Guidelines for the Use of Color in ATC Displays

DOT/FAA/AR-99/52
DOT-VNTSC-FAA-98-5

Office of Aviation Research
Washington, DC 20591

Kim Cardosi, Ph.D.

Dan Hannon, Ph.D.
Hannon Consulting

U.S. Department of Transportation
Research and Special Programs Administration
John A. Volpe National Transportation Systems Center
Cambridge, MA 02142

June 1999

This document is available to the public
through the National Technical Information
Service, Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

NOTICE

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE
	June 1999	Final Report

4. TITLE AND SUBTITLE

Guidelines for the Use of Color in ATC Displays

5. FUNDING NUMBERS

FA9L1/A9112

6. AUTHOR(S)

Kim Cardosi, Ph.D. – Volpe Center
Dan Hannon, Ph.D. – Hannon Consulting

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Department of Transportation
Research and Special Programs Administration
John A. Volpe National Transportation Systems Center
Cambridge, MA 02142-1093

8. PERFORMING ORGANIZATION REPORT NUMBER

DOT-VNTSC-FAA-98-5

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Department of Transportation
Federal Aviation Administration
Research and Development Service
800 Independence Avenue, SW
Washington, DC 20591

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DOT/FAA/AR-99/52

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

This document is available to the public through the National Technical Information Service, Springfield, VA 22161

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Color is probably the most effective, compelling, and attractive method available for coding visual information on a display. However, caution must be used in the application of color to displays for air traffic control (ATC), because it is easy to do more harm than good. The only thing that is truly obvious about the use of color on displays is that its benefits and drawbacks depend upon the task. This paper offers general guidelines on how color should, and should not, be used, but does not define a specific color-coding scheme. These guidelines are based on what is known about human vision, display capabilities, the knowledge gained from the lessons learned from the uses of color in the cockpit and ATC environments, and human factors "best practices."

The report also discusses a series of experiments that examined color production capabilities within and across five Sony DDM-2801C monitors and selected and validated an "ideal" color set for this monitor.

14. SUBJECT TERMS

Color displays, air traffic control displays, color guidelines, color coding

15. NUMBER OF PAGES

64

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

PREFACE

This work was sponsored by the Federal Aviation Administration's Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100), Integrated Product Team for En Route (AUA-200), and the STARS Product Team (AUA-310). The authors wish to thank Melissa Bud and Sarah Miescher for taking the measurements on the SONY monitors and conducting experiment two at the FAA Technical Center; Sarah Miescher for conducting experiment three at the Volpe Center; Robert DiSario and Sarah Miescher for conducting the data analysis; Gary Worrall for his technical assistance in setting up the program to present the stimuli and record the data; and Tom Robino for his logistical support. The authors are also very grateful to Lawrence Arend (NASA/Ames Research Center), Glen Hewitt (AAR-100), Jay Merkle (AUA-200), and Henry Mertens (CAMI) for their substantial contributions to an earlier draft of this report.

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in)	=	2.5 centimeters (cm)
1 foot (ft)	=	30 centimeters (cm)
1 yard (yd)	=	0.9 meter (m)
1 mile (mi)	=	1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2)	=	6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2)	=	0.09 square meter (m^2)
1 square yard (sq yd, yd^2)	=	0.8 square meter (m^2)
1 square mile (sq mi, mi^2)	=	2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he)	=	4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz)	=	28 grams (gm)
1 pound (lb)	=	0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb)	=	0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp)	=	5 milliliters (ml)
1 tablespoon (tbsp)	=	15 milliliters (ml)
1 fluid ounce (fl oz)	=	30 milliliters (ml)
1 cup (c)	=	0.24 liter (l)
1 pint (pt)	=	0.47 liter (l)
1 quart (qt)	=	0.96 liter (l)
1 gallon (gal)	=	3.8 liters (l)
1 cubic foot (cu ft, ft^3)	=	0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3)	=	0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] °F = y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm)	=	0.04 inch (in)
1 centimeter (cm)	=	0.4 inch (in)
1 meter (m)	=	3.3 feet (ft)
1 meter (m)	=	1.1 yards (yd)
1 kilometer (km)	=	0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2)	=	0.16 square inch (sq in, in^2)
1 square meter (m^2)	=	1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2)	=	0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2)	=	1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm)	=	0.036 ounce (oz)
1 kilogram (kg)	=	2.2 pounds (lb)
1 tonne (t)	=	1,000 kilograms (kg)
	=	1.1 short tons

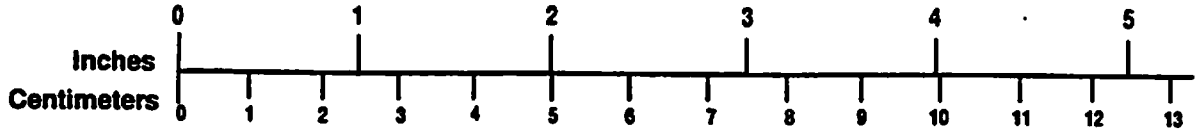
VOLUME (APPROXIMATE)

1 milliliter (ml)	=	0.03 fluid ounce (fl oz)
1 liter (l)	=	2.1 pints (pt)
1 liter (l)	=	1.06 quarts (qt)
1 liter (l)	=	0.26 gallon (gal)
1 cubic meter (m^3)	=	36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3)	=	1.3 cubic yards (cu yd, yd^3)

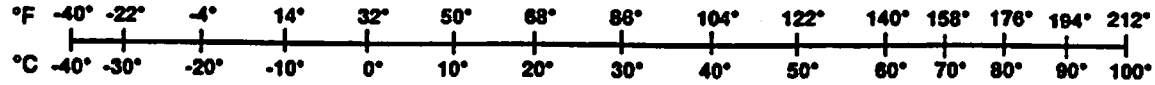
TEMPERATURE (EXACT)

[(9/5) y + 32] °C = x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSION



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 9/17/96

TABLE OF CONTENTS

Section — Page

1. INTRODUCTION 1
   1.1 Purpose 1
   1.2 Experiments Using the SONY DDM-2801C Monitor 1
2. CAUTIONS ON THE USE OF COLOR ON ATC DISPLAYS 3
   2.1 Questions on the Use of Color 3
   2.2 Cautions on the Use of Color 4
3. HOW SHOULD COLOR BE USED ON ATC DISPLAYS? 7
   3.1 General Guidelines 7
   3.2 Controllers' Suggestions for Color-Coded Information 9
   3.3 How Many Colors Can be Used Successfully? 9
   3.4 What are the Six Colors That Should be Used? 10
      3.4.1 Colors for the Sony DDM-2801C 11
   3.5 Specific Recommendations 13
      3.5.1 Background 13
      3.5.2 Data Blocks 15
      3.5.3 Alerts and Warnings 16
      3.5.4 Cursor 16
   3.6 Designing a Color Scheme for an ATC Display 16
4. HOW TO MEASURE THE EFFECTIVENESS OF A SPECIFIC USE OF COLOR ON AN ATC DISPLAY 19
5. USE OF COLOR IN AVIATION DISPLAYS 21
   5.1 Cockpit Displays 21
   5.2 ATC Color Displays in the U.S. 21
   5.3 ATC Color Displays – Lessons Learned from Other Countries. 22
      5.3.1 Canada 22
      5.3.2 The United Kingdom 23
      5.3.3 ODID (Operational Display Input Development) IV 25
6. CONCLUSIONS 29

APPENDICES

Appendix A: Determining the Ideal Color Set for the Sony DDM-2801C 31
Appendix B: Summary of Guidelines for the Use of Color on ATC Displays 47
Appendix C: Detailed Comments from the ODID IV Report 49

GLOSSARY 51

REFERENCES 53

LIST OF FIGURES

Figure Page

A-1 Log-log Plot of Input versus Luminance for the Sony DDM-2801C 33

A-2 CIE Plot of the Colors Produced by the Sony DDM-2801C 36

LIST OF TABLES

Table Page

1 Colors that Yielded 99% Accuracy in Color Identification and Legibility on a Black Background using the Sony DDM-2801C 12

A-1 Measurements from Nine Screen Positions on Five Sony DDM-2801C Monitors 34

A-2 Percent Accuracy for Color Identification for Experiment 2 38

A-3 Definition of the Color Set Tested in Experiment 3 41

A-4 Mean Percent Accuracies of Color Identification and Response Times (RTs) in Seconds for Correct Responses in Experiment 3 42

A-5 Color Identification Accuracies from Experiment 2 and Experiment 3 45

EXECUTIVE SUMMARY

Color is an effective and attractive tool for coding information on a display. It can help to organize and categorize complex information and can also have a profound effect on our ability to search for specific information. Color is both attractive and functional. However, caution must be used in the application of color to displays because it is easy to do more harm than good. The only thing that is truly obvious about the use of color on displays is that its benefits and drawbacks depend upon the task.

The primary purpose of this report is to provide general guidance to the STARS (Standard Terminal Replacement System) and DSR (Display System Replacement) program offices on the use of color on air traffic control (ATC) displays. However, the information in this report will also be useful to any program considering the use of color on its displays. It offers general guidelines on how color should and should not be used, but does not define a specific color-coding scheme. These guidelines are based on what is known about human vision, display capabilities, the knowledge gained from lessons learned about the uses of color in the cockpit and ATC environments, and human factors "best practices." The report discusses how the effectiveness of color displays can be assessed and provides an update on studies in progress about specific uses of color on ATC displays. Finally, the report summarizes what is known about the general benefits and limitations of the use of color on displays, the guidelines for use of color in the cockpit, and the existing recommendations about using color in ATC displays from air traffic organizations in other countries.

The guidelines presented in the report relating to the use of color in ATC displays are as follows:

- Displays must be designed for the tasks they need to support and the environment in which they will be used.
- Whenever color is used to code critical information it must be used along with another method of coding.
- When color is used to assign a unique meaning to specific colors (such as red for emergencies or green for aircraft under my control), it is imperative that no more than six colors be used.
- Care must be taken to ensure all color-coded text and symbols are presented in sufficient contrast.
- Cultural color conventions (such as red for danger and yellow for warning) should not be violated.
- Pure blue should not be used for text, small symbols, other fine detail, or as a background color.
- Bright, highly saturated, colors should be used sparingly.
- Color use needs to be consistent across all of the displays that a single controller will use.

- The specific colors that are selected for a display must take into account the ambient environment and the capabilities of the specific monitor.
- The entire set of displays that a controller will use must be designed and evaluated as a whole and not as a combination of parts.
- Any implementation of color needs to be tested in the context of the tasks that it is designed to support and the environment in which it is intended to be used.

This report also describes a series of experiments that examined the variability in the color producing characteristics of five "calibrated" SONY DDM- 2801C (20 x 20) monitors, and defined and validated an ideal color set for the Sony 20 x 20 monitor.

Data are presented for color identification and legibility for several different "reds," "greens," "blues," "yellows," "magentas," and "cyans" on a black, dark grey, medium grey and light grey background. In general, legibility was excellent with all of the colors at the eight and 12 pixel sizes tested. Color identification accuracy varied widely depending on the individual color/background combination. On the black background, there was at least one example of every color except cyan that was recognized correctly 99% of the time. Overall, color identification was very good on the black, dark grey, and medium grey backgrounds but suffered on the light grey background with only three of the colors tested (red, white, and magenta) yielding color identification accuracy of 99%.

1. INTRODUCTION

"In the majority of computer installations monochrome displays are virtually extinct and color is *de rigeur*, but when color is used in ATC displays it can be all too easy to impair controllers' performance rather than enhance it. In order to make the best use of color, careful consideration must be given to color vision and perception."

(Reynolds, 1997, p. 185)

Color is probably the most effective, compelling, and attractive method available for coding visual information on a display. It can be a very useful tool for organizing and segregating complex information and can have a profound affect on our ability to categorize and search out certain types of information. Color is both attractive and functional. However, this report will show we need to be very careful about using color on displays for air traffic control (ATC) because it is easy to do more harm than good. What is most obvious about the use of color on displays is that its benefits and drawbacks depend upon the task.

1.1 Purpose

The primary purpose of this report is to provide general guidance to the STARS (Standard Terminal Replacement System) and DSR (Display System Replacement) program offices on the use of color on ATC displays. However, the information in this report will be useful to any program considering the use of color on its displays. It will offer general guidelines on how color should, and should not, be used, but will not define a specific color-coding scheme, that is assign colors to meanings. These guidelines are based on what is known about human vision, display capabilities, the knowledge gained from the lessons learned about the uses of color in the cockpit and ATC environments, and human factors "best practices". The report will discuss how the effectiveness of color displays can be assessed and will provide an update on studies that are in progress on specific uses of color on ATC displays. Finally, it will summarize what is known about the general benefits and limitations of the use of color on displays, the efforts of other countries concerning the use of color ATC displays, the guidelines for use of color in the cockpit, and the existing recommendations about using color in ATC displays from air traffic organizations in other countries.

1.2 Experiments Using the SONY DDM-2801C Monitor

This report also describes a series of experiments conducted specifically to support the application of color on the Sony DDM - 2801C (20 x 20) monitor. Extensive research on human performance has determined that our ability to reliably identify colors is limited to five or seven. The seven colors that are easiest to identify on a CRT while using a black or dark grey background are, red, green, blue, yellow, cyan, magenta, and white. Given that a CRT such as the Sony DDM-2801C can produce thousands of "reds," "greens," and "blues," how should these colors be defined to minimize their confusion? Since the answer to this question depends on the color-producing capabilities of the specific type of monitor, a series of studies was conducted to determine the ideal color set for the Sony DDM-2801C. A set of colors is considered "ideal" if the accuracy of

identifying each of the colors is near 100% (thus minimizing the probability of mistaking one of the colors in the set for any other color in the set). It is well known that small patches of color are much harder to identify than large areas of color-filled areas. For this reason, all of the colors in the set must be able to be identified in the smallest size that might be color-coded in actual operations (such as a position symbol). A final requirement is that none of the colors negatively affects legibility. Certain combinations of colors and background (such as pure or "royal" blue on a dark background or yellow on a light background) are very difficult to read and can induce errors. For this reason, we also need to ensure that none of the colors in the proposed color set induce errors in legibility.

Three experiments were conducted. In the first experiment, the color-producing characteristics of the Sony DDM- 2801C (20 x 20) monitors was examined and the variability in the color producing characteristics was measured across five "calibrated" monitors. A set of colors to be tested was then derived from the colors that the monitor can produce. In the second experiment this proposed color set was tested for legibility and color confusability on the Sony monitor. In the third experiment a wider set of colors was tested. This set included the deviant measures found in the first experiment and other colors that had been selected by controllers. These colors were tested on four different backgrounds: black, dark grey, medium grey and light grey. These data describe the variability that can be expected within and between monitors, and describe the performance expected for color identification and legibility using different definitions of colors on several different backgrounds. A detailed description of these experiments is provided in Appendix A.

2. CAUTIONS ON THE USE OF COLOR ON ATC DISPLAYS

Clearly, when applying color to ATC displays, what we don't know can hurt us. There are many unanswered questions, some are more important than others while some are more complex than others.

2.1 QUESTIONS ON THE USE OF COLOR

Some of the unanswered questions that need to be considered are:

HOW EFFECTIVELY CAN A CHANGE IN COLOR (e.g., to red) ATTRACT ATTENTION?

There is reason to believe that blinking is more effective at attracting attention particularly under conditions of high workload or when the alert is in the periphery of the display. In a complex experiment, Thackray and Touchstone (1990) showed that controllers were able to notice flashing radar targets significantly faster than they were able to notice targets that were colored red. Experiments (in progress and discussed on page 17) examine the efficacy of flashing data blocks vs. data blocks that change color for attracting attention. The time required to notice these alerts will be measured under conditions of high and moderate task load.

WOULD A CONTROLLER BE LESS LIKELY TO RECOGNIZE AN IMPENDING CONFLICT BETWEEN TWO AIRCRAFT THAT WERE COLOR-CODED DIFFERENTLY THAN BETWEEN TWO AIRCRAFT THAT WERE COLOR-CODED THE SAME?

