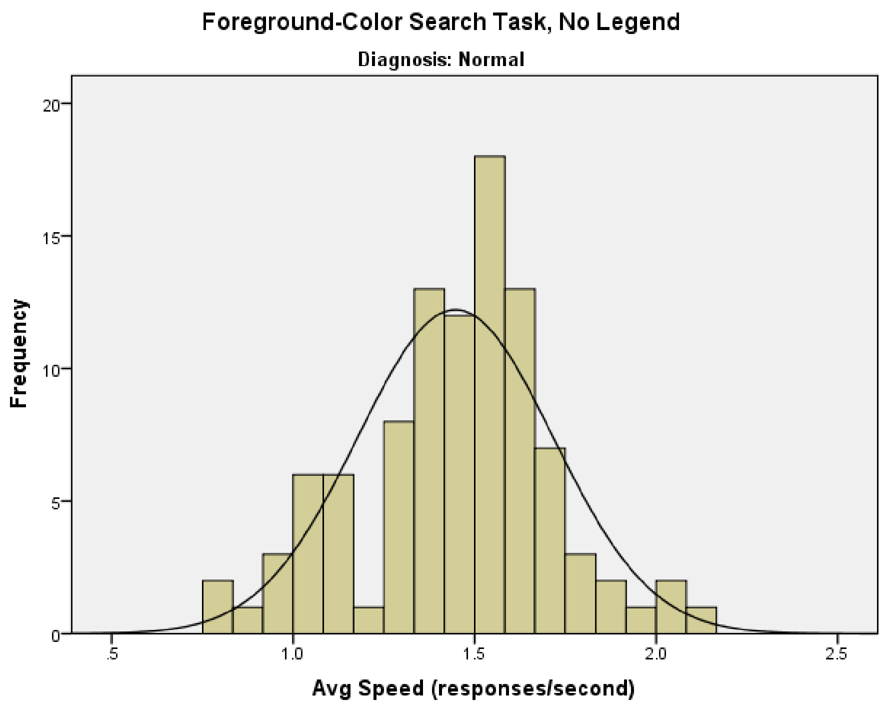

Figure 17. Histogram of color normals' average response speeds on the foreground-color search task without a legend present.

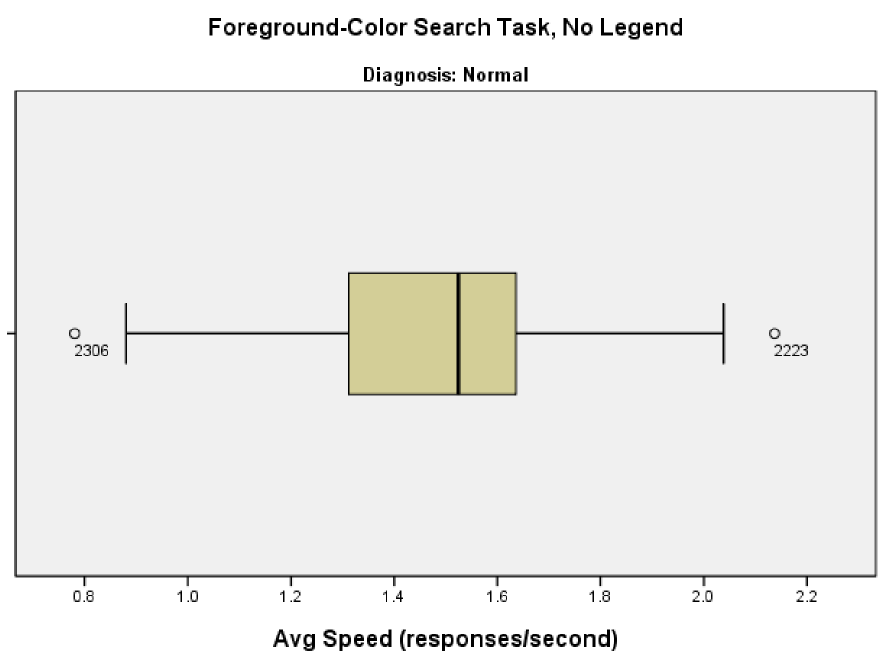

Figure 18. Boxplot of color normals' average response speeds on the foreground-color search task without a legend present. Numbered points are outliers.

Effects of legend.

The participants performed the weather-color, foreground-color, and redundant searches with and without having a legend present. We tested for effects involving Legend using three-way ANOVAs with Legend (present vs. absent), Task, and Diagnosis (normal, protan, or deutan) as the main effects. The Legend x Task interaction for Speed (see Figure 19) is significant, $p = 0.044$, but accounts for only 0.14% of the variance. Fisher's least significant difference (LSD) post-hoc paired-comparison test shows that the interaction consists solely of a slight benefit (+0.11 responses/s, or -18 ms/response) of a legend for the weather-color search task ($p < 0.05$) but not the other tasks. A difference of 18 ms is only 0.9% of the 2-s critical difference we adopted for response time. We conclude that adding the legend produced no important effects on search performance for stimuli that were color-coded using the standard palette. Consequently, we averaged over the Legend condition to produce a more stable dataset for the main analysis.

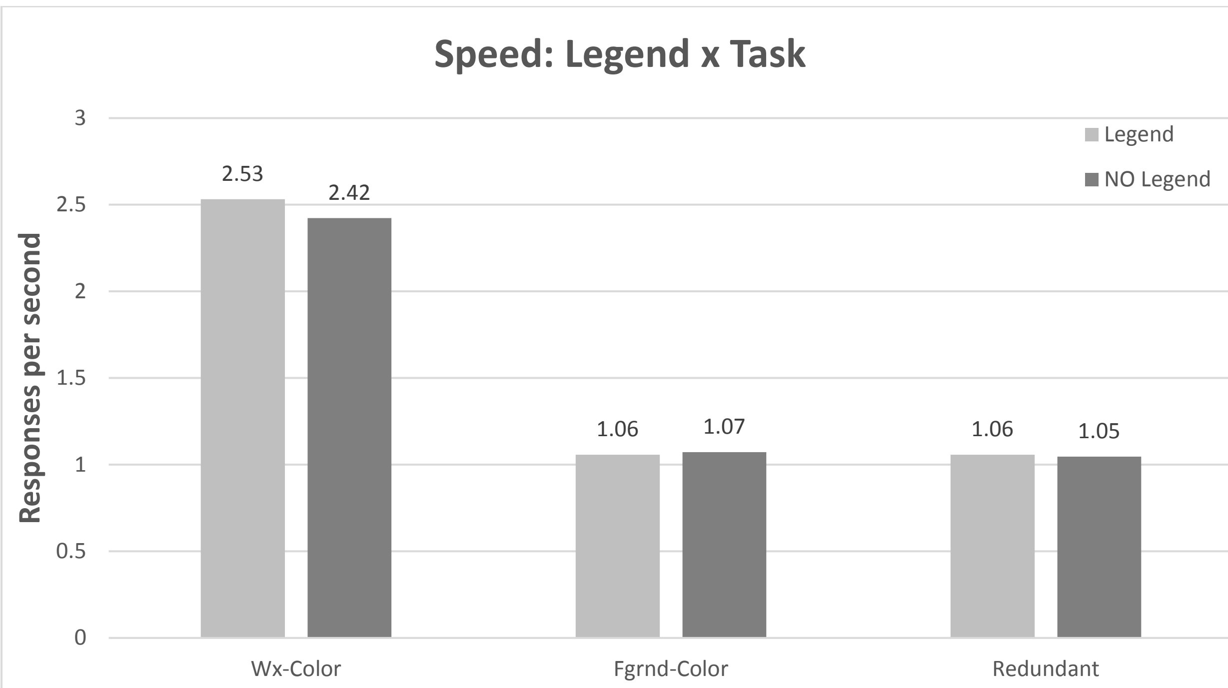

Figure 19. Legend x Task interaction for Speed.

Effects of display RGB spectra.

Eighteen of our protans performed all four search tasks in random order on Dell 2408 and 2210 displays. We tested for effects involving Display using two-way ANOVAs with Display and Task as the main effects. The main effect of Display and its interaction with Task on Score and Speed are not significant, $p \geq 0.105$. Power to detect a difference of 1 in Score is 0.99 for the main effect of Display. Power to detect a difference of 1 in Score for the interaction (for which $p = 0.105$), though, is only 0.13, and Figure 20 shows a sizable effect magnitude ($\Delta = 3.2$) for the Foreground-Color search. The statistical evidence for that effect is weak, so we averaged over Display condition for the protans to produce a more stable dataset for the main analysis. Our results suggest, though, that further exploration of the possible effect of display RGB spectra on CVD visual performance is warranted.

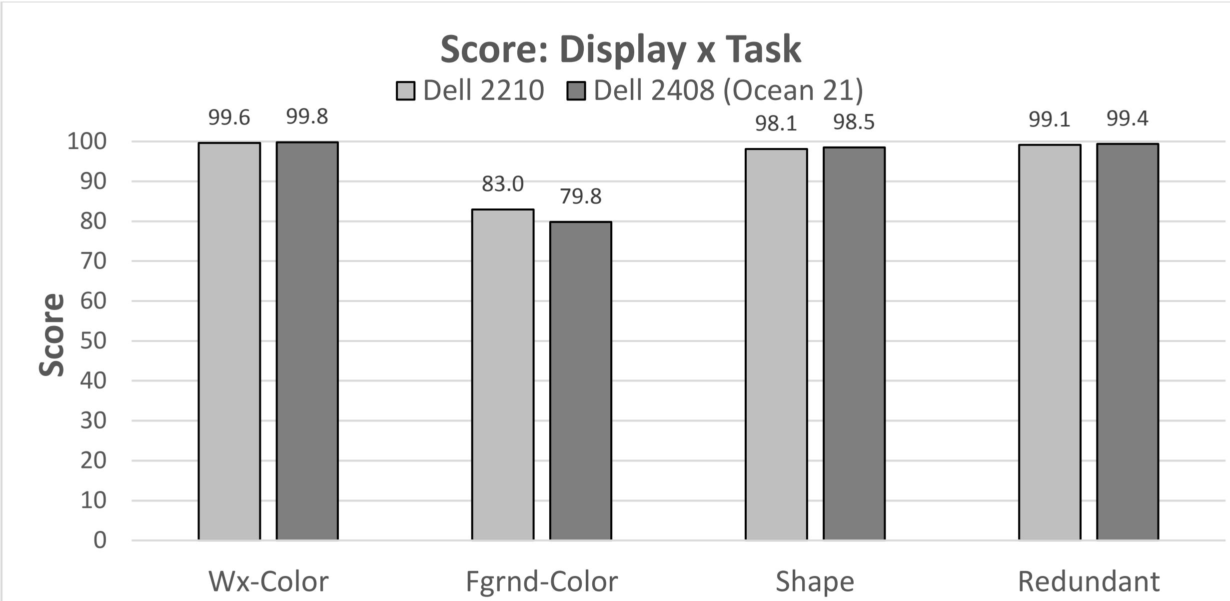

Figure 20. Display x Task interaction for Score.

Suitability of the palette as a standard (main analysis).

Of our 155 participants, 100 color normals, 19 protans, 31 deutans, and 1 tritan performed all four search tasks (4 did not complete all four tasks). We tested for effects of Task, Diagnosis, and their interaction using two-way ANOVAs. All three effects are significant for Score and Speed, $p < 0.001$. For Score, the effects of Task, Diagnosis, and their interaction account for 18, 8, and 26% of the variance, respectively (52% total). For Speed, those effects account for 83, 6, and 2% of the variance (91% total).

The interaction for Score is illustrated in Figure 21. Fisher's LSD test shows that all the pairwise differences among Diagnosis are significant for the Foreground-Color search, $p < 0.05$, but none are significant for the other three tasks. Color-normals' Scores for the weather-color and redundant searches are significantly higher than their Score for the shape search, $p < 0.05$; none of their other pairwise differences among tasks are significant. For protans and deutans, the Scores for the foreground-color search are significantly lower than for the other three tasks, $p < 0.05$; none of the differences among the other tasks are significant.

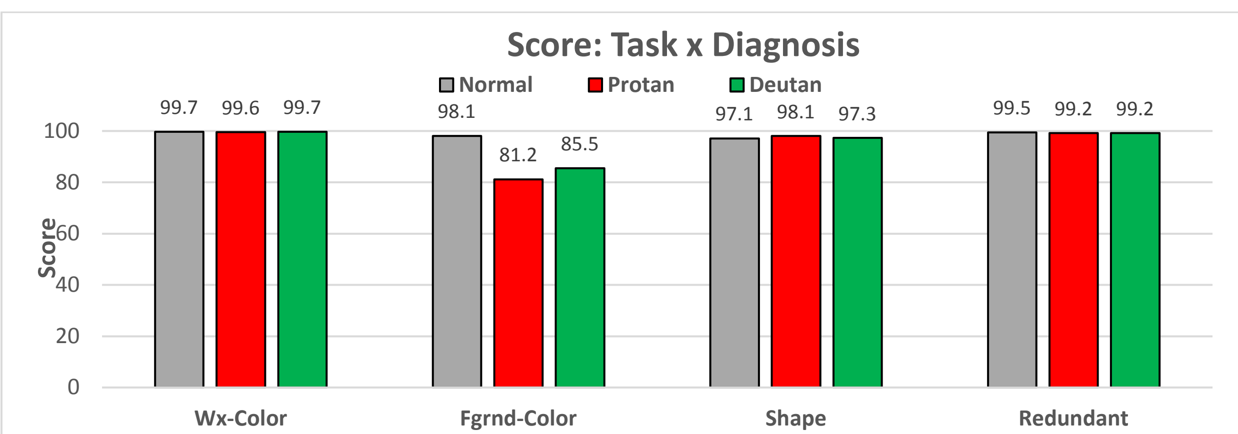

Figure 21. Task x Diagnosis interaction for Score.

The interaction for Speed is illustrated in Figures 22 and 23. The LSD test shows that, for color normals, protans, and deutans, Speed is significantly faster for the weather-color search than for the other three tasks, and Speed for the foreground-color and redundant searches is significantly faster than for the shape search, all $p < 0.05$ (see Figure 22). The interaction reflects differences among Diagnosis for the tasks (see Figure 23): All pairwise differences among Diagnosis are significant for the weather-color search, $p < 0.05$, none of those differences are significant for the shape search, and color normals are significantly faster than protans or deutans for the foreground-color and redundant searches.

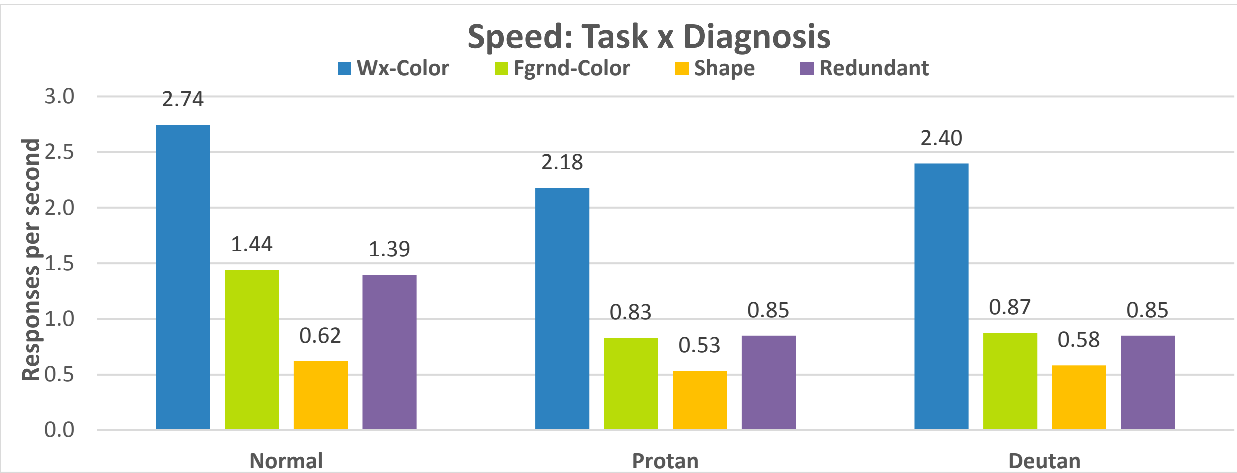

Figure 22. Task x Diagnosis interaction for Speed, graphed one way.