This is a critical question that needs to be answered before data blocks are color-coded on ATC displays. It may be the case that color-coding entire data blocks may have a detrimental effect on the controller's ability to predict potential conflicts between two aircraft that are coded in different colors. Color-coding part of the data block (such as the call sign, altitude, or position symbol) may help the controller to categorize the data blocks without the potentially detrimental effects on conflict prediction. This may preserve the ability to predict conflicts while using a redundant color code to help a particular set of data blocks stand out from the others. It may be useful in a variety of situations such as coding "my aircraft" from aircraft not under my control or separating flows of traffic, such as aircraft heading for different runways. Clearly, the risks and benefits of this potentially very useful type of color-coding need to be assessed for different situations.

COULD CONTROLLERS BECOME SO DEPENDENT ON THE COLOR-CODED CUES THAT THEY TEND TO PERCEPTUALLY "FILTER OUT" INFORMATION THAT THEY NEED? IF SO, HOW WOULD THIS AFFECT THEIR PERFORMANCE?

Controllers in the ODID (Operational Display Input Development) IV simulation conducted in Europe "appeared to be looking only for color [in the conflict alert tool they were provided][1] to indicate a traffic problem. This appears to have resulted in lack of memorization of textual and graphical [flight] detail which could reduce their traffic awareness" (Graham et al., 1994 p. 98).

---

[1] As a matter of form, clarifying comments from the author will be put in brackets. Thus, any material in brackets is not part of a direct quote.

We do not yet have a proven color scheme to help controllers do a better job than with using a monochromatic (single color) display. In fact, there is no evidence that adding color to a monochromatic display enhances controller performance. There is a subjective perception that color enhances performance and in several simulation studies controllers *report* that use of color on the situation display helps them do their job better. However, in the only study that examined both how controllers *did* and how they *thought they did* using color and monochromatic displays, controllers thought that they performed better with the color-coded display, when in fact, they did not (Connolly, Spanier, and Champion, 1975). One of the color-coding schemes examined in this study was the color-coding of altitude. Aircraft at level flight at odd altitudes were presented in green, aircraft at level flight at even altitudes were presented in yellow, and climbing and descending aircraft were presented in red. Controllers thought that this color scheme helped their performance. "Of 30 controllers, all but one were rather enthusiastic about color in air traffic displays.....Most controllers reported feeling that color helped them do a more effective job." (p. 24). However, despite the perception that color helped, this color scheme failed to have a measurable effect on the controller's ability to detect and correct potential violations of separation standards.

All of these questions require scientific investigation before a responsible application of color to ATC displays within the U.S. can be formalized. However, there is a vast amount of knowledge on color vision and use of color on displays that leads to logical conclusions about how color could be used on ATC displays in the U.S.

2.2 CAUTIONS ON THE USE OF COLOR

Effects of color-coding are not always intuitive. Many people believe that a change in color, for example changing a data block to red, will automatically attract attention. However, this has not yet been scientifically proven. Coding mechanisms proven to attract attention include dramatic differences or changes in contrast such as blinking. Any display item that blinks will "grab" attention as will any part of the display that is significantly brighter than surrounding items. While it may not be possible for a color change to attract one's attention as effectively as blinking, in all conditions, information that is color-coded well does stand out from the information that surrounds it. This makes color a very useful tool for categorizing information. It is important to remember that there is always an attentional trade-off in using such coding for alerts; what is effective at attracting our attention is also distracting if the alert does not convey useful information. For this reason, it is always recommended that controllers are able to suppress individual alerts.

Color-coding schemes will only be helpful if the "right" information is color-coded. As an example of how the wrong color scheme can be worse than no color at all, consider the following. Kinney and DeCicco (1982) gave 50 controllers at Washington ARTCC in Leesburg, Virginia the opportunity to use a color display while controlling live traffic. In the first "color" mode, weather was presented in orange, map information was presented in red, full data blocks were in green, and limited data blocks and history information were presented in yellow. In the second color scheme presented, weather was presented in red, and maps in orange, with the other coding the same. While using the color mode, controllers always had the opportunity to select the "all-green" mode, similar to what they are used to on the PVD. No objective performance data were taken during this study, but in general, "controllers considered the four-color mode to be worse than or at best, equal to, the all-green mode" (p. vii). While some controllers liked the color mode, most

controllers reported difficulty with the red, saying it was "hard to work with" (p. 2-1) and said that the differences between the red and the orange and the differences between the orange and the yellow were too small for easy recognition.

The appearance of a color is affected by the color around it. Colors can appear lighter, darker, or a different "color" (hue), depending on the color that is next to it ("simultaneous color contrast"), or the color that you looked at just *before* looking at the next one ("successive color contrast"). All of these factors can subtly affect the appearance of colors and need to be considered when selecting specific color schemes for a display.

Certain colors present special considerations. Three FAA flight test pilots encountered an interesting problem when they were examining a prototype cockpit display. One of them had trouble seeing some small blue symbols, another had no problem, and the third could see it, but could also see how someone else might have difficulty. This was puzzling to them because all three pilots had normal color vision.

The answer to this riddle lies in the age of the three pilots. The pilot who had the most trouble seeing the small blue symbol was in his early fifties. The one who had no trouble was much younger. From a display point of view, pure blue (the blue produced solely by the blue phosphor of a CRT) is problematic for a number of reasons, particularly for "older" observers. First, if we are looking at a display that contains several colors, the eye needs to accommodate (change the point of focus in depth) in order to bring a blue symbol into focus. This means that when short wavelengths are in focus, all other wavelengths are slightly out of focus and vice versa. **Small blue symbols or text can appear fuzzy and difficult to read.** When the blue text or symbols are in focus, the other information on the display that is not blue would be slightly out of focus. By approximately age 50, most of us will have permanently lost a portion of our ability to accommodate. This means that the blue symbols would be permanently out of focus and appear fuzzy. For these reasons, **pure blue should never be used for small symbols, text, fine lines, or as a background color.** (See page 8 of this document for more detail.) **Red text can also appear fuzzy and difficult to read (usually due to low contrast).**

There is another interesting age-related factor. As we age - even from 20 to 40 - the cells in the eye that respond to color become less sensitive and the lens of the eye yellows. As the lens yellows, **light blues will appear closer to white.**

Even individuals with normal color vision can have difficulty discriminating certain colors under optimal conditions. **White and yellow are easily confusable, particularly with small symbols (or, similarly, with larger symbols viewed from a distance)**[2]. For example, it could be very difficult to tell whether a small leader line was white or yellow.

Waivers have been granted to controllers with known color deficiencies. **We are uncertain of the exact number of color-deficient controllers currently in the workforce.** Approximately nine

---

[2]You can view this effect, known as small field tritanopia, for yourself on page 95 of Cardosi and Murphy (1995) or in the booklet of the CD. You can also view it by looking at the arrival/departure monitors at the new terminal at Ronald Reagan National Airport in Washington, D.C. or on the upper level at Dulles International Airport. As of this writing, some of the lines on these monitors are printed in white and some are printed in yellow. From a distance, these two colors look the same.

percent of the population has some sort of color vision deficiency (what is commonly called "color blindness"). Also, all color deficiencies are not created equal; there are many different types. The most common form of color deficiency is the inability to tell the difference between red and green. This deficiency is much more common among men than women, affecting eight percent of men and .04 percent of women (Pokorny, Smith, and Verriest, 1979). While most color vision anomalies are inherited, some can also be acquired through disease (for example, glaucoma or diabetes), injury, or as a side effect of certain medications (for example, streptomycin). There are also some minor color deficiencies that will not be picked up with traditional color screening tests. Because of these factors it is not possible to design a color scheme that can be used as effectively by people with abnormal color vision as by people with normal color vision. However, by specifying colors that are maximally discriminable, and by using these colors according to specified guidelines (such as the use of redundant coding), we can help to minimize the probability of errors induced by color-coding.

The appearance of colors on any given monitor will shift over time. All monitors are subject to shifts in color over time. At this time, we do not know enough to be able to predict when or how these shifts will occur. The report on the FAA/Eurocontrol's ODID IV simulation (that used a SONY 20 x 20 monitor) noted that the "...EEC [Eurocontrol Experimental Center] engineers noted that these monitors are frequently maintained due to color shifts over time." (p. 26). This means that displays using color must be periodically checked and recalibrated. **Maintenance procedures need to be in place to ensure the anticipated performance of the monitor and the efficacy of the color codes.**

Tower displays deserve special consideration. The tower (like the cockpit) environment is especially tricky to design because of its wide range of ambient lighting conditions. Daytime light levels in the tower have been measured in the vicinity of 6500fc. (Hannon, 1995). High ambient light, particularly direct sunlight, affects the appearance of colors on the monitor (making all colors appear closer to white) and takes a toll on contrast and readability. Contrast is the key determinant of legibility and some display technologies are more successful than others at producing the contrast necessary to make displays readable in sunlight. The brightness (level of light emitted from the display) of all displays designed for the tower must be easily adjustable, and separate daytime and nighttime color configurations are useful. Glare shields or anti-glare coatings may also be necessary, depending upon the location of the display in the tower. These may also affect color appearance. Design of color tower displays also has other special considerations, many of which are tower specific, such as physical placement of the display and the effect of the shades (many of which are blue) on the appearance of the colors. Another factor that will affect the appearance of the colors on the display is the characteristics of the sunglasses that many tower controllers wear. All of these factors combine to form a complex, but manageable, set of requirements for effective tower displays.

Like sunglasses, certain types of contact lenses can also change color appearance. Tinted contact lenses, particularly the ones designed to change the appearance of the color of your eyes (e.g., from blue to green or from brown to blue) are known to affect color perception. However, the precise effects of these lenses on color confusions has not been systematically studied and is largely unknown (Bill Wooten, personal communication, 1999).

3. HOW SHOULD COLOR BE USED ON ATC DISPLAYS?

It is clear that color could be a useful tool for categorizing information (such as inbound versus outbound aircraft) and helping to search for individual items in that category (such as a particular outbound aircraft). In the laboratory environment, the advantages of the use of color on visual displays is typically measured in visual search tasks. In a visual search task the observer must locate one or more particular objects, for example, letters or numbers; the time required to find the object(s) is measured. Almost without exception, the time required to locate such an object increases with the number of other objects ("distractors") in the display. In one particular visual search experiment, the time to find certain objects increased by 108% when the number of distractors was increased from 30 to 60. However, when the same items in the display were color-coded, the search time increased by only 17% as the number of differently colored distractors grew from 30 to 60 (Carter, 1979). It is this type of finding, along with the strong inherent appeal of color, that suggests an obvious benefit of the use of color for air traffic control displays.

Color is very useful in helping to separate out what we need to pay attention to. It is much easier to search for a specific data block if it is one of several that are color-coded, than if all the data blocks are the same color, because it narrows the number of data blocks that need to be examined. It is also easier to identify a group of data blocks (such as "my aircraft") if this is color-coded. However, this benefit may come at the price of hindering the controller's ability to detect potential conflicts. It may be the case that a controller is not as likely to detect a potential conflict between two aircraft that are color-coded differently than between two aircraft that are color-coded the same.

With the implementation of each new ATC system and subsystem, the amount of information that controllers must process grows at an alarming rate. It is clear that color may be useful, if not necessary, to help organize this information and to reduce the amount of physical and cognitive clutter on the displays. It is also clear that to achieve the expected benefits of color-coding, color must be used in accordance with known principles of color vision and human memory. Human factors "best practices" point to several general guidelines for the use of color on ATC displays. These guidelines are introduced below.

3.1 GENERAL GUIDELINES

Whenever color is used on a display, it should be used redundantly with another means of coding information. This means that there should be some indication, other than color, about the information that the color is to convey. It may be useful to think of redundant coding as a necessary scheme for color-blind users and a safety net for unknown color deficiencies. For example, if the activation of conflict alert caused the data block to turn yellow **and** CA to appear in the data block, this would be a redundant code (although it may not be sufficient for color deficient controllers.) If the conflict pair also appeared in the tabular list, this would give added redundancy.

Care must also be taken to ensure that color is used consistently across all of the displays that a single controller will use. All of the displays should use the same color conventions and meanings assigned to individual colors need to be compatible across displays. For example, if

"aircraft under my control" are color-coded in one color on the situation display, the same colorcoding strategy should be used for "my aircraft" on a conflict probe display.

When colors are assigned a meaning, such as green for "my aircraft" or yellow for "caution", the colors should be readily identifiable and each color should have only one meaning. This means that colors assigned a meaning should be able to be identified with near 100% accuracy, no matter which other color is present (or not present) in the display. Also, the color-coding scheme should be as intuitive as possible so that it is easy to use and remember. This does not mean that yellow could not be used for both Minimum Safe Altitude Warning (MSAW) *and* conflict alert, (since both are cautions) as long as there was a clear indication as to which warning was activated.

The number of colors assigned a different meaning should be limited to six, so that the colors, and their meanings, are not likely to be confused. Due to the limitations of both the color-producing mechanisms of displays and of our human visual system, we are not able to identify more than about six colors with 100% accuracy. Having fewer than seven color-meaning associations also helps to ensure that these meanings will be remembered accurately. This limitation of six colors, and which colors should be used, is discussed in detail in Sections 3.3 and 3.4.

When selecting colors for a display, it is important to consider the physical differences between the colors and the luminance contrast that particular colors will yield. Contrast is a key factor in determining whether or not items on a display will be legible. The American National Standards Institute (ANSI) recommends a contrast ratio of 7:1 for alphanumeric characters, and cites 3:1 as a minimum (ANSI, 1988). To ensure legibility, the International Civil Aviation Organization (ICAO,1993) recommends a contrast ratio of 8:1 for items (such as data blocks) that need to be read. For details that do not need to be read, such as maps and range rings, a contrast ratio of 3:1 (sometimes even less) is acceptable. While these guidelines help to ensure legibility, it should be noted that these contrast ratios may not always be achievable, particularly in the tower. Also, legibility can be demonstrated with lower contrast ratios in many conditions. However, in cases in which these guidelines can not be met, a thorough test of legibility and operational suitability in all anticipated lighting conditions is imperative.

Color-coding schemes should obey cultural color conventions. Each culture has certain associations between specific colors and meanings. In our culture, for example, red means "danger" or "stop" and green means "safe" or "go." These associations are very strong because of our constant exposure to them and our responses to them become somewhat automatic. Because of this they should never be violated in a color-coding scheme. This means that red should never mean anything but danger, alert, or warning. In fact, red is best preserved for situations in which an immediate action is required. This does not mean that all alerts *must* be red; however, when red *is* used, it should convey critical information. Similarly, green should indicate an "OK" status and yellow should be reserved for conveying caution.

From a display point of view, pure blue is problematic for a number of reasons, particularly for observers over 50. While blue text is usually not impossible to read, it is still best to avoid it for text, small symbols, and fine lines on a dark background. A low luminance or "dark blue" (which can be thought of as blue mixed with black) would be useable for a background or for symbols and text on a light background, as would black text on a very light blue background or light blue text

on a very dark background. In these cases, sufficient contrast can be created to support reading and other tasks which require resolution of detail. Pure blue can also be mixed with with green to create "cyan". While cyan is good for legibility, it can be confusable with blue, green, or white depending on how it is defined.

Pure, bright, highly saturated colors should be used sparingly. First, this is clearly a case where "less is more"; highlighting is only effective when there is little of it. Also, these colors should only be used for critical and temporary information so they are not distracting or visually disruptive. Finally, saturated red and blue, when presented simultaneously, can create a false perception of depth.

Any implementation of color will need to be thoroughly tested in the environment in which it is intended to be used. Prototype testing of individual color schemes, such as a scheme for weather, is highly recommended, but does not detract from the need to test the display in its entirety. For example, coding schemes for weather, and special use airspace may be successful when tested independently, but may be incompatible and confusing when presented together.