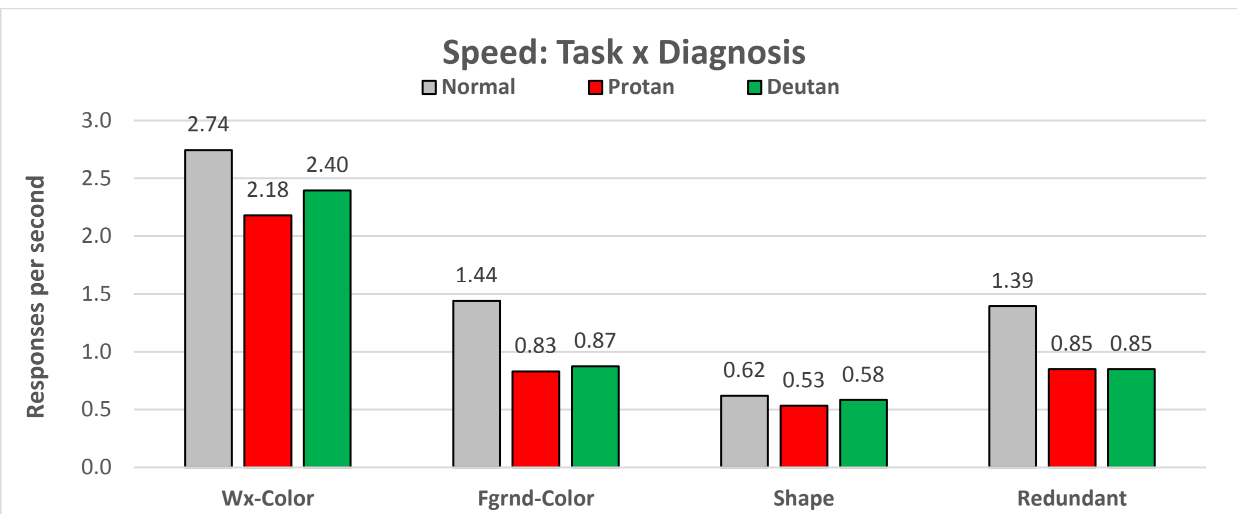

Figure 23. Task x Diagnosis interaction for Speed, graphed the other way.

Results for the tritan. The results for our tritan are shown in Figures 24 and 25. Tables 2 and 3 include the 95% confidence intervals for the color normals, so the statistical reliability of the differences can be assessed. Comparing the figures with the confidence intervals yields results slightly different from those for the protans and deutans: The tritan's mean Scores and Speeds are significantly lower than the color normals' for all but the Shape search, $p < 0.025$ one-tailed.

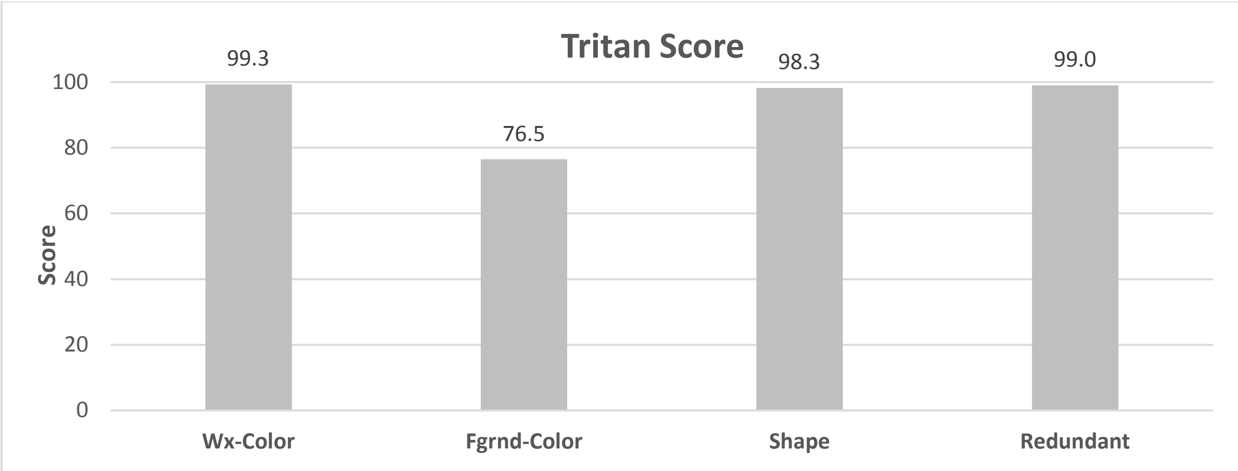

Figure 24. Main effect of Task on the tritan's scores.

Table 2. Summary statistics for the color normals' Scores.

Search Task	Mean	95% CI Lower	95% CI Upper	Median	5th percentile[3]	Cut score
Weather	99.7	99.6	99.8	100	98.3	98
Foreground	98.1	97.7	98.5	98.9	89.2	89
Shape	97.0	96.3	97.7	98.2	90.0	90
Redundant	99.4	99.3	99.5	99.6	97.9	98

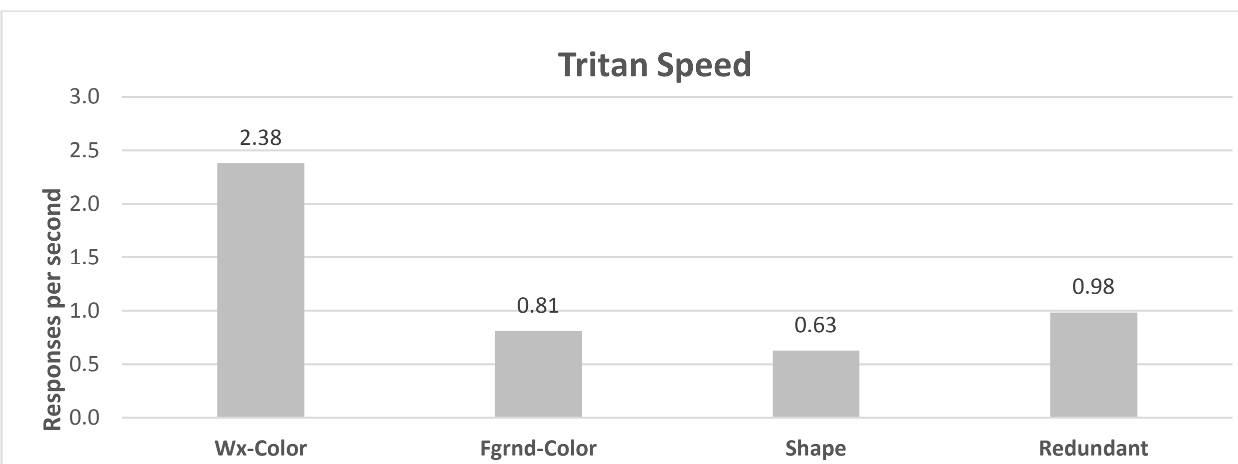

Figure 25. Main effect of Task on the tritan's speed.

[3] Determined without excluding outliers.

Table 3. Summary statistics for the color normals' Speed (responses/s).

Search Task	Mean	95% CI Lower	95% CI Upper	Median	5th percentile[4]
Weather	2.73	2.64	2.82	2.77	1.97
Foreground	1.43	1.38	1.48	1.44	1.05
Shape	0.62	0.60	0.63	0.62	0.48
Redundant	1.39	1.34	1.44	1.41	0.97

Summary. The CVD participants were slower than color normals for all but the shape-search task, but the largest difference in mean response time is 540 ms (normal vs. tritan for the foreground-color search). The largest differences for the weather-color and redundant searches are 94 and 457 ms, respectively. All of these lags are well short of the 2-s critical difference we adopted.

Foreground-color search was faster than shape search for all groups but significantly less accurate for the CVDs. Redundant search combined the high accuracy of shape search with the speed of foreground-color search for all groups.

The lack of an effect of Diagnosis on Score or Speed for the Shape search, combined with the near perfection of the mean scores (see Figures 21 and 24) is important. Those outcomes indicate that all eleven foreground colors were legible when they were used to draw alphanumerics against all four weather colors.

DISCUSSION AND CONCLUSIONS

All of our results indicate that the color palette we developed (documented in Appendix A) satisfies the criteria needed to justify its use as a standard for color coding ATC displays. The results for the weather palette show that our CVD participants matched the color-normals' accuracy with only slight (but statistically reliable) speed disadvantages. The results for the shape-search task show that the foreground colors were legible (as measured by accuracy and reading speed) for the NCV and CVD participants on all four weather backgrounds. The results for the foreground-color search show that all three CVD types benefitted more than might be expected from color coding, producing average accuracy scores of 81, 86, and 87 for the protans, deutans, and the tritan, respectively, versus 98 for the color normals. That benefit is evident also when the results for the shape and redundant searches are compared: All three CVD types (and the color normals) were faster for the redundant search while equaling or exceeding the high (> 98) scores they achieved for the shape search. For the two tasks that relate most closely to real ATC tasks, our CVD participants performed as accurately as color normals with only slight but reliable speed disadvantages; furthermore, the color normals performed well.