3.2 CONTROLLERS' SUGGESTIONS FOR COLOR-CODED INFORMATION

We can point to several general guidelines for the use of color on ATC displays, based on what we know about the workings of the eye, color perception, and human memory. Unfortunately, when it comes to *specific* uses of color on ATC displays, more is known about how color should *not* be used than about color codes that might help controllers do their job better. Nonetheless, a reasonable approach is to examine what controllers might like to color code on their display, prioritize this list, prototype an implementation to address an operational need (such as a need to segregate traffic on a display or to try to make overlapping data blocks legible), and then test the prototype.

The "wish list" of what controllers might like color-coded is quite extensive. When asked to brainstorm about color codes that might be useful, the STARS Operations working group generated a list that included, but was not limited to alerts (conflict alert, low altitude alert, aircraft squawking 7700, radio failure, hijack) inbound and outbound traffic, arrivals and departures, pointouts, handoffs, aircraft not under the controller's control versus "owned" aircraft, aircraft type, maps, and weather (notes from the Nov 3, 1997 meeting of the STARS Operations working group). There are a total of over 25 different types of information (characteristics of the aircraft or background (map/weather)) on the list. If everything on this list were coded with a different color, the coding scheme would be too complex to be usable. However, it does provide a useful list of the types of information that may be beneficial to differentiate on the display.

Another use of color on some controllers' wish list is to try to make overlapping data blocks legible. Henry Mertens at the FAA Civil Aeromedical Institute (CAMI) is currently researching this issue. However, V. David Hopkin, a world-renown ATC human factors specialist in the U.K., has done research in this area and reports that, "presenting different labels (U.K. terminology for data blocks) in different colors seldom makes overlapping labels readable" (Hopkin, 1995, p. 227).

3.3 HOW MANY COLORS CAN BE USED SUCCESSFULLY?

Given the fact that the "wish list" of what controllers might like to color code is considerably longer than what can be accommodated, the obvious question is, "How many colors can be used on a display and not detract from its effectiveness?" V. David Hopkin (1977) has recommended that only three or four colors be used on ATC displays, as did Kinney and Culhane (1978). Weitzman (1986) recommended four colors plus white. SAE guidelines for use of color on cockpit displays and other guidelines, such as National Air Traffic Services (NATS) guidelines for use of color on ATC displays in the U.K., recommend five to seven colors.

When color is used for the purposes of assigning specific meaning to specific colors (such as red for emergencies or green for aircraft under my control) it is imperative that no more than six colors be used. When more than six colors are used, two problems are likely to result. First, it is difficult to display the colors so that they are never confusable. Second, it becomes difficult to remember the entire color-coding scheme. Using no more than six colors for identification also capitalizes on the number of colors that are maximally discriminable on a CRT display.

If this limitation of six colors seems restrictive, it may help to understand that the number of useful colors expands greatly when the task depends not on absolute identification, but on discrimination, that is, being able to notice a difference between the colors presented. It has been estimated that the average person can discriminate, (say that two colors look different as opposed to looking the same) several million colors (Chapanis, 1996 p. 221). To illustrate this point, picture a box of 108 crayons. When held side-by-side, a person with normal color vision would be able to say that the crayon labeled "mango" looked different than the crayon labeled "melon". Yet, if asked to identify either crayon alone, they would probably identify each of them as "orange". If a person saw only one of these two crayons and was given the choice of "mango" and "melon" color names, they would probably be no better than chance at naming either one. When only discrimination, and not absolute identification is required, such as in reading a topographical map, the useful number of colors expands greatly. However, we must be careful not to mistake a task that we think depends on the ability to tell the difference between the colors for one that actually requires identification. For example, if weather were to be presented on a display in six shades of a single color, it could be relatively straightforward to tell where the levels of weather were - as long as all six were present. If less than six were present (with no key), the task would now require absolute identification of a subset of the six shades of the single color - a task that would be impossible in an ATC environment. Whenever colors are assigned a meaning, these colors must be able to be identified no matter what else is (or is not) displayed.

This does not mean that only six colors can be used on a display with a black background; it means that no more than six colors (red, green, yellow, blue, magenta, and white) should be assigned a specific meaning. More than six colors can be used successfully on a display if only color discrimination is required. That is, more than six colors can be used as long as the additional colors do not need to be identified, but only recognized as different from the other colors. This could be true for map lines, sector boundaries, or special use airspace. However, if more than these five colors are used for identification then the probability of confusing the meaningful color set on every potential background color (any color-filled areas such as for weather or special use airspace) needs to be investigated. Finally, adherence to the limitations of six colors does not detract from the need to adhere to the other guidelines (such as consistent use of redundant coding).

3.4 WHAT ARE THE SIX COLORS THAT SHOULD BE USED?

Colors used to convey a specific meaning (such as using green data blocks to indicate aircraft under my control) should not be confused with any other color. The six colors that are maximally discriminable on a CRT are:

- Red,
- green,
- blue,
- yellow,
- cyan, and
- magenta.

In addition the color set could include (depending on the background) the achromatic colors of:

- Black,
- grey, and
- white.

Yellow is a combination of red and green light. Cyan is a combination of blue and green light, and magenta is a combination of red and blue. These colors are chosen because they are the ones likely to be the most discriminable, based on their physical differences that correspond to separation in color space. However, precisely how discriminable these colors are will depend upon their precise definition in display producing terms. For example, R, G, B values determine the relative intensity of the red, green, and blue guns in a CRT display (as well as ambient illumination and other factors). The color-producing capabilities will vary as a function of the specific technology used to produce it (for example, flat panel versus CRT or characteristics of the CRT guns), the manufacturer, and model. For this reason, once the maximally discriminable color set is defined for a particular display, these colors should be defined in terms of chromaticity values so that they can be matched as closely as possible from one type of monitor to another. This is important whenever the same person is using more than one color display, so that the red on one display looks like the red on the other display.

3.4.1 Colors for the Sony DDM-2801C

A series of studies were conducted to determine the ideal color set for the Sony DDM-2801C (20x20) monitor. These studies are summarized here and a complete description can be found in Appendix A.

The first study examined five of the SONY DDM-2801C monitors used at the FAA's William J. Hughes Technical Center to see if there were any differences in the CIE (x, y) coordinates and luminance values of the RGB guns (that is, characteristics of the color producing mechanisms) across the different monitors. The next set of studies identified a proposed "ideal" color set based on the color production capabilities of the Sony DDM-2801C and what is known about human color vision; this is referred to as the "derived" color set. A set of colors is considered "ideal" if it minimizes the probability of mistaking one of the colors in the set for any other color in the set and

none of the colors in the set adversely affects legibility. Use of additional colors outside of this set should be thoroughly tested before implementation to ensure that the colors are not confusable; see Section 3.3 for details. The six colors are specified in terms of standardized units CIE *(x, y)* coordinates that can be reproduced by any monitor (that is capable of producing them). For another CRT, this means mapping the CIE *(x, y)* coordinates to R,G,B values (for the red, green, and blue guns). The values tested that were found to result in almost perfect identification of the colors with no detrimental effect of legibility are shown in Table 1.

Table 1. Colors that Yielded 99% Accuracy in Color Identification and Legibility on a Black Background using the Sony DDM-2801C

	CIE x	*CIE y*	*Luminance (fL)*
Red	.630	.341	3.63
	.628	.339	5.17
	.546	.357	7.42
	.593	.335	6.04
	.558	.367	7.53
Green	.289	.610	8.21
Blue	.149	.066	1.57
	.154	.070	3.34
White	.298	.281	13.82
	.274	.286	26.50
Yellow	.444	.488	12.18
Magenta	.298	.152	5.33

Coordinates were found for red, green, blue, yellow, white, and magenta that resulted in at least 99% accuracy in color identification and legibility when presented on a black background. While a total of three different coordinates were tested for cyan, the color identification accuracy on a black background did not exceed 95%. (However, identification of one of the values of cyan did result in 100% accuracy on a dark grey background.)

In addition to the derived color set, other values were also tested in this series of experiments. Deviations (that is, variations in how programmed colors appeared on different screens) that were found across the five monitors were tested. The values for magenta and yellow that were chosen based on preference by controllers working on DSR were also tested. None of the observed deviations adversely affected legibility on the black and dark grey background. However, some had a profound affect on color identification. For example, while the color identification accuracy on a black background for "Blue 1" was 100%, the accuracy for "Blue 2" was only 57%. The "preferred" yellow chosen by controllers resulted in performance that was as good as the "derived" (i.e., chosen based on mathematical derivation) yellow on a black background. However, the preferred magenta was only identified as magenta 81% of the time on a black background, while the derived magenta was identified correctly 98% of the time. For full details on the stimuli tested and their effects on color identification and legibility, see Appendix A.

3.5 SPECIFIC RECOMMENDATIONS

It is useful to think about a display in terms of background and foreground, both in terms of the types of information presented and in the calculation of the primary factor that determines legibility and contrast. Luminance contrast is a measure of the brightness of the foreground relative to the brightness of the background. It can be measured quickly and easily in precise scientific terms. It is even possible to mathematically determine the effect of introducing ambient light on the contrast of specific items on a display. This is useful, in determining whether a particular display would be legible in a tower environment.

When selecting colors for a display, it is important to consider the luminance contrast that particular colors will yield. Contrast is the key factor in determining whether or not items on a display will be legible. Sufficient contrast can make text that is presented in the same "color" (hue) as the background - such as light green on dark green - perfectly legible. If the background (such as weather) is one color and the foreground (data block) is another, but the contrast between them is too low, the information will be unreadable. Simply put, the higher the contrast, the more the information (text, symbol) will stand out. Therefore, the more conspicuous (or obtrusive) you want the information to appear, the greater the contrast ratio should be. It is also important to remember that high luminance, saturated colors need to be used sparingly for several reasons. First, the ability of these colors to "stand out" depends on them being rarely used. Clearly, in this case, less use is more effective. Second, what is good at capturing and maintaining visual attention can also be distracting and visually disruptive if used improperly. Third, saturated red and blue can induce an illusion of depth (this is called "chromostereopsis"), with red appearing closer to the observer than blue.

Finally, it should also be noted that it is never a good idea for controllers to be able to select their own color-coding scheme as a function of personal preference. First, with no conventions, it would be difficult for the controller to remember the meanings that he or she assigned. Second, the color-coding might not be meaningful to anyone else (either a relief controller or a supervisor) who may need to take over.

3.5.1 Background

A true background contains no information. Backgrounds are usually either very dark or very light to achieve maximum contrast. A mid-grey background may seem appealing, because it allows the

use of both lighter colors and darker colors on top of it. However, the contrast ratios that are able to be produced with a mid-grey background can never be as great as with a dark or light background. Black is not usually used as a background because of the glare problem it can produce. For this reason a dark grey background is preferred over black. A light background can have a few advantages over a dark background; glare is not nearly as noticeable on a light background as on a dark background, and the subjective contrast that is created with black text on a light background is greater than that created with light text on a dark background. However, a wider range of colors will be more easily identified on a black or dark grey background than a light background.

Another disadvantage to using a light background is that flicker is more noticeable at higher illuminations. To reduce the probability of flicker and the associated problems (such as distractions, eye strain, headaches), displays that use a light background should have a refresh rate of *at least* 65 cycles per second. In an environment with office level lighting and a display using a light background, a refresh rate of 75Hz or higher may be required to eliminate flicker for all users[3]. The careful selection of a background color is an important part of ensuring that the display is suitably designed for the specific environment, task, and implementation considerations. For example, a light background display may not be suitable for the TRACON if the new displays need to be used next to the old displays during the transition period, since a new display may emit too much light for an adjacent phosphor display to be usable.

Displays designed for the tower should have a daytime and nighttime configuration of background and display colors, and brightness controls that are easily accessible. Glare shields, or anti-glare coatings may also be necessary, depending on the location of the display in the tower. While a dark grey background is suitable for nighttime viewing in the tower, a light grey background may be preferred for daytime viewing. The color schemes chosen for the daytime and nighttime configurations need to be designed with reference to these backgrounds.

While the true background contains no information, it may be useful to think of the background of everything under the data block. For the purposes of this discussion, the background will contain information such as maps, special use airspace, and weather, that the controller would normally not want to "stand out" or be conspicuous. It should be presented in low contrast using muted colors. This is not always the case. For example, there are some tower displays in which the airport map is a level of detail that should be presented in high contrast to support the task. Greys (light grey for high ambient light conditions and dark grey for a dark environment) are good background colors because they are achromatic (technically colorless). All of these factors must be considered together. For example, a light grey map on a dark grey background might provide too much contrast and thus, be more conspicuous than is desirable. Very light (close to white) or very dark (close to black) blues can be used as background colors, as long as these colors are carefully designed. Again, such issues can only be considered in the entirety of the display design.

---

[3]Whether or not a display will appear to flicker depends on many variables, such as the amount of light emitted by the display, ambient illumination, age of the observer, and others. Agaki and Kelly (1991) found that in an office lighting environment where observers were allowed to set their own screen brightness levels, a refresh rate of 75 Hz was required for the monitor to appear flicker-free for 90% of the observers.

The decision about how to color-code certain information cannot be made out of the context of the choices of background color and colors for other information on the display.

3.5.2 Data Blocks

The traditional green is a good color for data blocks on a relatively dark background, because of our natural sensitivity to it, our cultural association between green and "normal" status, and the fact that controllers are accustomed to green data blocks. Green also maintains good brightness contrast over a wide range of saturation. Many controllers would like to see the aircraft under their control displayed in a color that is different from the other aircraft displayed on their screen. This type of color-coding would enable the controller to find their own data blocks very quickly when they are in an array of data blocks that are not theirs. While the benefits to this type of categorization are clear, the potential drawbacks also need to be considered. There is some evidence (ODID IV, p. 23) that this type of color-coding may make it more difficult to detect conflicts between aircraft that are presented in different colors than between aircraft that are presented in the same color. In effect, the fact that "unowned" aircraft are presented in a color that is different from "owned" aircraft makes it easier to unintentionally forget about them - at least when the entire data block is color-coded.

Color-coding data blocks certainly helps the controller (for better or worse) to focus on some aircraft and filter out others. However, it would not necessarily help to alleviate the problems of legibility of overlapping data blocks. David Hopkin has suggested that, "Although colour may reduce clutter if applied with care, its success depends on the successful manipulation of a series of visual layers with the most important information in the top layer, and it is imprudent to rely entirely on colour to separate items in this way: presenting different labels in different colours seldom makes overlapping labels readable" (Hopkin, 1995, pp. 226-227).

Whichever colors are chosen for the data blocks, it will be important to examine the color contrast and luminance contrast produced by each color chosen for data blocks when presented on top of each possible "background". In this sense, the "background" includes any color in a filled-in area such as levels of weather or special use airspace.

Data blocks should be able to be presented at medium intensity so that they stand out from the background. Other conspicuous information, such as alerts, should be presented at sufficiently higher intensity. This does not preclude the controller's ability to adjust the intensity of the data blocks relative to the background. In fact, the intensity of the maps and data blocks should always be able to be adjusted independently. It would be helpful if the intensity of one's own data blocks (coded with the position symbol) were on a separate intensity control. This would allow the controller to selectively adjust the intensity of his or her own data blocks while the intensity of everything else remains the same. This type of tool would help the controller to increase the intensity of their data blocks to quickly find his/her own aircraft among many others and to read their own data blocks (at the expense of the an overlapping one) or to decrease the intensity of their aircraft data blocks to read an overlapping data block.

It will also be important to limit the number of different colors within a data block. It may prove useful to color code portions of a data block (such as altitude). However, data blocks or elements of data blocks that are presented in the same color will have a natural grouping effect and be perceived as a group. The effects of using more than two colors in a data block are unknown. It

is possible that more than two colors within a data block could perceptually fracture the data block and impair performance. While multiple colors in a data block may seem appealing, specific color-coding schemes will need to be tested before they are implemented.