It was disappointing to find only one tritan for testing, despite considerable effort to canvass the Oklahoma City area over several months. Perhaps it will be possible to test the palette with more tritans in the future. For now, though, some reassurance can be had from the fact that tritans are rare – see Table 4, derived from Wyszecki and Stiles (1982, p. 464).

Table 4. Estimated frequencies of occurrence for the three main classes of color-vision defects.

	Protan	**Deutan**	**Tritan**
Male	2.0%	6.0%	0.002%
Female	0.04%	0.39%	0.001%
Average	1.02%	3.20%	0.002%

It is worth emphasizing that the standard palette was not designed or tested for use in tower environments during daytime. The colors' chromaticities might be suitable, but a gray or white background would be needed to represent weather severity 0 to combat screen reflections during daytime. That change would necessitate changing at least some of the foreground and weather color luminances to maintain the minimum 3:1 contrast ratio needed to ensure legibility and satisfy the color-difference requirements.

The FAA standard palette may be useful for color coding other applications, unrelated to ATC. It may be usable as is or provide a good starting point for developing other, special purpose color sets. Its advantages for people who have color-vision deficiencies expand its potential utility. Removing colors to obtain smaller sets should pose no problems, but adding or changing colors should be checked by testing of the sort demonstrated in this project.

REFERENCES

Barbur, J., Rodriguez-Carmona, M., Evans, S., & Milburn, N. (2009). *Minimum color vision requirements for professional flight crew, part III: Recommendations for new color vision standards* (Report No. DOT/FAA/AM-09/11). Washington, DC: Federal Aviation Administration.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley, CA: University of California Press.

Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research and Application, 12,* 94-105.

Chidester, T., Milburn, N., Lomangino, N., Baxter, N., Hughes, S., & Peterson, L. (2011). *Development, validation, and deployment of an occupational test of color vision for air traffic control specialists* (Report No. DOT/FAA/AM-11/8). Washington, DC: Federal Aviation Administration.

Chidester, T., Milburn N., Peterson, L. S., Gildea, K. Roberts, C., & Perry, D. (2013). *Development, validation, and deployment of a revised air traffic color vision test:*

Incorporating advanced technologies and oceanic procedures and en route automation modernization systems (Report No. DOT/FAA/AM-13/15). Washington, DC: Federal Aviation Administration.

CIE (2004) *Colorimetry* (3rd edition). Publication 15. Vienna, Austria: Commission Internationale de l'Eclairage.

Crawford, T. D. (1982). Defining basic color terms. *Anthropological Linguistics, 24,* 338-343.

Derefeldt, G., & Swartling, T. (1995). Colour concept retrieval by free colour naming. Identification of up to 30 colours without training. *Displays, 16,* 69-77.

FAA HF-STD-010 (2017). *Color use in air traffic control displays.* Washington, DC: US Federal Aviation Administration.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

IEC (1999). *Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management – Default RGB colour space – sRGB* (IEC 61966-2-1:1999). Geneva, Switzerland: International Electrotechnical Commission.

Post, D. L., & Calhoun, C. S. (1989). An evaluation of methods for producing desired colors on CRT monitors. *Color Research and Application, 14*, 172-186.

Post, D. L., & Calhoun, C. S. (2000). Further evaluation of methods for producing desired colors on CRT monitors. *Color Research and Application, 25,* 90-104.

Post, D. L., & Goode, W. E. (2017). A color-code design tool. In *Proceedings of the 19th International Symposium on Aviation Psychology,* 77-82. Dayton, OH: Wright State University.

Smallman, H. S., & Boynton, R. M. (1993). On the usefulness of basic colour coding in an information display. *Displays, 14,* 158-165.

Wilson, E., Wilson, D., & Jha, P. D. (2007). *Airport traffic control tower lighting and viewing measurements.* (Report No. DOT/FAA/TC-07/09). Washington, DC: Federal Aviation Administration.

Wyszecki, G., & Stiles, W. S. (1982). *Color science* (2nd ed.). New York: Wiley.

APPENDIX A: FAA STANDARD COLOR PALETTE

Table A1. FAA standard color palette: Foreground colors.

	Color name	u'	v'	%Y	sR	sG	sB	Hex
	White	0.1978	0.4683	100	255	255	255	FFFFFF
	Pink	0.266	0.418	41	246	132	216	F684D8
	Gray	0.1978	0.4683	45	179	179	179	B3B3B3
	Blue	0.17	0.348	28	94	141	246	5E8DF6
	Orange	0.294	0.541	42	254	147	13	FE930D
	Red	0.368	0.457	26	255	58	134	FF3A86
	Green	0.13	0.54	55	35	225	98	23E162
	Yellow	0.193	0.55	80	223	243	52	DFF334
	Purple	0.276	0.304	23	216	34	255	D822FF
	Aqua	0.142	0.428	50	7	205	237	07CDED
	Brown	0.241	0.519	34	197	149	91	C5955B

Table A2. FAA standard color palette: Weather colors.

	Color name	Severity	u'	v'	%Y	sR	sG	sB	Hex
	Black	0	---	---	0.0	0	0	0	000000
	Wx-Green	1 & 2	0.15	0.5	3.2	23	57	40	173928
	Wx-Yellow	3 & 4	0.23	0.54	7.1	90	74	20	5A4A14
	Wx-Red	5 & 6	0.26	0.4	5.0	93	46	89	5D2E59

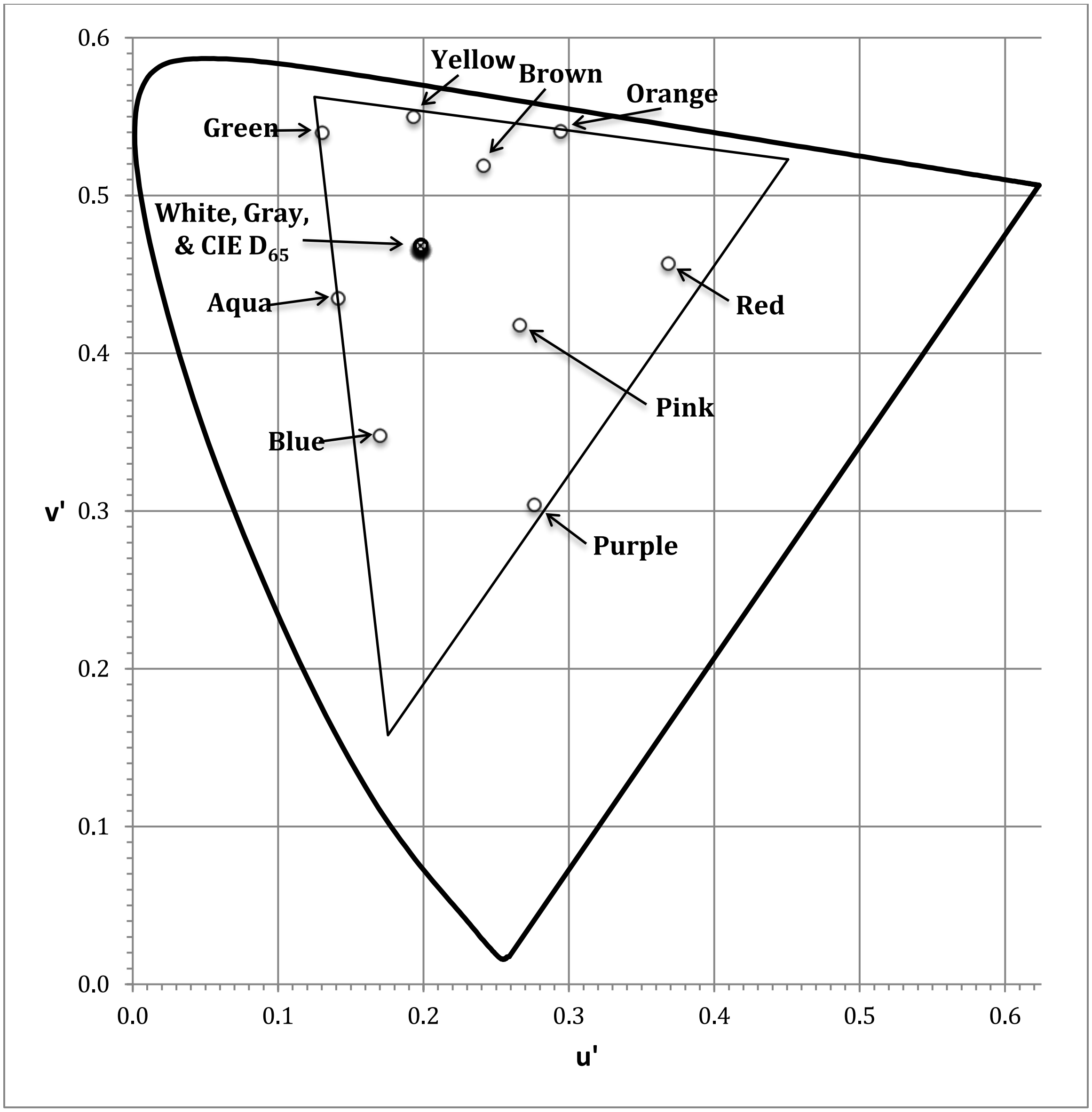

Figure A1. FAA standard color palette: Foreground colors, plotted on the CIE 1976 u′v′–chromaticity diagram.