3.5.3 Alerts and Warnings

Alerts and warnings should be presented at high contrast. If they are color-coded, the colors used should also be highly saturated. This mode of presentation should be reserved for critical warnings that require an immediate response and not be used for routine system messages (such as "message waiting"). Red and yellow are customarily chosen to present alerts and warnings because of the learned association of these colors with danger and caution. While it is not necessary to present alerts and warnings in yellow or red, these colors should be reserved for alerts and warnings because of their cultural associations. While controllers may tend to like the idea of alerts presented in red, it is important to note that red is a color that usually can only be produced at low contrast. The implications of low contrast on legibility have already been discussed. Also, if yellow is being used as an alerting function, it should be reserved for this purpose alone. If other (non-alerting) information is displayed in yellow (for example, if the controller draws a map in yellow) then yellow could lose its effectiveness to indicate an alert. Finally, the choice as to whether yellow is used as an alerting color or not should determine the relative luminance used for yellow. If yellow is used for alerting purposes, the luminance of the yellow should be higher than if it is used to present routine information.

3.5.4 Cursor

Since the cursor needs to be visible at all times, it should be presented at a higher intensity than any other information on the screen, to provide maximum contrast with the background.

3.6 DESIGNING A COLOR SCHEME FOR AN ATC DISPLAY

Color displays need to be designed for the tasks that they support and the environments in which they will be used. Ambient lighting is a major determinant of the characteristics of the colors that should be selected for a display. Colors are very difficult to see on a CRT in bright light. Even the brightest colors will appear desaturated ("washed-out") in bright light; this makes the differences between these colors much less noticeable and the colors easier to confuse. In these conditions, bright, highly saturated, colors are necessary, particularly if the display may need to be used in direct sunlight. High ambient light also makes a black background undesirable because of the problem with glare. On the other hand, current TRACONS typically have very low ambient lighting (although this is not uniform). Because of this, the use of bright, highly saturated colors should be minimized so as to provide useful information without being distracting or annoying. The en route (Display System Replacement, DSR) lighting environment is closer to the office level lighting than to either the tower or TRACON. While it is still true that the use of highly saturated colors should be minimized, this lighting level offers more flexibility in the selection of color configurations.

Another factor that must be known before choosing a color scheme is the characteristics of the specific monitor; this will identify the colors that the display is capable of producing. The information, along with what is known about the human ability to identify and remember colors, sets the stage of what is possible. Next, a limited amount of information must be selected to be

color-coded; the fewer the number of color codes, the more effective the color-coding will be. The color-coding scheme must be developed as a whole. For example, a coding scheme for weather needs to be considered in light of all of the other color-coding schemes (e.g., for conflict alert, special use airspace, etc) to be used. **The most important component of a plan for choosing a color-coding scheme for an ATC application is thorough operational testing using representative users, tasks, environment, and operating conditions.**

4. HOW TO MEASURE THE EFFECTIVENESS OF A SPECIFIC USE OF COLOR ON AN ATC DISPLAY

The final step in any plan to implement a specific color-coding scheme should be a thorough test of its effectiveness and operational suitability. The validity of the test results will depend upon the appropriateness, completeness, and sensitivity of the measures used. Any specific application of color can be measured to assess its effectiveness. For example, a study in progress at the Volpe Center is examining whether an alert that changes color is as effective as one that blinks at attracting attention to a conflict alert. In this part-task simulation, participants control traffic in a modification of a computer simulation that was designed by CAMI as a screening test. (Controllers have validated this task realistic [Dana Broach, personal communication,1996].) When conflict alert is activated, it either blinks, changes color, or both. Participants are instructed to respond to these alerts as soon as they notice them by selecting one of the data blocks. In this way, the time required to respond to each type of alert is measured. We also examine the subjective effectiveness of each type of alert by asking the participants to rate their effectiveness.

Scientifically speaking, the only way to isolate the effectiveness of color on a particular display is to have controllers control traffic (in a simulation) using the color display and using the same display with no color. However, adding color to a monochromatic display is not the best way to design a color display. Color displays have special considerations, such as the sizes and shapes of symbols and alphanumerics. Still, it is wise to evaluate a particular color scheme, and make the necessary adjustments, long before it is implemented and used to control live traffic.

Each application of color should be carefully tested before it is implemented. The methods used to measure the effectiveness of color codes on a display are no different than methods used to assess the effectiveness of any other aspect of a display. If certain information needs to attract the user's attention immediately, then aspects of the display, such as contrast, can be measured, and simulations can be devised to measure the time required to notice such alerts. Legibility can be assessed by measuring the accuracy of identifying alphanumerics.

All colors on the display must be readily identifiable and the coding scheme simple and intuitive so that the meanings of the colors are easy to remember. Everything on the display should be easy to read and understand; this can be subjectively evaluated by asking controllers if this is the case.

Finally, the most important measure of the effectiveness of a display is how well controllers are able to do their job using it. This is best measured in a full-mission simulation study where controllers are required to perform all of their usual tasks such as communicating with pilots and coordinating with other controllers using both objective and subjective measures of performance. Objective measures of performance include errors, omissions, and time required to perform certain tasks. Subjective measures include controllers' estimation of their own workload and their assessment of the clarity, legibility, and ease of use of the display. For further information on human factors testing and evaluation refer to Chapter 10 of *Human Factors in the Design and Evaluation of ATC Displays* (Cardosi and Murphy, 1995).

5. USE OF COLOR IN AVIATION DISPLAYS

5.1 COCKPIT DISPLAYS

Color has been used successfully in aircraft cockpit displays for many years. There are many similarities between cockpit and ATC displays. First, both types of displays present complex information required for complex cognitive tasks. Second, this information must be presented in a limited amount of space. Third, both cockpit and tower displays must be able to be used under ambient lighting conditions ranging from bright sunlight to nighttime conditions. The displays used at night must emit light at a low level that does not interfere with the controllers' or pilots' ability to see in the dark. The similarities between the cockpit and ATC environments make the lessons learned and knowledge gained from use of color in the cockpit useful to considerations of use of color in ATC.

Several experts in the field of color vision and displays have compiled recommendations on the use of color in electronic aircraft displays based on lessons learned in the development and use of color displays for the cockpit. The Society of Automotive Engineers (SAE) 1988 Aerospace Recommended Practice, recommends a conservative and consistent use of color, using no more than six color codes for symbols: white, red, green, yellow, magenta, and cyan, while reserving red and yellow for warnings and cautions. These particular colors were chosen to maximize the physical differences between them so that they are not likely to be confused. They note that, "The use of more than six symbol colors may degrade performance on search, identification and coding tasks due both to poorer discriminability (especially under high ambient light) and a loss of organizational value." (p. 5) The document also describes the poor legibility that characterizes blue phosphors and recommends that blue not be used for alphanumerics or symbols containing fine spatial detail.

It is notable that there have been no serious problems reported to the Aviation Safety Reporting System attributable to the color scheme chosen for cockpit displays used for navigation and weather avoidance (Mertens, 1997). Again, it is important to remember that the usefulness of a specific color code is dependent upon what information is coded. In an experiment that color-coded traffic on a cockpit display of traffic information (CDTI), pilots were actually slightly more successful at maintaining self-separation with a monochromatic (single color) display than when the aircraft were color-coded according to proximity to own aircraft (Scallen, Smith, and Hancock, 1997). It is likely that performance in the color-coded condition would have been at least as good, if not better, as the monochromatic condition if more useful information (such as closure rate) was coded instead of distance.

5.2 ATC COLOR DISPLAYS IN THE U.S.

The first descriptions of specific colors to be used in the U.S. are for colors to be used on TRACON displays in the U.S. and can be found in the Advanced Automation System (AAS) specifications and a later document referred to as "CDRL EN09" (or "EN09" for short). Both of these documents suggest specific colors (in chromaticity coordinates) but neither document presents information as to why these particular colors were chosen, nor did they present any

studies to show how these colors could be operationally useful in an ATC environment. For a detailed discussion of these two color sets and color display issues, particularly as they relate to the tower environment (Hannon, 1995).

Color is used on many ATC displays that are used for purposes other than separating aircraft. For example, the interim situation display (ISD) used to display calculated aircraft position in oceanic sectors and the touch screens used for the voice switching and control system (VSCS) are color displays. Only one civilian facility in the U.S., High Desert TRACON (HDT) uses color displays to control traffic. Co-located on Edwards Air Force Base in California, HDT handles a mix of civilian and military traffic. The color scheme used was designed by controllers at the facility. The alerting scheme they chose was red for emergencies (such as hijack) and yellow for conflict alert and minimum safe altitude warning (MSAW). They also chose to color code the data blocks using gold for limited data blocks, white for handoffs, cyan (greenish blue) for aircraft being handled by the controller, and green for aircraft being handled by other controllers. While controllers who use this display like it very much, it is important to remember two key points. First, no studies have been done to determine what effect, if any, this coding scheme has on performance. Second, what works at one facility may not be the best display at busier facilities. Each facility has traffic characteristics that may be different from the needs of other facilities. For example, while their system was being upgraded, HDT did not have some functions that other facilities would find it unacceptable to work without - automated handoffs, conflict alert, and MSAW.

5.3 ATC COLOR DISPLAYS - LESSONS LEARNED FROM OTHER COUNTRIES

5.3.1 Canada

Canada was the first country to publish a report on their approach to the issues associated with the use of color in ATC displays. Transport Canada commissioned Thompson-Hickling Aviation, Inc. to prepare a report on the use of color in ATC displays (Campbell, White, and Hamilton; 1990). This report examined key aspects of color production and perception, the activities conducted in other countries on this topic, and other relevant information, to formulate a list of advantages and disadvantages of color-coding information on ATC displays and develop guidelines for the use of color. Their list of potential disadvantages of color-coding information on ATC displays was as long as the list of possible advantages.

Advantages included:

- "Some increased speed of recognition and visual segregation"
- "Superior recognition and visual segregation of tabulated items in alphanumerics"
- "Perceptual grouping of like-colored items in a random display"
- "Useful for assisting in attention getting under restricted conditions"
- "May reduce the appearance of visual clutter if coding schemes were efficiently used"
- "Probably more aesthetically pleasing (at least in the short term)"

Disadvantages included:

- "Colored objects and icons must meet minimum size requirements that are generally larger than monochrome objects/icons"
- "Limited number of useable colors (probably less than five)"
- "Visual search is not improved if the color-coding of items must be ignored in order to perform the search"

The list of recommendations did not include specific color assignments but consisted of general recommendations such as: using desaturated colors for areas or backgrounds, a minimum size for colored symbols (15 minutes of arc for red and green symbols and 30 minutes of arc for blue and yellow symbols), and keeping the number of colors used to a minimum. Adopting the SAE recommendation, they stated, "More than six simultaneous colors is thought to detract from performance." (p. 58). Again, these guidelines were based on sound principles and state of the art information on the topic, but no independent research on ATC tasks.

Color displays are currently being used to provide air traffic services in Canada in addition to Australia, the United Kingdom, and other countries. While controllers prefer the new color displays over the monochromatic displays, there are no objective data on how the use of color affects controller performance.

5.3.2 The United Kingdom

The United Kingdom's Civil Aviation Authority's (CAA), National Air Traffic Services (NATS) established a group under the Research and Development Directorate to look at how color should be applied to ATC. The group presented a comprehensive summary of research findings relevant to the use of color on CRT displays and various recommendations of the use of colors on displays. The extensive 1990 document was the foundation for the 1992 *Guidelines for the Use of Colour on Air Traffic Control Displays* by Reynolds and Metcalfe. Again, the conclusion was clear that there was no evidence that color aided performance on ATC tasks. However, the sound approach taken by this group and their findings are worth reviewing in detail.

The NATS guidelines present a color palette and display presentation principles developed by Linda Reynolds of the Royal College of Art and controllers using en route and traffic management display scenarios. The colors selected for the display and the "mid-luminance" background were chosen with "normal office lighting levels in mind" (Section 4.2.1), which is important to remember when assessing the applicability to our low illumination TRACONS and towers with their changing light levels. In fact, the authors note that two sets of display colors may be necessary for towers because of the extremes in ambient light experienced in the tower.

The guidelines present a conceptual framework for color-coding ATC information based on the importance to the controller. These "conceptual layers" or classifications of information are assigned layers on the display. There are three basic levels: background, foreground, and alerts.

These basic levels are further subdivided into seven layers where the higher the layer number the more important the information and the more the displayed item must "stand out". The first layer is the background with sector boundaries indicated either with lines or filled-in areas. The second layer is the background detail such as range rings and airways. The third and fourth layers are for radar data; the third layer is for data on aircraft that the controller is not handling and the forth is

for aircraft that the controller is currently handling. The fifth layer is for low-level alerts such as aircraft that require special monitoring. The sixth layer is for aircraft that require the controller's immediate attention. The seventh and highest layer is for the cursor, since the cursor must be visible at all times. In this scheme, muted colors are used except for the alerts.

The procedures given in the NATS document for the use of this color standard represent a sound, principled, approach to the use of color. This is reprinted below from Section 3.2:

(i) Determine High-Level Display Requirements. The objective is to identify the display requirements from an analysis of the ATC task/application. At this stage the requirements will be at a strategic level, i.e., how many levels of alert will be required, what information will be required in the track data block, what detail of map information is required, etc.

(ii) Derive List of Items/Objects to be Displayed. This will be a list of items/objects which will be required to be selectable on the display as determined by the ATC task to be performed. At this stage, the description will be in terms of the ATC requirements and will be independent of display attributes, i.e., airways, airports, beacons, tracks, data blocks (with description of actual data required in the block and any variants), etc.

(iii) Divide Identified Items/Objects into Seven Layers. This stage is still independent of the display attributes. The allocation of an item/object to a particular layer will be made according to the visual priority placed on that item/object (e.g., background map data on Layers 1 and 2, alerting information on Layer 6), and also the graphic form (e.g., background infills on Layer 1, background lines/symbology on Layer 2).

(iv) Allocate Colors to Items/Objects in each Layer. Using the color specification and guidelines in Section 5, and taking into account the high-level requirements identified at step (I), allocate colors to the identified items in each level.

(v) Prototyping. Ideally, a prototyping system should be available so that the color selection can be checked.

This approach to the use of color was applied to the design of the displays for the NATS Swanwick en route air traffic control center in the U.K. The center is not yet operational and the color scheme has yet to be finalized. However, the color scheme proposes a mid-grey background and color-coded data blocks. The data blocks are presented as black text on top of rectangles of color. The exception is if the data block goes into conflict alert. In that case, the text is white on a red background (since black text on a red background would not provide sufficient contrast). The advantage to presenting data blocks on top of colored blocks is that it makes the colors easier to identify. First, it is much easier to identify a large area of color than a smaller one. Second, the color code of the data block remains constant; whereas if the data block was presented alone (and not on top of a colored rectangle), the appearance of the color of the data block would be influenced by the color on the background (which would change as it moved across the display). Presenting data blocks on filled-in rectangles keeps this perception constant. The guidelines also note that "infills of the same color will be more readily perceived as a group than data blocks presented without infills" (Section 4.3.1). The pros and cons of such groupings have already been

discussed. Another possible disadvantage is that these filled-in blocks probably take up more space than the traditional data block, although this has not yet been measured.

Although the advantages of this particular color scheme has not yet been proven in the scientific sense, the approach is consistent with a sound human factor approach to coding information and is an excellent foundation for applying color to ATC displays.

5.3.3 ODID (Operational Display Input Development) IV

Another color ATC display that has received a great deal of attention is the ODID IV. This display is used at the Eurocontrol's Experimental Center to test advanced ATC tools, such as conflict detection tools and alerts. When considering this system, it is important to remember that it is not currently used for air traffic control at any facility. Still, the results of the simulations using the system provide some useful information for the use of color in ATC displays.

The first ODID IV simulation (Graham et al., 1994) evaluated some of the human factors aspects of an ATC system that uses advanced automation tools and electronic flight strips. This simulation did not record any data as it relates to controller performance with the use of the color display, but recorded the controllers' comments with respect to the use of color in this system. In general, controllers thought that color was very useful. They particularly liked the use of yellow as a warning color. However, there was one cautionary note with particularly interesting and serious implications. Controllers "appeared to be looking only for color [in the conflict alert tool they were provided] to indicate a traffic problem. This appears to have resulted in lack of memorization of textual and graphical [flight] detail which could reduce their traffic awareness." (p. 98). More detailed comments from the report can be found in Appendix B.

Nine out of ten controllers who participated in this ODID simulation indicated that they were never confused by the meanings of the colors, but five out of ten controllers said they occasionally experienced difficulty in reading colored text, such as the radar label or Sector Inbound Lists (the other five controllers checked the box between "occasionally" and "never"). These difficulties were primarily attributed to the grey and red text.

Controllers did find the use of color very helpful in categorizing information and coding alerts. For example, an aircraft calling in for the first time would have a pink data block. This would significantly reduce the time required to find the data block among the others. However, this useful categorization may come at a high price if this coding scheme also has the effect of making a controller less likely to notice a potential conflict between two aircraft with different color codes than between two aircraft with the same color code, a factor that was noted, but not investigated, in a later ODID simulation. The distractions noted with large blocks of color should also be taken into consideration in color applications.

Another study using ODID was a joint FAA/Eurocontrol simulation to assess the usefulness of some of the ODID tools for U.S. controllers (Krois and Marsden, 1997). While it is impossible to separate out the effects of color-coding from the effects of the use of these tools (such as short-term conflict alert and medium-term conflict assistance), it is again useful to look at the controllers' subjective assessment of the uses of color they saw. Again, this simulation made no attempt to objectively examine the effect of color on controller performance, nor did they obtain any objective data that could be used to address this point. Still, controller opinion is quite

valuable, as long as we remember that controllers may *think* that color helps their performance in cases where it has been shown to have no effect (Connolly et al., 1975).

The color scheme in this ODID IV simulation used four different non-alerting or "state" color codes for data blocks. A description is presented here, not as guidance, but as background material necessary to understand the controllers' comments that follow.

> GREY - "transferred" (also known as not my aircraft). Grey was used to indicate an aircraft that was outside the controllers sector or one that had been handed off to another frequency.
>
> PINK - "advanced information" (and coordination) state. Pink was used to indicate that the aircraft would be entering the sector in 10 minutes and the "inbound flight could commence entry negotiations, if required." (p. 24)
>
> WHITE - "assumed". White was used to indicate that the aircraft was on the frequency and that the controller had, or was in the process of, taking control.
>
> MUSTARD - "concerned". "Concerned mustard", as they called it, was used to indicate that the aircraft had been transferred to another controller, but was still in the sector (and thus the controller must still "maintain situational awareness of this traffic").

In addition to these "state" colors, two warning colors were used:

> RED - "short term conflict alert". Aircraft call signs within the data block were coded red when there was an "imminent loss of radar separation based on a two minute warning". The rest of the data block stayed the color that it was (Galushka, 1998).
>
> YELLOW - "alert". Aircraft call signs within the data block could turn yellow for one of two reasons. This alert was automatically activated when an aircraft had been transferred to the next sector before receiving a clearance to its exit flight level. Call signs would also turn yellow when a conflict pair (as detected by the system) was selected by the controller to be highlighted.

Another use of color was the "Medium Term Conflict Assistance" tool. In this window, red was used to indicate a conflict, yellow was used to indicate a risk of conflict and grey was used to indicate a potential conflict.

Results of the ODID IV Simulation

Simulation results mention that "controllers supported the use of color to assist them in understanding the ATC situation" and that the use of color displays "contributed to improved effectiveness" (Krois and Marsden, 1997, p. 77). Again, this is based solely on subjective assessment and not on any objective performance data. Still, some of the comments offered by controllers provide useful information on what they found useful and problematic. Controllers "unanimously endorsed" the use of color to indicate status such as "not concerned" (not my aircraft) and warnings. They also liked the use of green to indicate the flight leg with red to show the area of conflict. There were also a number of concerns (noted by Krois and Marsden, 1977) with specific applications of color some of which have serious implications:

> "The Mustard labels did not sufficiently attract the attention of some controllers during their visual scan. For example, *the controller could climb an aircraft and inadvertently not*

be aware of another aircraft because it was not white" [emphasis added] (p. 26) [White was used to code "my aircraft".]

"Grey labels were not conspicuous enough for some controllers and *were sometimes overlooked* [emphasis added] (p. 27). In fact, *one of the operational errors noted during the simulation was due to a controller not being aware of a track displayed as grey* (p. 76).

Of the data blocks displayed in pink, "Controllers stated specifically that they were often unaware of the position of an aircraft at the first frequency contact as a result of the absence of the handoff facility." (p. 26) [In the United Kingdom, controllers do not routinely hand-off aircraft as we do in the U.S. In fact, in London, handoffs are the exception, not the rule. It is interesting to note that coding these aircraft pink was not an adequate substitution for our current handoff procedures for U.S. controllers.]

"Some controllers considered the red call signs "fuzzy" and difficult to read..." (p. 26)

"Some controllers indicated that there should be fewer number of colors used to denote aircraft status, for example, not to differentiate pink and mustard states from white."

6. CONCLUSIONS

The effective use of color on a display requires careful attention to the definitions of the colors used and the ways in which the colors are used. The development of a coding scheme begins with an understanding of the tasks to be performed and identification of subtasks for which color coding may be particularly helpful (such as immediate identification of certain streams of traffic). From this identification of tasks and the information required to complete the tasks, we can construct a hierarchy of display information. Such a hierarchy would include all of the information to be displayed ranging from the background information (e.g., a map) that should be present, but unobtrusive, to critical components of a display (e.g., warnings that require immediate action, the cursor) that require some form of highlighting.

While this hierarchy of display information will help to determine how the information is color-coded, the first decision must be a choice of a background color. As previously discussed, black (or other very low luminance) backgrounds are very good for color identification and help to minimize display flicker, but can suffer from noticeable glare. Light gray backgrounds minimize glare, but color identification can suffer and flicker is more apparent (and potentially distracting) at higher luminances. The selection of a background color should consider the task and operational environment. In a dark environment, a dark background is usually preferred. In high ambient light (such as in an ATC tower), with a display that uses minimal color coding, a light gray background may be preferred. In office level lighting, either light or dark backgrounds are usable, when properly designed.

Once a background color is selected and a hierarchy of information has been constructed, color names can be assigned to specific groups of information. While user input is important at all stages of display development, an efficient strategy is to have structured user input at the first stages of design (e.g., the construction of the information hierarchy) and then at the final stages (evaluating the prototype displays). Human factors specialists are in a better position than users to design the prototype color schemes, since their designs will be based on human factors "best practices" rather than user preferences. There is a tremendous amount of variation in personal preferences for selection of display colors. As we have seen, however, the colors that users prefer are not necessarily the ones that will optimize their performance. Thus, having a prototype color scheme based on human factors guidance not only minimizes variability, but also precludes designs that may have a detrimental effect on performance.

Once a color scheme that matches color names to groups of information has been developed, these color names will need to be defined to be the most maximally discriminable for a specific monitor. Once the color set has been defined, it should be validated by testing to ensure that no two colors in the set are confusable and none of the colors in the set adversely affect legibility. Again, these are tasks for human factors engineers. The next and final steps are a critical review of the color scheme by users and, finally, operational testing and evaluation.

In developing a prototype color scheme, it is very useful to examine the color schemes that other air traffic organizations use. It could be argued that since specific color schemes have been in use at foreign ATC facilities with no disastrous outcomes, the same color scheme should be tried in the U.S. *However, since none of these air traffic control facilities have exactly the same*

amount, mix, and complexity of traffic as the U.S. facilities, it could be a mistake to adopt these color schemes for the U.S. without adequate operational testing.

It is equally important to consider the concern first raised by V. David Hopkin in 1977 and noted in the ODID simulation. *Color-coding the entire data block may be so effective at dividing the aircraft into groups (departure and arrival) that controllers may not notice potential conflicts between two aircraft color-coded differently as well as they would notice potential conflicts between aircraft whose data blocks are the same color.* This is a serious question that requires further study.

There are many cautions about the use of color that need to be considered; it is all too easy to do more harm than good. *However, color remains an effective, compelling, and attractive method for coding visual information on a display.* The use of color in ATC displays presents exciting opportunities, as well as challenges. Color, when used properly, is an extremely useful tool for organizing complex information. Well-designed color displays for ATC are likely to have profound advantages over monochromatic displays. The careful design and testing of specific color-coding schemes with attention to the guidelines presented in this report will help to realize the potential benefits of the use of color while minimizing the potential drawbacks.

APPENDIX A

DETERMINING THE IDEAL COLOR SET FOR THE SONY DDM-2801C

As a result of a request from the FAA's Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100), Integrated Product Team for En Route (AUA-200), and the STARS Product Team (AUA-310), a series of studies was conducted to determine the ideal color set for the Sony DDM-2801C (20 x 20) monitor. In the first experiment described, five of the SONY DDM-2801C monitors used at the FAA's William J. Hughes Technical Center were assessed for variations in color production characteristics. The second and third experiments identified and validated a proposed "ideal" color set based on the color production capabilities of the Sony DDM-2801C and what is known about human color vision. A set of colors is considered "ideal" if it minimizes the probability of mistaking one of the colors in the set for any other color in the set and none of the colors in the set adversely affects legibility. Use of additional colors outside of this set should be preceded by tests to ensure that the colors are not confusable; see Section 3.3 of this report for details. The recommended colors are specified in terms of standardized units (i.e., CIE *x,y* chromaticity coordinates and luminance values) that can be reproduced by any monitor capable of producing them.

In 1995, the National Information Display Laboratory (NIDL) at the Sarnoff Research Center conducted a comprehensive evaluation of a single Sony DDM-2801C monitor (NIDL, 1995). This report identifies many performance parameters and compares the Sony monitor to other monitors that were available at the time. The Sony monitor was the largest in size (19.6 x 19.6 inches) and had the lowest luminance tested (23fL). It also had a uniform gamma curve (relating input to output luminance) that was almost straight on a log-log plot from 1 fL to 20 fL (p.x). The refresh rate is listed in the specifications as 60Hz. Luminance was found to vary as a function of position on the screen by as much as 27% at the highest luminance setting. With respect to color, chromaticity coordinates varied as much as 5% (*x*) and 10% (*y*) with the location on the screen. This means that the appearance of a color could change slightly depending where the color was presented on the screen, particularly at the corners. This study detailed the changes in chromaticity coordinates and luminance level that were recorded at each of 12 positions on the screen. However, since measuring human performance was outside the scope of that study, it could not address how these changes in chromaticity and luminance might affect color identification or legibility. It did also not address how the colors on a monitor might shift over time, or the variability that could be found from one monitor of the same model to another.

A.1 EXPERIMENT ONE: ASSESSMENT OF THE VARIABILITY ACROSS SONY MONITORS

The purpose of this experiment was to determine the variability that could be expected across several different Sony DDM-2801C monitors. Colors are created on a CRT through the excitation of phosphors. Luminance (roughly correlated to the perception of brightness) is determined by the amount of excitation of these phosphors. Software controls which phosphor is excited and the amount of the excitation. Due to factors beyond the scope of this report, variations in phosphor output occur over the surface of the screen. This means that the same

software input may result in chromaticity coordinates that are different for different areas of the screen.

A.1.1 Method

Five Sony DDM-2801C monitors used in the integration and interoperability facility at the FAA's William J. Hughes Technical Center in Atlantic City, NJ were assessed for variations in color production characteristics. All of the monitors were reported to be properly calibrated. Measurements (CIE *x,y* and luminance) were taken using a Photo Research PR-650 spectraphotometer on each of the guns (Red [180, 0, 0], Green [0, 180, 0] and Blue [0, 0, 180]) at nine different positions on the screen. In this way, the variabilities due to the location on the screen and the different monitors could be assessed.

A.1.2 Results

Figure A-1 shows the relation between the software control input and the output luminance observed for one monitor. The resulting slope of 2.72 on log-log coordinates is consistent with what was found by NIDL (1995). The figure shows the CIE plot of the colors able to be produced by the Sony monitor. Also shown are the seven colors chosen for the derived color set. The area bounded by the triangle represents the total color gamut (in CIE coordinates) of the Sony monitor.

The results of the variations of *x, y* coordinates and luminance values across the different screen positions are shown in Table A-1 and are also consistent with the NIDL findings. In addition, variations were found across the five monitors. Documentation concerning operationally acceptable tolerances for *x* and *y* coordinates could not be found. However, the FAA has previously used differences of plus or minus .02 for colors using only one primary (red, green, or blue) and plus or minus .03 for colors using two or more primaries (yellow, cyan, and magenta) [Tarka, 1998, personal communication]. Using these criteria, only seven measurements exceeded these tolerances; three "reds," two "magentas," and one instance each of "green" and "blue." (In Experiment 3, these "deviant" colors were tested to see if they would have a detrimental effect on performance.)

A.2 EXPERIMENT TWO: DETERMINATION AND VALIDATION OF THE IDEAL COLOR SET FOR THE SONY DDM-2801C

If an individual color is assigned a specific meaning, then the information presented in this color is said to be "color-coded." Information (such as text, symbols, or areas) is color-coded when it is presented in a color that has been assigned a meaning (such as red for conflict alert). For this use of color to be successful, a set of reliably identifiable colors must be established. This means that the probability of any color in the set being confused with any other color in the set is near zero when presented on any of the possible backgrounds, such as color-filled areas.

Extensive research on human performance has determined that our ability to reliably identify colors is limited to five or seven. The seven colors that are easiest to identify on a CRT while using a black or dark grey background are red, green, blue, yellow, cyan, magenta, and white. Given that a CRT such as the Sony DDM-2801C can produce thousands of "reds," "greens," and "blues," how should these colors be defined to minimize their confusability? The answer to this question depends on the color-producing capabilities of the specific type of monitor.