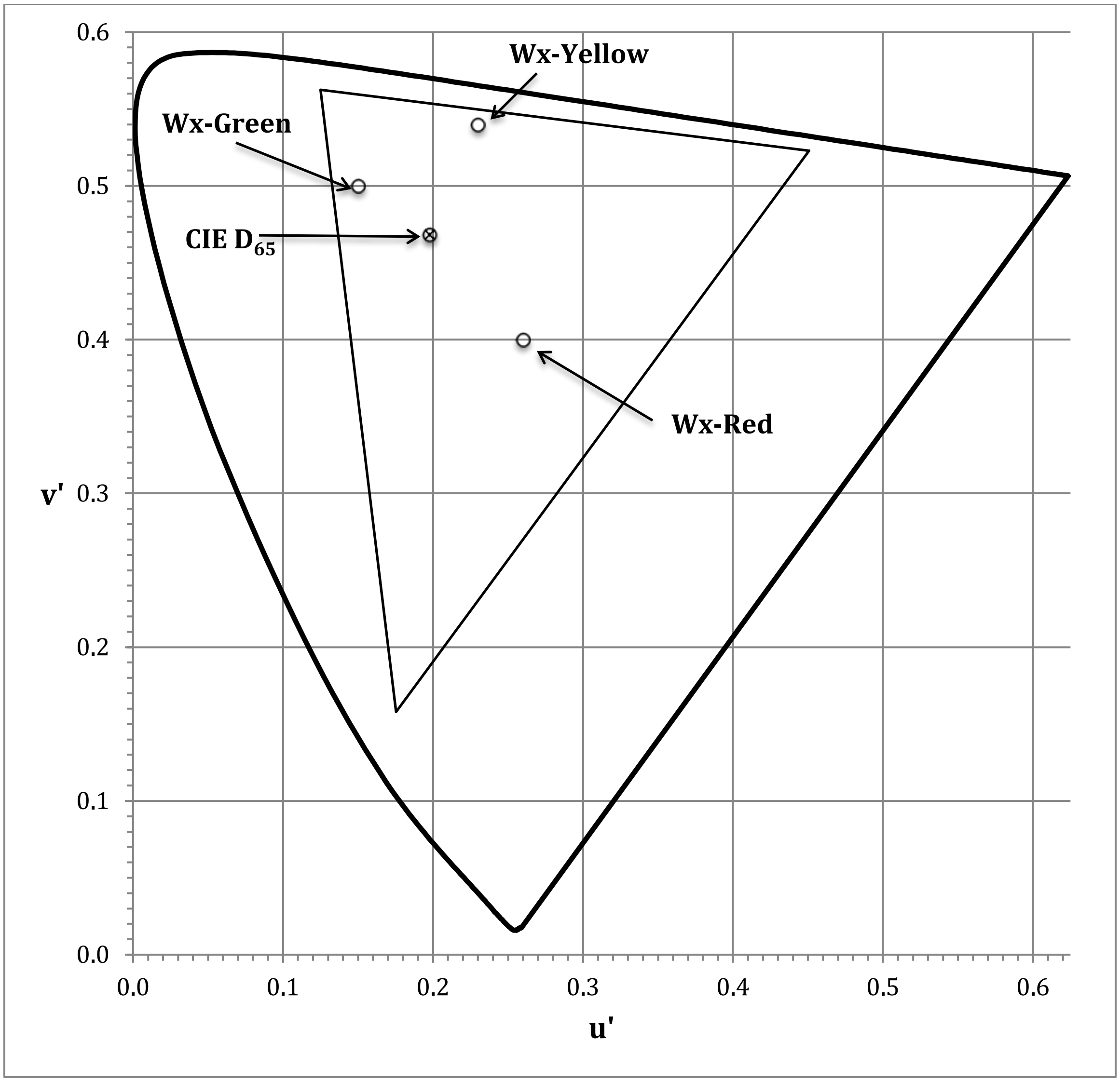

Figure A2. FAA standard color palette: Low-, moderate-, and high- severity weather colors, plotted on the CIE 1976 u′v′–chromaticity diagram.

APPENDIX B: FAA STANDARD PALETTE'S CONFORMANCE TO FAA HF-STD-010 REQUIREMENTS

FAA HF-STD-010 (2017) includes three colorimetric requirements that address the recognizability, discriminability, legibility, and conspicuity of a color set. We assessed the standard palette's conformance to those requirements using Post and Goode's (2017) Palette Designer program.

Discriminability

Discriminability is assessed in the color standard by computing the color differences between all pairings of the foreground colors plus the background color, using Equation 1:

$$\Delta E^*_{ab} = ((\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2)^{0.5} \qquad (1)$$

where ΔE^*_{ab} is the color difference and ΔL^*, Δa^*, and Δb^* are computed in accordance with the conventions of the CIE 1976 ($L^*a^*b^*$) color space (CIELAB) described in CIE (2004). The standard requires that ΔE^*_{ab} for each color difference be greater than or equal to 9.9. That criterion is taken as a conservative estimate of the minimum acceptable color difference among spatially adjacent colors that must be discriminable.

Tables B1 and B2 show the results from Eqn. 1 for all pairings of the palette's foreground color and all pairings of the weather colors. The calculations take account of the 40 lux of ambient illumination that was used in our experiments. All values in both tables exceed 9.9 by large margins. Consequently, the discriminability of the colors in the standard palette is predicted to be satisfactory.

Table B1. Discriminability of the standard palette's 11 foreground colors, per Eqn. 1.

Color	White	Pink	Gray	Blue	Orange	Red	Green	Yellow	Purple	Aqua
Pink	66.01									
Gray	26.97	59.03								
Blue	71.57	51.71	60.83							
Orange	85.24	99.18	80.18	131.94						
Red	86.80	38.13	77.53	86.49	81.67					
Green	87.32	143.34	85.01	137.41	105.33	153.19				
Yellow	85.96	134.42	87.57	148.35	64.20	132.65	54.35			
Purple	120.54	58.72	113.35	73.24	153.46	75.19	198.12	192.94		
Aqua	47.22	87.52	40.90	61.88	117.20	114.36	80.77	105.81	131.37	
Brown	51.71	74.09	39.00	94.05	42.91	72.35	82.68	64.66	131.05	75.54

Table B2. Discriminability of the standard palette's four weather colors, per Eqn. 1.

Color	Black	Wx-Green	Wx-Yellow
Wx-Green	27.07		
Wx-Yellow	45.19	31.11	
Wx-Red	42.64	51.71	57.33

Legibility

The standard requires that the luminance-contrast ratio (i.e., L_{max}:L_{min}) between each object on the display screen and its background be greater than or equal to 3:1. That criterion is taken as the minimum needed to ensure symbol legibility. Tables B3 - B6 show the contrast ratios for all 11 foreground colors against the 4 weather backgrounds. The calculations take account of the 40 lux of ambient illumination that was used in our experiments.

All ratios exceed 3:1 – often, by large margins. Consequently, the legibility of the standard palette's foreground colors is predicted to be satisfactory.

Table B3. Luminance-contrast ratios between foreground colors and black weather background.