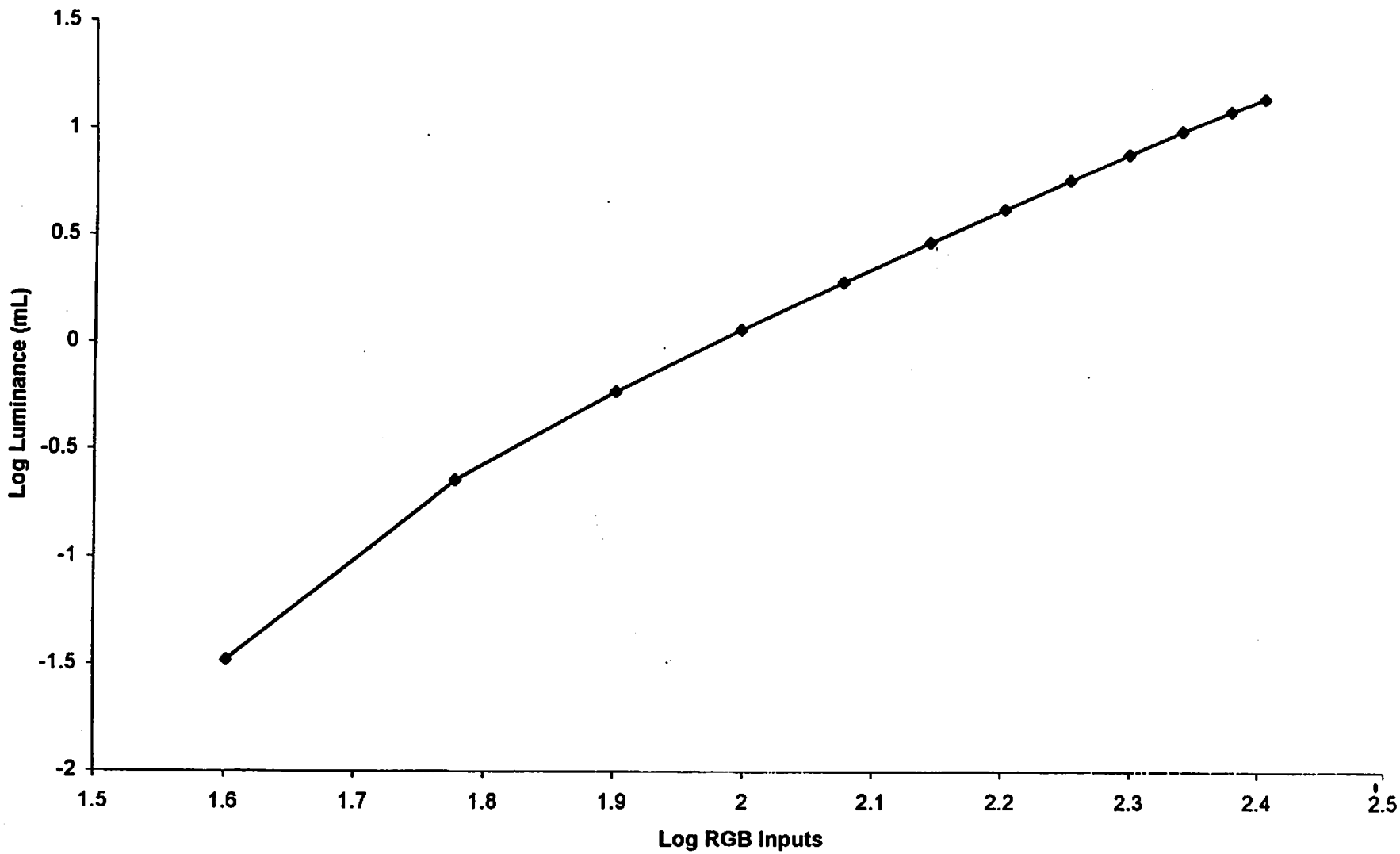


Figure A-1. Log-log Plot of Input versus Luminance for the Sony DDM-2801C

Table A-1. Measurements from Nine Screen Positions on Five Sony DDM-2801C Monitors

Monitor	Color		Left Luminance	Left x	Left y	Center Luminance	Center x	Center y	Right Luminance	Right x	Right y
1	Red (180,0,0)	Top	0.221	0.605	0.346	0.240	0.615	0.342	0.227	0.602	0.342
		Middle	0.228	0.607	0.339	0.273	0.611	0.334	0.226	0.608	0.336
		Bottom	0.217	0.608	0.338	0.231	0.602	0.337	0.206	0.592	0.337
	Green (0,180,0)	Top	0.905	0.292	0.595	0.937	0.293	0.599	0.894	0.291	0.594
		Middle	0.939	0.293	0.599	1.051	0.291	0.603	0.927	0.292	0.602
		Bottom	0.871	0.292	0.593	0.863	0.292	0.598	0.835	0.289	0.598
	Blue (0,0,180)	Top	0.141	0.154	0.070	0.149	0.153	0.070	0.135	0.153	0.070
		Middle	0.148	0.153	0.072	0.158	0.153	0.068	0.144	0.152	0.071
		Bottom	0.143	0.153	0.070	0.141	0.151	0.069	0.140	0.152	0.071
2	Red (180,0,0)	Top	0.157	0.610	0.347	0.172	0.604	0.338	0.153	0.607	0.342
		Middle	0.152	0.625	0.349	0.191	0.621	0.345	0.163	0.608	0.345
		Bottom	0.156	0.605	0.340	0.156	0.603	0.338	0.142	0.595	0.332
	Green (0,180,0)	Top	0.837	0.294	0.597	0.866	0.294	0.598	0.826	0.292	0.597
		Middle	0.843	0.295	0.601	0.977	0.294	0.601	0.829	0.293	0.599
		Bottom	0.832	0.294	0.599	0.777	0.293	0.596	0.770	0.292	0.596
	Blue (0,0,180)	Top	0.150	0.153	0.070	0.154	0.153	0.070	0.147	0.154	0.072
		Middle	0.151	0.153	0.072	0.167	0.152	0.068	0.149	0.152	0.069
		Bottom	0.830	0.292	0.596	0.149	0.154	0.071	0.139	0.153	0.071
3	Red (180,0,0)	Top	0.191	0.575	0.363	0.344	0.524	0.383	0.324	0.512	0.377
		Middle	0.186	0.576	0.359	0.187	0.590	0.359	0.188	0.577	0.360
		Bottom	0.170	0.579	0.356	0.185	0.564	0.354	0.176	0.576	0.355
	Green (0,180,0)	Top	0.803	0.309	0.585	0.978	0.334	0.551	0.960	0.328	0.549
		Middle	0.864	0.305	0.585	0.942	0.304	0.587	0.857	0.303	0.590
		Bottom	0.821	0.303	0.586	0.805	0.303	0.584	0.791	0.301	0.585
	Blue (0,0,180)	Top	0.185	0.174	0.095	0.324	0.213	0.144	0.303	0.209	0.139
		Middle	0.180	0.169	0.090	0.179	0.165	0.084	0.175	0.168	0.090
		Bottom	0.158	0.166	0.086	0.170	0.168	0.089	0.164	0.166	0.089
4	Red (180,0,0)	Top	0.247	0.558	0.349	0.271	0.588	0.351	0.256	0.586	0.350
		Middle	0.271	0.581	0.351	0.249	0.591	0.353	0.265	0.589	0.352
		Bottom	0.233	0.586	0.354	0.246	0.580	0.346	0.238	0.580	0.348
	Green (0,180,0)	Top	1.046	0.307	0.580	1.199	0.302	0.591	1.134	0.300	0.588
		Middle	1.191	0.303	0.587	1.108	0.301	0.590	1.130	0.300	0.589
		Bottom	1.043	0.301	0.586	1.076	0.302	0.588	1.026	0.298	0.587
	Blue (0,0,180)	Top	0.220	0.173	0.095	0.206	0.164	0.083	0.193	0.165	0.085
		Middle	0.219	0.168	0.088	0.206	0.164	0.082	0.198	0.166	0.085
		Bottom	0.186	0.165	0.086	0.186	0.164	0.083	0.178	0.162	0.081
5	Red (180,0,0)	Top	0.316	0.536	0.369	0.286	0.600	0.348	0.241	0.592	0.343
		Middle	0.245	0.589	0.351	0.347	0.582	0.347	0.265	0.602	0.345
		Bottom	0.264	0.595	0.342	0.273	0.593	0.346	0.242	0.594	0.347
	Green (0,180,0)	Top	0.994	0.321	0.561	1.075	0.300	0.588	0.906	0.299	0.592
		Middle	0.973	0.301	0.587	1.311	0.297	0.582	0.992	0.297	0.592
		Bottom	0.973	0.299	0.591	0.958	0.298	0.595	0.898	0.296	0.589
	Blue (0,0,180)	Top	0.278	0.196	0.121	0.207	0.164	0.083	0.159	0.160	0.079
		Middle	0.184	0.165	0.083	0.245	0.164	0.085	0.170	0.157	0.077
		Bottom	0.182	0.159	0.078	0.186	0.160	0.079	0.174	0.159	0.078

The purpose of this study was to determine the ideal color set for the Sony DDM-2801C.- A set of colors is considered "ideal" if the accuracy of identifying each of the colors is near 100% (thus minimizing the probability of mistaking one of the colors in the set for any other color in the set).

It is well known that small patches of color are much harder to identify than large areas of color-filled areas. For this reason, it is also a requirement that all of the colors in the set must be able to be identified in the smallest size that might be color-coded in actual operations, such as a position symbol. A final requirement is that none of the colors negatively affects legibility. Certain combinations of colors and background (such as pure or "royal" blue on a dark background or yellow on a light background) are very difficult to read and can induce errors. For this reason, we also need to ensure that none of the colors in the proposed color set induce errors in legibility on the black background. In addition to the black background two dark grey backgrounds were also tested, since dark grey is often preferred to black because of the reduction in noticeable glare.

A.2.1. Method

A.2.1.1 Derivation of the Ideal Color Set

First, the CIE chromaticity coordinates (in *x, y* values) of each of the guns were mapped to define the colors (in color space) produced by the monitor[4] (see Figure A-2). From this information, *x, y* coordinates of candidate colors were identified for testing.[5]

A.2.1.2 Validation of the Color Set

The resulting color set was validated by testing for confusions of colors within the set and for detrimental effects on legibility.

A.2.1.3 Participants

Seven federal employees of the William J. Hughes Technical Center participated in the study. They ranged in age from 32 to 51 with an average age of 42. All had normal or corrected-to-normal vision and had no known color deficiencies.

A.2.1.4 Task

Each of the colors in the set was used to display an "x" contained in a circle or a filled-in circle on one of five positions (four corners and the center) on the Sony monitor at the Integration and Interoperability Laboratory at the William J. Hughes Technical Center. Participants were asked to name the symbol and its color, that is, whether they saw a filled-in circle (●) or an "x" enclosed in a circle (⊗) and the color of the symbol. Each symbol subtended .27 degrees of visual angle at a distance of 20 inches. The participants were given the names of the colors that they could expect to see and the unfamiliar color terms such as magenta and cyan were described. (Magenta was described as a reddish blue or "purple" and cyan was described as a greenish blue or bluish green.) After a response, another item (filled-in circle or "x" enclosed in a circle) was presented in another color on another portion of the screen.

---

[4] You do not have to minor in the CIE color diagram to understand the basics of what was done; it is only important to know that it is a standardized way of describing colors. However, for those of you who may be after my job, a chromaticity diagram and an example of its use can be found in Cardosi and Murphy (1985), p. 91-92.

[5] The McAdam ellipses were not needed to determine potential confusable color spaces because of the separation between the candidate colors, although the distances between the colors were measured using deltaE* and found to be acceptable.

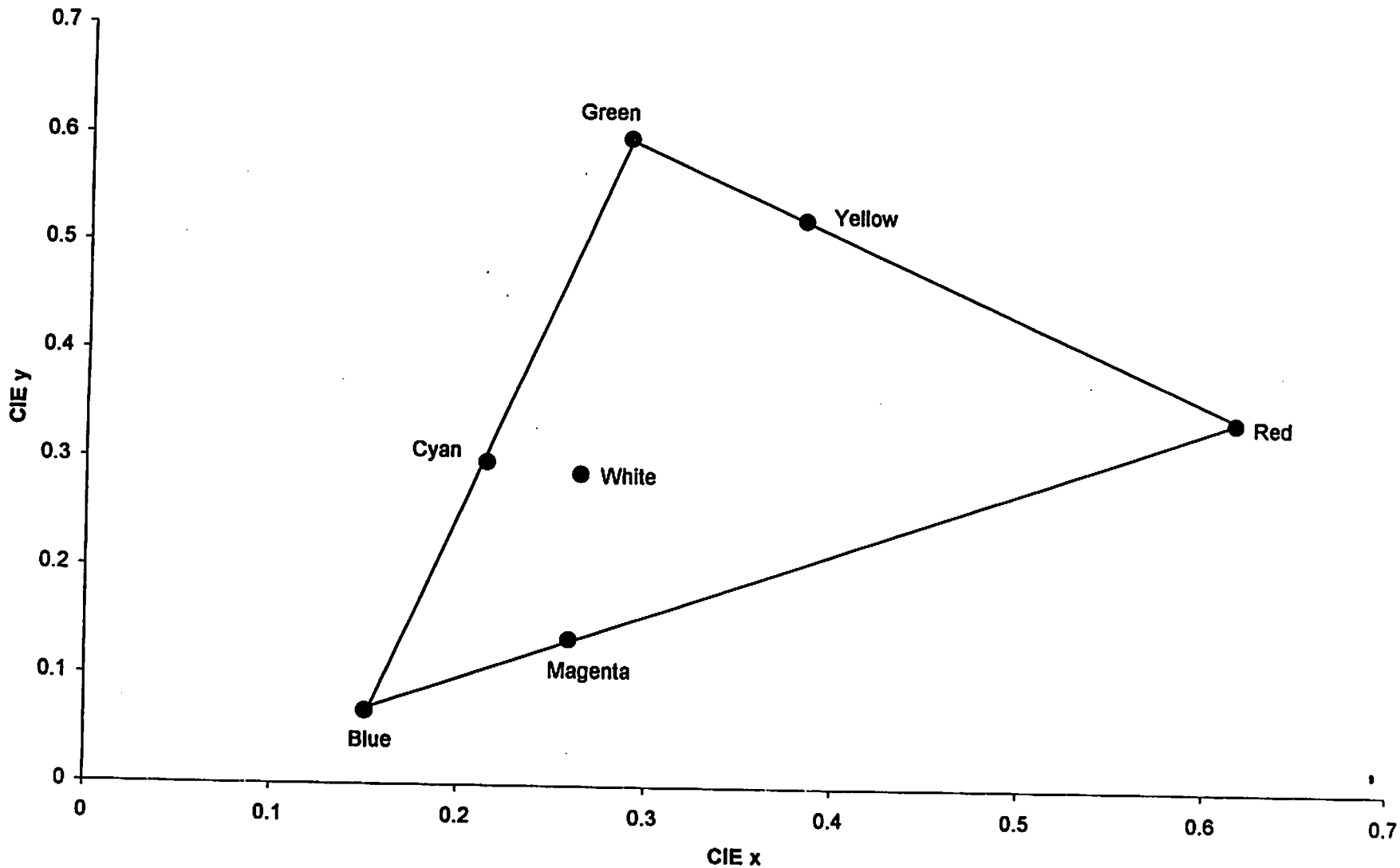


Figure A-2. CIE Plot of the Colors Produced by the Sony DDM-2801C

This procedure was repeated until each person saw each color 50 times (half as the filled in circle and half as the "x" enclosed in a circle) on each of the three backgrounds: black, and two variations of grey in trials blocked by background color. The testing with the black background was conducted with the room lights off to simulate the lighting in a typical TRACON. The testing with the grey backgrounds was conducted with the room lights on.

A.2.1.5 Stimuli

The filled-in circle and the x enclosed in a circle were presented in equal numbers in two sizes - drawn with either eight or twelve pixels. This was used to simulate the two smallest font sizes currently available in STARS.

The backgrounds were defined as follows:

	CIE *x*	CIE *y*	Luminance (fL)
Black	not measurable	not measurable	0
Grey 1	.300	.287	3.46
Grey 2	.299	.285	7.42

The derived color set tested in Experiment 2 was as follows:

	CIE *x*	CIE *y*	Luminance (fL)
Red	.630	.341	3.63
Green	.298	.610	8.21
Blue	.149	.066	1.57
White	.298	.281	13.82
Yellow	.444	.488	12.18
Magenta	.298	.152	5.33
Cyan	.203	.262	9.74

In addition to this set, an additional value was tested for magenta (called "Magenta 2" and defined as x = .334, y = .270, fL = 9.01). This value was tested because it was the one selected by controllers who were helping to choose the color set for DSR.

A.2.2 Result

Legibility (symbol identification) was at least 99% accurate with all of the colors tested on the black and darker grey (Grey 1) background. Legibility was at least 99% accurate with all of the colors tested on the Grey 2 background except for Magenta 2, which yielded an accuracy of 98%.

The accuracy of color identification as a function of color and background is shown in Table A-2. Color identification on the black background was at least 99% accurate for all colors except Magenta 2 (93%) and Cyan (95%). Magenta 2 was most likely to be confused with white and

cyan was most likely to be confused with blue. Results were similar with the Grey 1 background where color identification was at least 99% accurate for all colors except Magenta 2 (98%) and Cyan (98%). Color identification was slightly worse with the Grey 2 background; the same problems with Cyan (96% accuracy) and Magenta 2 (96%) were found and the identification of green was also slightly worse (98%). On the Grey 2 background, only red, blue, white, yellow, and magenta were accurately identified at least 99% of the time.