Color	White	Pink	Gray	Blue	Orange	Red	Green	Yellow	Purple	Aqua	Brown
Black	257.4:1	106.1:1	116.4:1	72.8:1	108.7:1	67.6:1	142.0:1	206.1:1	59.9:1	129.2:1	88.2:1

Table B4. Luminance-contrast ratios between foreground colors and green weather background.

Color	White	Pink	Gray	Blue	Orange	Red	Green	Yellow	Magenta	Aqua	Brown
Wx-Green	34.0:1	14.0:1	15.4:1	9.6:1	14.4:1	8.9:1	18.8:1	27.2:1	7.9:1	17.1:1	11.7:1

Table B5. Luminance-contrast ratios between foreground colors and yellow weather background.

Contrast ratio	White	Pink	Gray	Blue	Orange	Red	Green	Yellow	Magenta	Aqua	Brown
Wx-Yellow	16.6:1	6.9:1	7.5:1	4.7:1	7.0:1	4.4:1	9.2:1	13.3:1	3.9:1	8.4:1	5.7:1

Table B6. Luminance-contrast ratios between foreground colors and red weather background.

Contrast ratio	White	Pink	Gray	Blue	Orange	Red	Green	Yellow	Magenta	Aqua	Brown
Wx-Red	22.9:1	9.4:1	10.3:1	6.5:1	9.7:1	6.0:1	12.6:1	18.3:1	5.3:1	11.5:1	7.8:1

Recognizability and Conspicuity

Recognizability and conspicuity is assessed in the standard by computing the color differences between color pairings using Equation 2:

$$\Delta E^{*}uv\text{-}sc = \left((K_{L^*} * \Delta L^*)^2 + (K_{u^*} * \Delta u^*)^2 + (K_{v^*} * \Delta v^*)^2 \right)^{0.5} , \quad (2)$$

where ΔE*uv-sc is the size-corrected color difference, the coefficients KL*, Ku*, and Kv* are computed as shown below, and ΔL*, Δu*, and Δv* are computed in accordance with the conventions of the CIE 1976 (L*u*v*) color space (CIELUV) described in CIE (2004).

$$K_{L^*} = 1.0366 - e^{0.15263 - 0.05766A} \quad \text{for } 0 < A < 60 , \quad (3)$$

$$K_{u^*} = 0.008991A - 0.0065 \quad \text{for } 0 < A \le 32 , \quad (4)$$

$$= 0.0257A - 0.5403 \quad \text{for } 32 < A < 60 , \quad (5)$$

$$K_{v^*} = 0.005446A - 0.042 \quad \text{for } 0 < A \le 32 \text{ , and} \quad (6)$$

$$= 0.031A - 0.8594 \quad \text{for } 32 < A < 60 , \quad (7)$$

where A is the visual angle subtended by the stimulus in arc-minutes. For A ≥ 60 arc-minutes, KL* = Ku* = Kv* = 1. These coefficients model reductions in color recognizability that occur for stimuli subtending < 60 arc-minutes of visual angle.

The standard recommends but does not require that Equation 2 yield a value greater than or equal to 28. That criterion is taken as the minimum color difference needed to obtain asymptotic visual search time.

Tables B7 and B8 show the results from Eqn. 2 for all pairwise combinations of the foreground colors and all pairwise combinations of the weather backgrounds. The calculations take account of the 40 lux of ambient illumination that was used in our experiments. The pink highlighting shown in the tables denotes the many color pairs that do not meet the minimum color difference recommendation. That outcome predicts that search times for many colors will be longer than the minimum possible. Whatever the outcome, though, whether the search times will be acceptable can only be learned empirically.

Table B7. Conspicuity of the standard palette's foreground colors, per Eqn. 2.

Color	White	Pink	Gray	Blue	Orange	Red	Green	Yellow	Purple	Aqua
Pink	22.79									
Gray	18.04	11.23								
Blue	27.63	16.23	11.25							
Orange	25.10	8.80	16.00	23.01						
Red	35.61	14.18	24.23	26.50	12.04					
Green	19.00	24.72	13.53	18.78	27.71	37.13				
Yellow	8.58	20.77	14.11	24.82	21.38	32.82	13.99			
Purple	32.36	11.14	17.18	13.91	17.05	14.44	29.77	30.18		
Aqua	18.74	20.74	10.10	12.90	25.96	33.89	7.73	16.06	24.44	
Brown	24.32	8.11	8.72	13.97	9.85	16.70	20.69	19.56	12.95	18.35

Table B8. Conspicuity of the standard palette's weather colors, per Eqn. 2.

Color	Black	Wx-Green	Wx-Yellow
Wx-Green	13.77		
Wx-Yellow	21.31	8.49	
Wx-Red	18.10	7.32	5.06

Confusion lines

Palette Designer draws confusion lines for each type of color-vision deficiency. The results for the standard palette are shown in Figures B1 – B3. The confusion lines for many foreground-color pairs are nearly colinear. That outcome predicts that luminance differences must be provided to help CVD viewers distinguish those pairs. The weather colors, however, appear to be discriminable for CVD viewers without need for luminance differences.

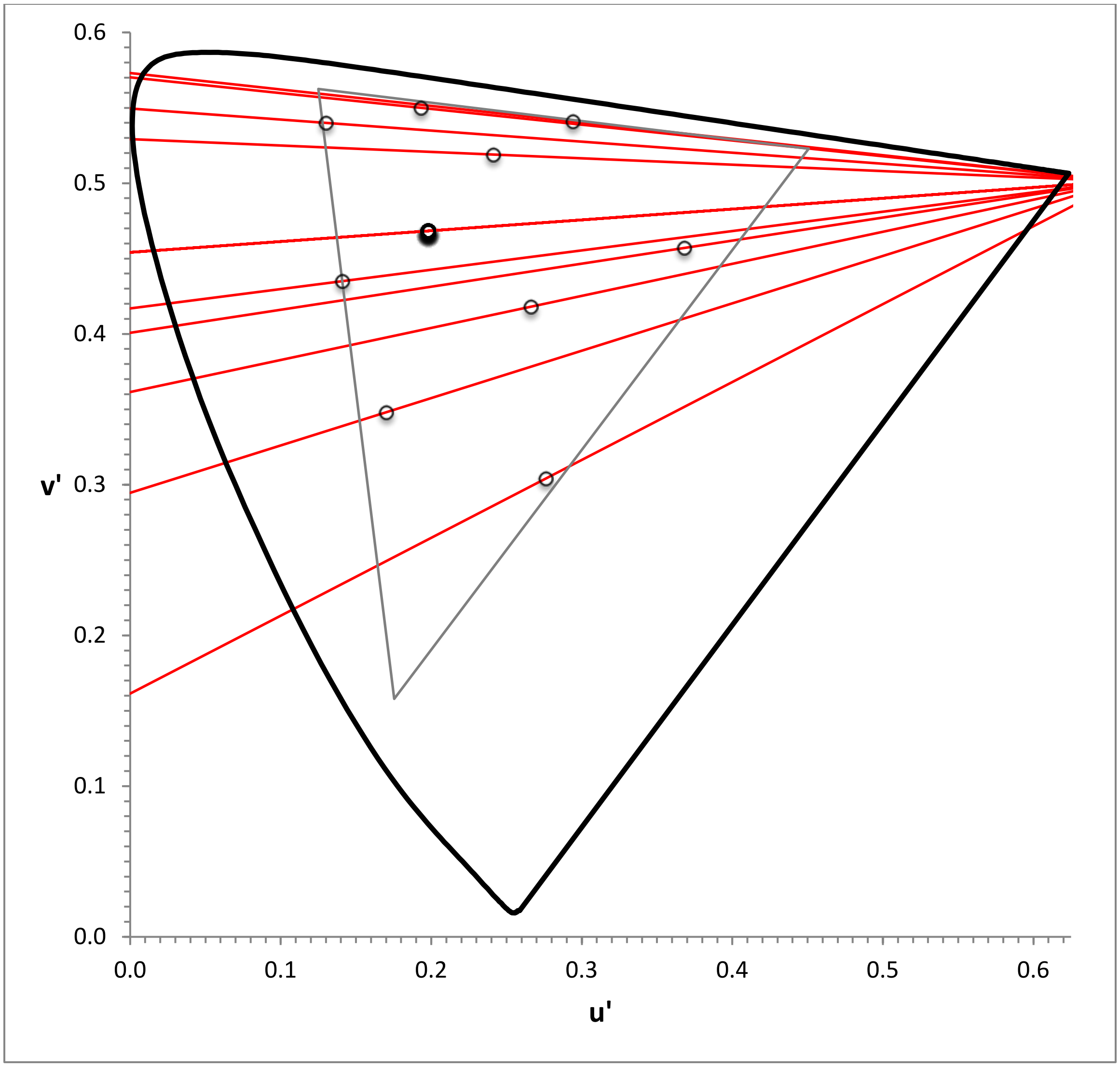

Figure B1. Protan confusion lines for the FAA standard palette’s foreground colors.