Table A-2. Percent Accuracy for Color Identification for Experiment 2

	BACKGROUND		
Color Name	Black	Grey 1	Grey 2
Red	100	100	99.7
Green	100	100	98.5
Blue	100	100	99.2
White	99.8	100	100
Yellow	99.3	99.3	99.2
Magenta	99.3	99.8	99.5
Magenta 2	93.5	97.8	96.5
Cyan	95.3	98	96.5

With the exception of cyan, all of the colors in the derived color set were validated on the black background with at least 99% accuracy in color identification and symbol recognition. The magenta proposed for DSR (based on controller preference) was less accurately identified than the derived magenta (93.5% vs. 99.3%). Cyan was identified as cyan only 95% of the time. When there were errors, cyan was most often called "blue" which it is physically very similar to. However, there were not an equal number of errors calling the blue "cyan". If the errors were truly a perceptual confusion of blue and cyan, then we would expect errors in both directions, that is, there would be roughly as many errors in which "blue" was called "cyan" and "cyan" was called "blue". The fact that the errors were found only in the labeling of cyan as "blue" suggests that maybe the unfamiliar color name "cyan" may have contributed to the error rate. Also, since the colors were not shown to the participants beforehand, only described, participants had to rely on their natural (untrained) responses for color naming. It is possible that fewer errors would have been made in the identification of "cyan," had the participants been shown an example beforehand. Experiment 3 was conducted to determine whether cyan is actually difficult to identify because of its similarity to blue and green, or simply a difficult name to recall.

A.3 EXPERIMENT THREE: EXPANSION OF THE RECOMMENDED COLOR SET FOR THE SONY DDM-2801C

As a result of the variability found across five monitors thought to be calibrated, the decision was made to test the range of coordinates that exceeded the FAA tolerance. If these coordinates resulted in criterion performance for color identification and legibility, they would yield a range of usable coordinates for the colors in the set. Such a range would be more useful to designers than a single value for each color name. Additionally, participants would be shown large patches of the colors before the testing was conducted and given a list of the colors that they could expect to see. This would allow us to determine whether the difficulties found with cyan in Experiment 2 were due to its similarity to blue and green or because "cyan" was an unfamiliar color name. At least two sets of values for each color were tested. This included all of the values observed in Experiment 1 that exceeded the established tolerances and the DSR "preference" values for cyan, magenta, and yellow to determine their effect on color identification and legibility.

A.3.1 Method

A.3.1.1 Participants

Twelve paid volunteers ranging in age from 19 to 35 (average age of 26) participated in the study. All had normal or corrected-to-normal vision. After the color names were listed and described (if necessary), the participants were given a pre-test in which they were asked to name large blocks of colors presented on the screen and identify each symbol ("0" or "8") in a pre-test. Two of the participants who suspected they had a color vision anomaly were unable to successfully name the colors in the pre-test and their data were not included in the results presented here. This left a total of 10 subjects on which the data below are based.

A.3.1.2 Task

Each of the colors in the set was used to display a zero (0) or an eight (8) at either one of the four corners or the center of the display. The display used for this experiment was a Pixel Link calibrated color monitor that simulated the colors of interest on the Sony monitor. The CIE x, y coordinates were replicated but the luminance for some of the values was higher than those produced on the Sony. Participants were asked to press the space bar when they were ready to say whether they saw the "0" or the "8" and to name the color it was. After they responded, another number (zero or eight) was presented in another color on another portion of the screen. This procedure was repeated until each person saw each color 50 times (half as zeros and half as eights) on each of four backgrounds: black, dark grey, medium grey and light grey. The trials were blocked by background color.

The requirements for this study were to determine the ideal color set using a black background and low ambient light conditions. In addition, we also tested the proposed color set in office level lighting on a dark grey background (which may be preferred because of the reduced glare), a medium grey, and a light grey background (which may be preferred in high ambient light environments). The purpose was to provide information on the effect of increasing the ambient light level (turning up the lights) on the accuracy of color identification. Testing with the black background was conducted with the room lights off to simulate a typical current TRACON environment; testing with the grey backgrounds was conducted with the room lights on (2fc).

A.3.1.3 Stimuli

The "0" and "8" were drawn with eight and twelve pixels to simulate the two smallest available font sizes in STARS. The height of the "0" and "8" subtended .27 degrees of visual angle when drawn with eight pixels and .36 degrees when drawn with twelve pixels at a viewing distance of 20 inches. The colors used are defined in Table A-3. "OD" is used to indicate that this value was an "observed deviant", meaning this value was observed in Experiment 1 on one of the Sony monitors measured, and exceeded the tolerance of +/-.02 for red green or blue, or +/- .03 for magenta, yellow, and cyan. That is, there was a difference between the color that was "programmed" to be displayed and the color that was observed on the display. "DSR" indicates that these were the values selected by controllers for the DSR program (based on preference) that were significantly different from the derived values for the same color.

The backgrounds used are also defined in Table A-3. The "Dark Grey" background produced a contrast of 5:1 for the red targets. The "Medium Grey" yielded a contrast of at least 2.3:1 for the yellow and red targets; yellow was brighter than the background (positive contrast) and red was dimmer than the background (negative contrast). With the "Light Grey" background, different values for green, yellow, and cyan were chosen so that all colored targets were presented with negative contrast, meaning all of the colored "0"s and "8"s were dimmer than the background.

A.3.2 Results

Legibility, defined as the ability to distinguish between the zero and eight, was at least 99% accurate with all of the colors tested on all of the backgrounds. There was no effect of stimulus size (8 or 12 pixels), meaning that accuracy was no different for the smaller sized numbers than the larger one.

Color identification varied widely with individual colors on the different backgrounds ($F(3, 33) = 12.96, p<001$). Table A-4 shows the accuracy for color identification and average response times (RTs) for the correct responses as a function of the background. Recall that accuracy for symbol identification was consistently at 99% or above. Generally, color identification was very good on the black, dark and medium grey backgrounds but limited with the grey background. On the light grey background only four colors, red, white, magenta 1 and green were identified with an accuracy of at least 95%. The effect of individual colors had a significant effect on the accuracy of color identification ($F(15,164) = 13.68, p<001$) as did the interaction of the color and the background ($F(45,476)= 3.25\ p <.001$). **These results show once again that the ability to identify colors is a function of both the individual colors and the background. Colors cannot be chosen for use on a display without considering the background, nor can the background of a display be chosen without considering the foreground colors to be used.**

The most dramatic differences within a color name were seen with blue and magenta. While "Blue 1" was correctly recognized as blue 100% of the time on a black background, "Blue 2" (an observed deviant) was only identified as blue 57% of the time. Similarly, "Magenta 1" was correctly identified 98% of the time on a black background, while "Magenta 2" (a color previously selected by controllers based on preference) was only identified as magenta 81% of the time. Results were similar with a dark grey background. "Blue 1" was correctly identified as blue 99% of the time, whereas "Blue 2" was identified as blue only 58% of the time. "Magenta 2" was identified as magenta only 75% of the time, while "Magenta 1" was identified as magenta 100% of the time.

Table A-3. Definition of the Color Sets Tested in Experiment 3

Color Name	CIE x	CIE y	Luminance (fL)
Red 1	.628	.339	5.17
Red 2 (OD)	.546	.357	7.42
Red 3 (OD)	.593	.335	6.04
Red 4 (OD)	.558	.367	7.53
Green 1	.289	.596	24.20
Green 2 (OD)	.313	.555	25.10
Blue 1	.154	.070	3.34
Blue 2 (OD)	.201	.131	7.39
Yellow 1	.388	.522	29.20
Yellow 2 (DSR)	.356	.474	29.80
Cyan 1	.219	.299	25.60
Cyan 2 (DSR)	.218	.332	26.20
Magenta 1	.261	.136	8.54
Magenta 2 (DSR)	.399	.208	6.67
Magenta 3 (OD)	.290	.178	10.70
White	.274	.286	26.50

The above colors were used for the following backgrounds:

	CIE x	CIE y	Luminance (fL)
Black	not measurable	not measurable	0
Dark Grey	.270	.284	1.21
Medium Grey	.274	.284	12.1

For the Light Grey Background:

	.274	.281	22.3,

the values for the Reds, Blues, White, and Magenta were the same; the Green, Yellows, and Cyans were defined as follows:

Color Name	CIE x	CIE y	Luminance (fL)
Green	.291	.603	11.00
Yellow 1	.412	.509	9.15
Yellow 2	.411	.509	8.98
Cyan 1	.212	.305	10.30
Cyan 2	.228	.365	10.20

Table A-4. Mean Percent Accuracies of Color Identification and Response Times (RTs) in Seconds for Correct Responses in Experiment 3

	BACKGROUND							
	Black		Dark Grey		Med. Grey		Light Grey	
Color Name	Accuracy	RT	Accuracy	RT	Accuracy	RT	Accuracy	RT
Red 1	100	.95	100	1.18	99.3	1.15	100	1.07
Red 2	100	.96	100	1.05	99.3	1.51	97.2	1.25
Red 3	100	.93	100	.85	98.6	1.24	99.3	1.21
Red 4	100	.97	100	.95	99.3	1.66	97.9	1.28
Green 1	97.2	.96	98.6	.74	93.8	1.26	95.8	1.41
Green 2	93.8	.94	96.5	.84	94.4	1.12	96.5	1.48
Blue 1	100	.96	98.6	1.21	91.0	1.18	85.4	1.15
Blue 2	56.9	.94	57.6	1.18	86.8	2.10	79.2	1.56
White	100	1.03	99.3	.86	99.3	1.04	99.3	3.98
Yellow 1	97.9	.93	98.6	.75	98.6	.99	65.3	1.47
Yellow 2	97.9	.98	97.9	.77	97.9	1.06	63.9	2.06
Magenta 1	97.9	1.01	100	.88	97.2	2.11	98.6	1.40
Magenta 2	81.3	1.23	75.0	.99	59.7	1.62	70.1	1.53
Magenta 3	97.2	1.01	99.3	.86	92.4	3.67	91.7	1.65
Cyan 1	93.1	.98	96.5	.82	98.6	1.05	66.0	1.55
Cyan 2	91.7	.96	100	.78	98.6	1.06	59.0	1.93
Mean	94.0	.98	94.7	.91	94.0	1.47	85.2	1.62
Standard Deviation	18.6	.51	16.9	.88	18.3	1.39	28.4	1.37

Response time was measured by asking the participants to press the space bar when they were ready to respond with the number (0 or 8) and color name. At this point the stimulus disappeared from the screen. Response times reported here are for correct color identifications only. Response times increased significantly with the luminance of the background with the lighter backgrounds associated with longer response times. (F (3, 33) = 23.94, $p < .001$). Response times also varied significantly as a function of the individual colors presented (F (15,164) = 26.07, $p < .001$)

There was also a significant interaction between individual colors and backgrounds in response times. (F (45, 476) = 35.61, $p < .001$). This means that response times were dependent upon combinations of colors and backgrounds and were not solely dependent on either the individual color or the specific background. In fact, even the nature of the color confusion errors changed as a function of the background. For example, when errors in identifying the two values for cyan were made, cyan was mistaken for "white" when presented on the black background and either "blue" or "white" when presented on a dark grey or medium grey background. When confusions were made on the light grey background cyan (using slightly different values and a much lower luminance than those used for the other backgrounds) was confused with "blue" and "green". Again, the successful use of any individual color depends as much upon the characteristics of the background as the characteristics of the foreground. Average response times for all of the correct color identifications are presented in Table A-4. (Note: The means and standard deviations at the bottom of the table are based on all responses.) With the black background, response times ranged from .93 seconds (for Red 3 and Yellow 1) to 1.23 (for Magenta 2) seconds with an average of .98 seconds. Performance was comparable with the dark grey background; average response times ranged from .74 seconds (for Green 1) to 1.21 seconds (for Blue 1) with an average of .91 seconds. With the medium grey background, response times ranged from .99 seconds (for Yellow 1) to 3.67 (for Magenta 3) with a mean of 1.49 seconds. With the light grey background, response times ranged from 1.07 (for Red 1) to 3.98 seconds (for White) with a mean of 1.62 seconds.

In actual operations, the accuracy with which colors can be identified and colored symbols and text can be read is usually more important than a one second difference in how long this process may take; however, such issues are task dependent. The data provided in this study will allow individual questions regarding the use of specific colors on black and grey backgrounds *for specific uses* to be addressed.

In summary, *the light grey background was associated with longer response times and more errors in color identification compared to the black, dark grey and medium grey backgrounds. At least one value for each color except cyan was able to be identified for the Black, Dark Grey, and Grey 1 backgrounds to yield at least 99% accuracy in color identification with no detrimental effect on legibility. With the black background, cyan was the only color for which none of the coordinates tested resulted in criterion performance of 99% accuracy for color identification and legibility. It is also the case that the cyan tested is likely to be confused with white by people with the most common form of "color-blindness". This suggests that cyan should not be used to code critical information (i.e., to convey an important meaning). It also supports the recommendation of using no more than six colors (e.g., red, blue, green, yellow, magenta, and white on a black background) to code information.*

The results of Experiments 2 and 3, as shown in Table A-5, provide color identification accuracies on several different color/background combinations so that programs can set their own criteria (depending on the task) and choose the colors accordingly. "Best" color identification accuracy for each color name/background combination is shown in bold. The "Mean of "best" instances" is the average percent correct of the values (one for each color name) that resulted in the highest identification accuracy. If more than one set of coordinates for a single color name yielded equally high performance, only one of these values was used in the calculation of this mean. All of the "best" instances of colors in the color set for the black background are ones that can be replicated with the Sony DDM-2801C.

In cases where three or more coordinates for a color name resulted in criterion performance (of at least 99% accuracy in color naming with at least 99% accuracy in symbol identification), the values can be plotted on the CIE diagram to define a recommended color space for that color name. This means that any values within that space can be assumed to minimize the probability of color confusions with other colors in the set and have no detrimental effect on legibility (when presented at the same contrast tested). However, red was the only color for which three or more tested values resulted in 99% accuracy on the black background. Using other performance criteria such as a color identification accuracy of 97% or 98% would widen the usable color gamut.

Finally, it should be noted that this study found deviations in color appearance as a function of the location on the screen and as a function of the variability from monitor to monitor. Some of these differences were found to result in a decrease in accuracy for color identification. For example, while the derived blue resulted in 100% accuracy on the black background, the observed deviant of this blue resulted in 57% accuracy. *(We must also expect the appearance of colors on any given monitor to shift over time.)* All monitors are subject to shifts in color over time. We do not know enough to be able to predict when or how these shifts will occur. The report on the FAA/Eurocontrol's ODID IV simulation (that used a SONY 20 x 20 monitor) stated, that the "...EEC [Eurocontrol Experimental Center] engineers noted that these monitors are frequently maintained due to color shifts over time." (p. 26). Displays using color must periodically be checked and recalibrated. **Maintenance procedures need to be in place to ensure that monitors are recalibrated on a schedule that sustains reliable color production.**

Table A-5. Color Identification Accuracies from Experiment 2 and Experiment 3

		BACKGROUND					
Color Name	Experiment Number	Black	Dark Grey	Grey 1	Grey 2	Med. Grey	Light Grey
Red 1	3	100	100			99.3	100
Red 2	3	100	100			99.3	97.2
Red 3	3	100	100			98.6	99.3
Red 4	3	100	100			99.3	97.9
Red 5	2	100		100	99.7		
Green 1	3	97.2	98.6			93.8	95.8
Green 2	3	93.8	96.5			94.4	96.5
Green 3	2	100		100	98.5		
Blue 1	3	100	98.6			91.0	85.4
Blue 2	3	56.9	57.6			86.8	79.2
Blue 3	2	100		100	99.2		
White 1	3	100	99.3			99.3	99.3
White 2	2	99.8		100	100		
Yellow 1	3	97.9	98.6			98.6	65.3
Yellow 2	3	97.9	97.9			97.9	63.9
Yellow 3	2	99.3		99.3	99.2		
Mag 1	3	97.9	100			97.2	98.6
Mag 2	3	81.3	75.0			59.7	70.1
Mag 3	3	97.2	99.3			92.4	91.7
Mag 4	2	99.3		99.8	99.5		
Mag 5	2	93.5		97.8	96.5		
Cyan 1	3	93.1	96.5			98.6	66.0
Cyan 2	3	91.7	100			98.6	59.0
Cyan 3	2	95.3		98.0	96.5		
Mean of "best" instances:		99.1	99.3	99.6	98.9	96.9	87.3

APPENDIX B

SUMMARY OF GUIDELINES FOR THE USE OF COLOR ON ATC DISPLAYS

1. DISPLAYS NEED TO BE DESIGNED FOR THE TASKS THAT THEY NEED TO SUPPORT AND THE ENVIRONMENT IN WHICH THEY WILL BE USED.

All information presented on the display, including the color-coding scheme, needs to be designed to support specific tasks. Color displays also need to be designed for the lighting environments in which they will be used. Design of tower displays requires special consideration for optimum legibility in bright sunlight and low ambient light.

2. USE REDUNDANT CODING.

Whenever color is used to code critical information (such as Conflict Alert, handoff status, MOA), there must be some other signal to the meaning other than color (for example, blinking with "CA" or "H").

3. LIMIT THE NUMBER OF COLORS THAT NEED TO BE IDENTIFIED TO SIX.

When color is used for the purposes of assigning a specific meaning to specific colors (such as red for emergencies, green for aircraft under my control), it is imperative that no more than six colors (for example, red, green, blue, yellow, magenta, and white on a black background) be used. Beyond this, it is difficult to display the colors so that they are never confusing and it is difficult to remember the entire color-coding scheme. This does not preclude the use of more than six colors on a display. However the use of additional colors should be limited to tasks that depend on being able to tell the difference between two or more colors that are always present, rather than tasks that depend upon identifying any single color.

4. ENSURE ADEQUATE CONTRAST.

When selecting colors for a display, it is important to consider the chromatic and luminance contrast that particular colors (foreground and background) will yeild. Contrast is a key factor in determining whether or not items on a display will be legible. For items that need to be read, such as data blocks, a contrast of 8:1 is recommended (but not necessary) to ensure legibility. For details that do not need to be read, such as maps and range rings, a contrast ratio of 3:1 (sometimes less) is acceptable. These guidelines, originally developed by ICAO (1993), are sound principles that ensure legibility. While these contrast ratios may not be achievable in all conditions, especially in the tower, they point to the need for careful testing to ensure operational suitability of displays whenever these standards may not be able to be met.

5. OBEY COLOR CONVENTIONS.

Red should not mean anything but danger, alert, or warning. This does not mean that alerts must be red, only that whenever red is used, it should be used only to convey critical information. Similarly, green should indicate an "OK" status. Yellow is typically used to convey caution.

6. PURE BLUE SHOULD NOT BE USED FOR FINE DETAIL OR BACKGROUND.

From a display point of view, blue is problematic for a number of reasons. First, when short wavelengths are in focus, all other wavelengths are slightly out of focus and vice versa; small blue symbols or text can appear fuzzy. While blue is usually not impossible to read (usually because of the amount of white or green combined with it), it is still best to avoid it for text, small symbols, and fine lines, particularly on a dark background. Dark blue symbols and text would be usable on a light background, because the contrast would be sufficient to support reading and other tasks requiring resolution of fine detail. It is also necessary to avoid pure blue as a background color, although very dark blues (close to black) or very light blues (close to white) could be used as background colors, as long as these colors are carefully designed.

7. BRIGHT, HIGHLY- SATURATED COLORS SHOULD BE USED SPARINGLY.

To preserve the conspicuity of high luminance, highly saturated colors should be used sparingly. Also, these colors should only be used for critical and temporary information so they are not visually disruptive. Finally, saturated red and blue should never be presented simultaneously to avoid a false perception of depth.

8. USE OF COLOR NEEDS TO BE CONSISTENT ACROSS ALL OF THE DISPLAYS THAT A SINGLE CONTROLLER WILL USE.

All of the displays that a controller will use should use the same color conventions; meanings assigned to individual colors need to be compatible across displays. For example, if aircraft under my control are color-coded in one color on the situation display, the same color-coding should be used for "my aircraft" on a conflict probe display.

9. COLOR SET SHOULD BE SELECTED FOR EACH TYPE OF MONITOR AND FOR THE AMBIENT ENVIRONMENT.

The precise definition of the colors that are maximally distinct need to be defined for each monitor to ensure that colors are never confused. Selection of the background color is also an important consideration. For dimly lit environments, such as TRACONs, a dark background is preferred, although an absolute black background should be avoided because of the glare that it invites. A light background (such as light grey) can offer better contrast and significantly reduce the problem of glare. For this reason, it may be preferred for the tower environment during the day, with the understanding that the number of usable colors is lower for a light grey background than for a dark grey background.

10. ALL COLOR SCHEMES SHOULD BE TESTED BEFORE IMPLEMENTATION.

Any implementation of color will need to be tested in the environment in which it is intended to be used. Prototype testing of individual color schemes (such as one for weather) is highly recommended, but does not lessen the need to test the display in its entirety. Coding schemes for weather and conflict alert, for example, may be successful when tested independently, but may be incompatible and confusing when presented together. The entire display (particularly integrated displays) must be designed and evaluated as a whole and not as a combination of parts.

APPENDIX C

DETAILED COMMENTS FROM THE ODID IV REPORT

The first ODID IV simulation (Graham et al., 1994) evaluated some of the human factors aspects of an ATC system that uses advanced automation tools and electronic flight strips. This simulation did not record any data as it relates to controller performance using the color display, but recorded the controllers' comments relating to the use of color in this system.

Some of the positive comments with respect to color included:

> "ODID participants felt that the use of color can support the ATC task, specifically when determining task priorities and recognizing coordination or urgency situations." (p. vii)
>
> "Participants found the role of color-coding in the radar window to be positive. The use of color helped to determine task priorities, e.g., when an aircraft makes its first call the controller searches only the pink radar labels." (p. VII-8)
>
> "Participants would like to maintain the use of red and yellow to distinguish conflict and risk of conflict information." (p. VII-10)
>
> "The consensus of opinion indicates that a color-coded and interactive radar label provides a positive and exploitable interface for controller/system dialogue." (p. 46)
>
> "The controllers stated that the color-planning states, coordination indications, urgency situation warnings, and label shape assisted them significantly in their control task." (p. 46)
>
> "The use of red to indicate short-term conflict alert immediately attracted the controllers attention." [Recall, however, that no performance data were obtained to verify this subjective report.] "However, the usefulness of this warning was lost due to the illegibility of the red call sign. Consideration was given to the other colors to overcome this problem, but the participants felt that red should be retained due to its special significance and that a readable version could be defined." (p. 32)
>
> "The warning color (yellow) was valued as a highlighting tool." (p. 32)
>
> "The use of color to indicate significant operating areas was considered to be extremely useful and visually "distinctive," requiring little interpretation." (p. 32)
>
> "All of the participants considered the "warning" selection (yellow call signs on conflict pairs) to be an extremely useful method for highlighting conflicts. It was suggested that this function would be improved if the warning could also be deselected on individual call signs." (p. 70)

Some of the cautionary comments on the use of color included:

> "The use of too many colors for primary functions can make it difficult for the controller to assimilate the meaning of individual colors." (p. vii)
>
> "The red call signs resulting from activation of the short-term conflict alert function were difficult to read. The red color gave the call sign a "halo" effect." (p. VII-7)

"Controllers did not like the use of large blocks of strong color which were considered to be distracting and without meaning. It was often difficult to read call sign data due to overlapping text." (p. 98)

Controllers in the ODID simulation "appeared to be looking only for color [in the conflict alert tool they were provided] to indicate a traffic problem. This appears to have resulted in lack of memorization of textual and graphical [flight] detail which could reduce their traffic awareness." (p. 98)

In the conclusion, the following comments were offered on the use of color:

"The ODID IV participants felt that the use of color supports the ATC task, specifically when determining priorities such as identifying an aircraft's planning status or recognizing coordination and urgency situations." (p. 87)

"The use of too many colors for primary functions can have a negative outcome. Controllers found it difficult to assimilate color-planning states following the introduction of a fifth-color state in the approach area." (p. 87)

"In general, displaying several colors within the radar data block appears to be acceptable, but there are potential areas of concern. For example:

the use of similar colors (shades) of white to identify the speed vector and the lead line (link between data block and position symbol) caused confusion in maintaining traffic identity (the speed vector of one aircraft could appear to link data blocks and position symbols belonging to other aircraft);

the use of white to indicate coordination in the radar label occasionally confused the EC [their equivalent to our radar controller] as it was not apparent without reference to the message windows whether the coordination was incoming or outgoing. ECs [controllers] were observed trying to accept outgoing coordination." (p. 88)

GLOSSARY

ambient light	Light originating from sources other than the controller's visual display. The general level of illumination in the control room due to sunlight, overhead lights, or individual lamps at the workstation.
brightness	The appearance of a display, or part of a display that ranges from "dim" to "bright". The perception that varies (although not (linearly) with display illumination.
color	A visual sensation that can result when light waves of various wavelengths strike the retina. The main determinants of this perceptual experience are hue, saturation, lightness, and brightness.
color contrast	A comparison of a foreground image to its background based on their relative hue/saturation ratios.
color discrimination	The task of discerning the difference between two or more colors. The ability to differentiate between colors that are present at the same time. Compare to *color identification.*
color identification	The task of associating a unique color name with a specific display color whether or not it is the only color present.
contrast	(also known as luminance contrast) The difference in luminance between foreground objects and their background. See *contrast ratio.*
contrast ratio	A measure of luminance contrast. The luminance of the foreground divided by the luminance of the background.
hue	The component of the experience of color, primarily associated with different wavelengths, that is the primary determinant of the color name such as red, green, yellow, etc.
illumination	The amount of light striking a surface. Measured as the output level of the light source (in candlepower) divided by the square of the distance of the observer from the source. See also *ambient light.*
intensity	The strength of an input or stimulus. In vision, intensity is a main determinant of the perception of "brightness"; in audition, it is associated (again, not 1:1) with "loudness."
legibility	The extent to which alphanumeric characters and text are easy to read.
luminance	The amount of light reflected from a surface in the direction of an observer.
redundant coding	Using more than one means or cue to convey the same information. For example, if conflict alert was conveyed by color-coding *and*

blinking the data block, color would be used redundantly with blinking.

refresh rate The number of cycles per second that the displayed contents of the computer screen are periodically regenerated (usually by a scanning electron beam).

saturation The extent to which a hue is "pure", as opposed to mixed with white, black, or grey. Saturated colors such as pure red appear "vivid", while desaturated colors such as pink (red mixed with a lot of white) appear "washed out."

REFERENCES

Akagi, K. and Kelly, B., 1991. *The Perception of Flicker in Various Ambient Light Conditions and at User Adjustable Screen Brightness Levels.* Proceedings of the Human Factors Society 35th Annual Meeting 1991.

American National Standards Institute (ANSI), 1988. American National Standard for Human Factors Engineering of Visual Display Terminal Workstations. ANSI/Human Factors Society Standard No. 100-1988. Santa Monica, CA: The Human Factors Society, Inc.

Campbell, P., White, R., and Hamilton, A., 1990. *The Use of Colour on Air Traffic Control Displays.* Transport Canada, Publication Number TP 10532E.

Cardosi, K and Murphy, E. (Eds.), 1995. *Human Factors in the Design and Evaluation of Air Traffic Control Systems.* Federal Aviation Administration, Office of Aviation Research, DOT/FAA/RD-95/3.

Carter, R.C., 1979. *Visual Search and Color Coding.* Proceedings of the Human Factors Society 23rd Annual Meeting, pp. 369-373. Santa Monica, CA: Human Factors Society

Chapanis, A., 1996. *Human Factors in System Engineering.* New York, John Wiley and Sons, Inc.

Connolly, D., Spanier, G., and Champion, F., 1975. *Color Display Evaluation for Air Traffic Control.* Federal Aviation Administration Technical Center, FAA-RD-75-39.

EN09, *Controller Interaction and Task Analysis.* CDRL EN09. Revised draft. August 30, 1993. FAA-RE-75-39.

Graham, R.V., Young, D., Pichancourt, I., Marsden, A., and Irkiz, A., 1994. *ODID IV Simulation Report.* Eurocontrol Experimental Centre, France, EEC Report No. 269/94.

Hannon, D. J., 1995. *Human Factors Considerations for Color Displays for ATC.* In *Human Factors in the Design and Evaluation of Air Traffic Control Systems.* Federal Aviation Administration, Office of Aviation Research, DOT/FAA/RD-95/3.

Hopkin, V.D., April 1977. *Colour Displays in Air Traffic Control,* International Conference on Displays for Man-Machine Systems, p. 46-49, University of Lancaster, London.

Hopkin, V.D., 1982. *Human Factors in Air Traffic Control,* NATO Advisory Group for Aerospace Research and Development, AGARD-AG-275.

Hopkin, V.D., 1995. *Human Factors in Air Traffic Control.* Taylor and Francis, Bristol, Pennsylvania.

International Civil Aviation Organization (ICAO), 1993. *Human Factors Digest Number 8: Human Factors in Air Traffic Control.* Circular 241-AN/145. Montreal, Canada, International Civil Aviation Organization.

Kinney, G. and Culhane, L., March 1978. *Color in Air Traffic Control Displays: Review of the Literature and Design Considerations.* MITRE Corporation MITRE Report Number MTR-7728.

Kinney, G. and DeCicco, M., April 1982. *Discussion of Trials of Four-Color Plan View Displays at Washington Air Route Traffic Control Center.* MITRE Corporation MITRE Report Number MTR-82W37.

Krois, P. and Marsden, A., 1997. *FAA ODID IV: En Route Baseline Comparison Simulation Final Report.* Crown Document Number C012-001-008. Crown Communications, Washington, D.C.

Lalomia, M. and Happ, A., 1987. *The Effective Use of Color for Text on the IBM 5135 Color Display.* Proceedings of the Human Factors Society - 31st Annual Meeting, 1987. [Manipulated the colors of foreground and background, measured response time to identifying letters and took preference ratings. Does not cite luminance, contrast or CIE coordinates.]

National Air Traffic Services, Directorate of Research, Chief Scientist's Division, Feb 1990. *Notes on the Use of Colour on CRT Displays*, Internal CSD Report No. 9211, Civil Aviation Authority, London.

National Information Display Laboratory, 1995. *Evaluation of the Sony DDM-2801C Monitor.* Publication No. 572895-043, David Sarnoff Research Center CN 5300, Princeton, NJ.

Pokorny, J., Smith, V., and Verriest, G., 1979. "Congenital Color Defects" in Pokorny, J., Smith V., Verriest, G., and Pinckers, A., *Congenital and Acquired Color Vision Defects.* Grune and Stratton, New York.

Reynolds, L., 1997. *Colour for ATC Radar Displays - A New Approach.* Air Traffic Technology International '97, pp 185-189.

Reynolds, L., 1994. *Colour for Air Traffic Control Displays.* Displays, Volume 15, Number 4, pp 215-225.

Reynolds, L., and Metcalfe, C., 1992. *The Interim NATS Standard for the Use of Colour on Air Traffic Control Displays.* National Air Traffic Services (NATS), CS Report 9213, Civil Aviation Authority, London.

Scallen, S., Smith, K., and Hancock, P., 1997. *Influence of Color Cockpit Displays of Traffic Information in Pilot Decision-Making in Free Flight.* Proceedings of the Ninth International Symposium on Aviation Psychology, April 27 - May 1, 1997. pp. 368-373. Ohio State University, Department of Aviation, Columbus, Ohio 43210.

Society of Automotive Engineers, 1988. *Human Engineering Considerations in the Application of Color to Electronic Aircraft Displays.* Aerospace Recommended Practice, ARP4032.

Thackray, R. And Touchstone, R.M., 1990. *Effects of Monitoring under High and Low Taskload on Detection of Flashing and Colored Radar Targets.* Federal Aviation Administration, Office of Aviation Medicine, DOT/FAA/AM-90/3.

Weitzman, D., May 1986. *The Use of Color in Complex Information Displays.* MITRE Corporation. MITRE Report Number MP-86W16.