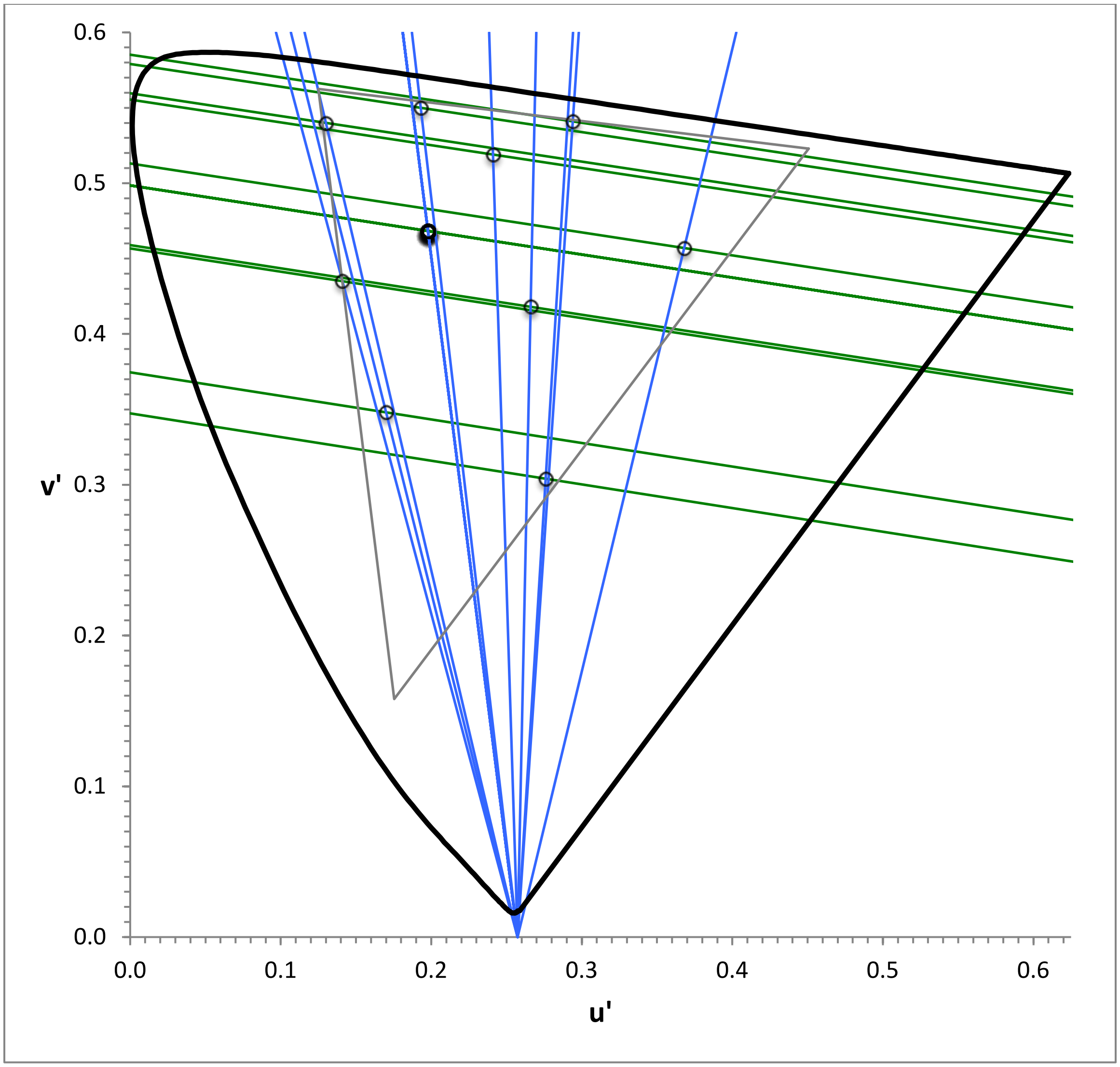

Figure B2. Deutan (green) and tritan (blue) confusion lines for the FAA standard palette's foreground colors.

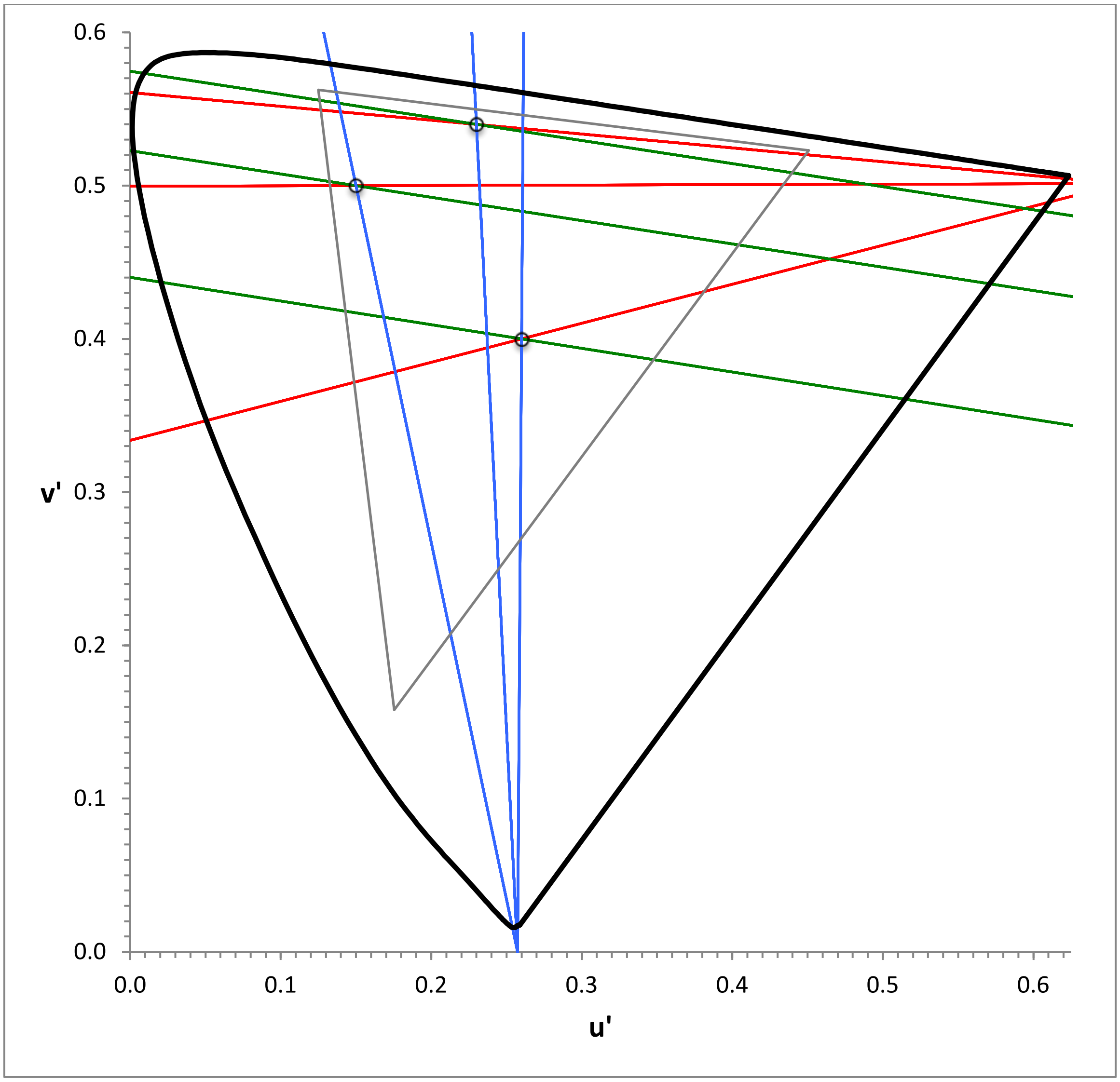

Figure B3. Protan (red), deutan (green), and tritan (blue) confusion lines for the FAA standard palette's weather colors.

APPENDIX C: INFORMED CONSENT FORM

FEDERAL AVIATION ADMINISTRATION

PROJECT: EVALUATION OF A COLOR PALETTE FOR AIR TRAFFIC CONTROL

AAM-500 Research Task No. AHRR521

Principal Investigator: Dr. Kevin Gildea, Ph.D.

A Cooperative Agreement between:

the Aerospace Human Factors Research Division (AAM-500).

Civil Aerospace Medical Institute (CAMI) and

David Post, Ph.D.

Wright State University (WSU) Dayton, OH

INDIVIDUAL'S CONSENT TO VOLUNTARILY PARTICIPATE IN A RESEARCH PROJECT

I, _____________________________, understand that this study, entitled **"EVALUATlON OF A COLOR PALETTE FOR AIR TRAFFIC CONTROL AND COLORED LED LIGHTS FOR NORMAL AND CVD PARTICIPANTS"** is sponsored by the Federal Aviation Administration's (FAA) Civil Aerospace Medical Institute (CAMI) in Oklahoma City, OK and is under the direct supervision of Dr. Kevin Gildea of the Aerospace Human Factors Research Division of CAMI and Dr. David Post of the Wright State University, Dayton, OH.

1. Purpose. The objective of this study is to assess a proposed color palette for air traffic control displays and to ensure that it is usable by individuals with normal color vision and those with some types and degrees of color vision deficiency known to be in the current workforce. Further, the goal is to determine if the ability to distinguish the colors of the color palette differs between display types, including those currently in use and/or newly proposed display types. Approximately 100 participants will be involved in this research study.

2. Procedures used in this study. I understand that the data-collection protocols used in this study involve computer-driven assessments that require response by computer mouse point/click or keystroke, response pad, some paper-and-pencil forms, and some color vision assessments that require me to record the color name or colored numbers. In some cases, I may be asked to respond to verbal instructions given through headphones. The performance data gathered during the study will be used for subsequent analyses. In no way will my name be associated with the coded subject identification number for any presentation or publication. Research results will be reported as group data.

3. Description of study requirements. There are various types of color vision deficiencies and various degrees or those deficiencies. For this study, the researchers need a certain number of each type and degree of color vision deficiency and a certain number of participants with normal color vision. All volunteers will complete a color vision assessment to determine their color vision status: normal color vision or the type and degree of their color vision deficiency. I understand that I will be asked to complete several color perception tasks and that, collectively, those tasks may last between 2 to 5 hours, and the length of time that I will be asked to participate depends upon the fit of my color vision diagnosis to the requirements of the study. If my participation is longer than 2 hours, I will be given a short break after 2 hours of work and a 30-minute lunch break after not more than 4 hours of work.

I understand that I will be asked to complete a visual acuity screening (10 minutes) to ensure that I have normal, or corrected-to-normal, near and far vision, with of a minimum of score of 20/30 visual acuity for participation in this study. Next, I will be asked to complete the Colour Assessment and Diagnosis (CAD) (30 minutes) to determine whether my color vision type is needed for this study.

I understand that if I am accepted for this study, I will be asked to complete any or all of the following color naming, color identification, color discrimination, or color matching tasks: The Air Traffic Color Vision (ATCOV) assessment (60 minutes), the Palette Evaluation Tool (PET) (60 minutes); and (e) an interactive color discrimination task working with the researcher (not more than 60 minutes). I may be asked to work interactively with Dr. Post or Dr. Gildea on a color discrimination task during which the researcher presents a pair or group of colors on the computer display and I will be asked if the colors are distinguishable as separate colors or if any two or more look the same, or whether they appear so similar that I would easily confuse them. That task may take as long as 60 minutes to complete. I understand that if I have a protan-type of color vision deficiency, I may be asked to complete the PET assessment once again on a different computer display, thus adding 60 more minutes of work. I understand that I will be asked to identify the color of two types of LEDs (single-die and RGB). RGB (red, green, blue) refers to the way that colors are made with lights. TVs, LED Christmas lights, and computer monitors use a similar method for making colored lights.

As a person who is in reasonably good health, I know of no medical or other conditions that would prevent me from attentively operating a computer workstation, operating a keyboard, and a mouse, for up to 120 minutes in one sitting. My health condition allows me to intensively work at a computer workstation for 5 to 6 hours in a day. I do not have any prior experience of seizure due to exposure to flicking lights, moving lights, flashing displays, etc. I do not have diabetes that is controlled by medicine (as the medicine may affect one's color vision). Additionally, I have no medical conditions which would prevent me from standing or sitting in a dark room for up to 30 minutes, or prevent me from listening to audio instructions through headphones.

I understand that:

a. I must perform to the best or my ability on all of the color vision tasks to allow accurate assessment of my color vision performance including accuracy and response time.
b. It is essential that I am attentive and follow the instructions during the course of this study.
c. Completing all of the tasks described above will take between 2 and 5 hours and that I will be given short breaks periodically, including a 30- minute lunch break after 4 hours of work.
d. I may be asked to work using a computer display in a dark or dimly lighted room.
e. I may work interactively with the researcher to identify colors that are easily discriminable to me and that my task is to communicate that information to the researcher.
f. If I have any questions, I may ask the experimenter.
g. I agree to immediately report any injury or suspected adverse effects from this study to Dr. Kevin Gildea at (405) 954-7481.

4. Possible risks. The risks involved in this study are minimal. Sitting for approximately 120 minutes at a time performing the computer-based tasks may produce some fatigue. The overall fatigue experienced from a day in the laboratory should be similar to that experienced during the course of an intensive work day at a computer.

5. Benefits. The major benefit from this study for the aviation community will be the development of a standard color palette for air traffic control displays that is usable for people with normal color vision and for those individuals with some types of color vision deficiencies. This study will also aid in developing new methods of assessing color vision sensitivity of air traffic controller applicants. Furthermore, I understand that my participation is strictly voluntary and I benefit by acquiring an understanding of my color vision. I understand that I will be paid for my time participating in this study.

6. Conditions of participation. I understand that I am free to withdraw my consent and to withdraw from the project at any time. I understand that my participation in this study may be terminated at any time at the experimenter's discretion. Examples of reasons for termination would include an intentional lack of cooperation or disregard of the experimental equipment or procedures.

7. Assurances. Furthermore, I understand that:

a. The CAMI and Wright State University researchers, Dr. Kevin Gildea, and Dr. Dave Post and their colleagues, will take every precaution to utilize proven research procedures.
b. If any new findings develop during the course of this study which may relate to my decision to continue participation, I will be informed.
c. By signing this consent form, I have not waived any legal rights or released CAMI or any individual from liability for negligence. I may revoke my consent and withdraw from the study at any time.
d. The results will be treated as confidential and will receive a code number so that they will be anonymous when filed with the Laboratory Manager. In no case will any use be made of these results other than the application of experimental analyses unless I provide explicit written permission.
e. I understand that Dr. Nelda Milburn, Dr. Kevin Gildea, Dr. Dave Post or their colleagues will be available to answer my questions concerning procedures throughout this study.

8. Qualification for participating in the study. I understand the following requirements for qualifying me to participate this study:

1. Required age: 18 -55 years.
    - This requirement represents the eligible age range to work as an air traffic controller. The FAA requires that a controller applicant holds a high school diploma plus three

years of post-high school experience. This results in the age of the youngest controllers in the current work force being roughly 20 years old. The official retirement age for controllers is 56 years.

2. Required gender:
   - Both men and women will be allowed to participate.
3. Required medical conditions:
   - I do not have any prior experience of the following conditions:

     (a) Seizure induced by optical stigma due to exposure to flicking lights, moving lights, flashing displays, etc.;

     (b) Diabetes that is controlled by medicine, as the medicine may affect my color vision;
   - My health condition allows me to intensively work for 5-6 hours in a day with breaks, and I understand that some color assessment tasks will be administered in a dark or dimly-lighted room.

9. Compensation / Injury (FOR CONTRACT SUBJECTS). By signing, I understand that the contract company that recruited me is responsible for the compensation of my participation in this study and potential injury during the study.

10. Signatures. By signing, I understand that I am free to withdraw this consent and discontinue participation at any time, for any reason. If I do withdraw, I will be paid for my time worked.

I have read this consent document. I understand its contents and I freely consent to participate in this study under the conditions described here. I will receive a copy of this form up on request.

________________________________ Investigator's Signature

________________________________ Participant's Signature

________________________________ Participant's Signature Witnessed by

________________________________ Date